



Modelo de predicción del cambio incremental dentro del ciclo de vida tecnológico

Juan Andres Niño Peñalosa

Tesis para optar al título de doctor en Gestión de la Tecnología y la Innovación

Director

Luciano Gallón, PhD

Universidad Pontificia Bolivariana

Escuela de Ingenierías y Escuela de Economía, Administración y Negocios

Doctorado en Gestión de la Tecnología y la Innovación

Medellín, Colombia

2023

Originalidad

Medellín, 2023

Yo, JUAN ANDRES NIÑO PEÑALOSA

Declaro que esta tesis no ha sido presentada con anterioridad para optar a un título, ya sea en igual forma o con variaciones, en esta o en cualquiera otra universidad.

A handwritten signature in black ink, appearing to read 'J. Niño Peñalosa', is written over a horizontal line.

JUAN ANDRES NIÑO PEÑALOSA

Dedicatoria

A quienes creyeron que es posible llegar tan lejos como uno se propone.

Mi amor y gratitud eterna

“Hasta el cielo y tres góticas”

Agradecimientos

Pareciera que esto es un camino que se recorre solo, pero nada más alejado de la realidad, cuando la oscuridad y las dificultades aparecen siempre están los tuyos alrededor, de diferentes maneras y en diferentes momentos todos han contribuido, quiero presentar mi profundo agradecimiento, por acompañarme a soñar y materializar esta etapa de mi vida:

- A mi amada esposa Yohana, por resistir y persistir, gracias por la paciencia ante tantos debates y largas horas de escucharme y guiarme en mis locuras e ideas.
- A mi hijo Andres, un ser de infinita virtud.
- A mis padres Margarita y Nel Francisco, su apoyo y formación tallaron en mi la fuerza para afrontarlo todo.
- A mi hermana Paola, bastión de fortaleza y ejemplo de vida.
- A mis sobrinos por alegrar la vida e iluminar el camino.
- A mi tío Hermes.
- A mis suegros Marleny y Cesar.
- A mi cuñado Julián.
- A mi incondicional amigo William Alejandro, gracias parcerito.
- Al Dr. Luciano Gallón, por ser el faro en medio de la tempestad de afrontar el doctorado, su paciencia y sabiduría me guiaron siempre.
- A todas las personas que no alcanzó a nombrar, los llevo en mi corazón.

Este logro es nuestro, es de todos.

Gracias.

Tabla de contenido

Resumen.....	15
Abstract.....	16
1 INTRODUCCIÓN.....	17
1.1 Problema.....	21
1.2 Hipótesis.....	22
1.3 Objetivo General.....	22
1.4 Objetivos Específicos.....	22
2 ESTADO DEL ARTE.....	24
2.1 Tecnología.....	24
2.2 Cambio Tecnológico.....	25
2.3 Evolución Tecnológica.....	28
2.4 Trayectoria Tecnológica.....	30
2.5 Curva Sigmoidal.....	32
2.6 Ciclos de Vida.....	37
2.6.1 Ciclo de Vida de Onda Corta:.....	38
2.6.2 Ciclo de Vida de Onda Larga:.....	38
2.7 Sistema Sociotécnico.....	39
3 IDENTIFICACIÓN DE ETAPAS EN EL CICLO DE VIDA TECNOLÓGICO	
44	
3.1 Etapas del ciclo de vida tecnológico.....	50

3.2	Creación de límites:	50
3.3	Método 1 de identificación – Distribución Normal	52
3.4	Método 2 de identificación – Derivadas	55
3.5	Prueba de la metodología de identificación de la etapa	57
3.5.1	Frecuencias absolutas y acumuladas	59
3.5.2	Cálculo de la derivada y sus curvas de tendencia	62
3.5.3	Puntos críticos y etapas en la sigmoide	65
4	CARACTERIZACIÓN DEL SISTEMA SOCIOTÉCNICO	69
4.1	Pilares del sistema sociotécnico.....	70
4.2	Variables del sistema sociotécnico	72
4.3	Depuración de variables	76
5	MÉTODOS PARA EL MODELO DE PREDICCIÓN	95
5.1	Modelo 1 – red LSTM	97
5.1.1	Definición de parámetros en el modelo LSTM	98
5.1.2	Pasos del modelo LSTM	99
5.2	Modelo 2 – ARIMA.....	112
5.3	Modelo 3 – Bosques Aleatorios.....	118
6	PRUEBA OPERATIVA DEL MODELO.....	132
6.1	Descripción de la variable a predecir.....	132
6.2	Identificación de la etapa en el ciclo de vida de A01B.....	136
6.3	Aplicación de red LSTM.....	139
6.4	Aplicación de ARIMA	143
6.5	Aplicación de Bosques Aleatorios	147

6.6	Optimización de red LSTM.....	153
6.7	Optimización de ARIMA	156
6.8	Evaluación del modelo.....	158
7	CONCLUSIONES	166
7.1	Limitaciones	170
7.2	Aportes	170
7.3	Trabajo Futuro	172
8	REFERENCIAS	175
	ANEXO - DIVULGACIÓN.....	192
	ANEXO – PROCESO METODOLÓGICO	194

Lista de figuras

Figura 2.1 Etapas del ciclo de vida tecnológico.....	33
Figura 2.2 Relación entre distribución normal y curva sigmoideal	36
Figura 2.3 Ciclos de onda corta y onda larga para TLC.....	39
Figura 3.1 Efectos de sesgos en las curvas de frecuencia absoluta y acumulada	45
Figura 3.2 Restricciones en el ciclo de vida de la tecnología	46
Figura 3.3 Diferentes curvas de ciclo de vida dependiendo de su duración.....	47
Figura 3.4 Distorsión en la curva de ciclo de vida con un punto de introducción erróneo	48
Figura 3.5 Efectos de una distorsión en la emergencia de curvas de ciclo de vida.....	49
Figura 3.6 Proyección de curva normal a sigmoide, con puntos críticos	54
Figura 3.7 Modelo de identificación de etapa en el ciclo de vida.....	57
Figura 3.8 Frecuencia absoluta y acumulada de la tecnología de prueba Pager.....	60
Figura 3.9 Frecuencia absoluta y acumulada de la tecnología de prueba VHS	60
Figura 3.10 Frecuencia absoluta y acumulada de la tecnología de prueba FAX	61
Figura 3.11 Frecuencia absoluta y acumulada de la tecnología de prueba Máquina de escribir.....	61
Figura 3.12 Frecuencia absoluta y acumulada de la tecnología de prueba Floppy	62
Figura 3.13 Uso de derivadas para identificación de las etapas en la tecnología Pager....	63
Figura 3.14 Uso de derivadas para identificación de las etapas en la tecnología VHS.....	63
Figura 3.15 Uso de derivadas para identificación de las etapas en la tecnología FAX.....	64
Figura 3.16 Uso de derivadas para identificación de las etapas en la tecnología Máquina de escribir	64
Figura 3.17 Uso de derivadas para identificación de las etapas en la tecnología Floppy .	65

Figura 3.18 Puntos criticos y etapas identificadas en la tecnología de prueba Pager	66
Figura 3.19 Puntos criticos y etapas identificadas en la tecnología de prueba VHS	66
Figura 3.20 Puntos criticos y etapas identificadas en la tecnología de prueba FAX	67
Figura 3.21 Puntos criticos y etapas identificadas en la tecnología de prueba Máquina de escribir.....	67
Figura 3.22 Puntos criticos y etapas identificadas en la tecnología de prueba Floppy	68
Figura 4.1 Pilares del sistema sociotécnico	71
Figura 4.2 Proceso depuración de variables.....	77
Figura 4.3 Identificación visual de celdas vacías.....	80
Figura 4.4 Eliminación de variables límite de vacíos 60%.....	81
Figura 4.5 Eliminación de variables límite de vacíos 50%.....	82
Figura 4.6 Eliminación de variables límite de vacíos 40%.....	82
Figura 4.7 Eliminación de variables límite de vacíos 20%.....	83
Figura 4.8 Sistema sociotécnico de las categorías de variables incluidas en el modelo ..	91
Figura 4.9 Radar de áreas en el pilar B.....	92
Figura 4.10 Radar de áreas en el pilar C.....	93
Figura 4.11 Radar general de áreas y variables en el modelo.....	93
Figura 5.1 Distribución de las variables D1 A D24	103
Figura 5.2 Distribución de las variables D25 A D49	103
Figura 5.3 Distribución de las variables D50 A D74	103
Figura 5.4 Distribución de las variables D75 A D100	103
Figura 5.5 Distribución de las variables D101 A D124.....	104
Figura 5.6 Comportamiento de las ventas de predicción.....	107

Figura 5.7 Comportamiento de la variable a predecir	114
Figura 5.8 Primera predicción ARIMA	117
Figura 5.9 Desempeño para n_features	125
Figura 5.10 Desempeño para n_estimators.....	127
Figura 5.11 Resultado de la predicción Bosques Aleatorios V1.....	129
Figura 5.12 Importancia de variables	130
Figura 6.1 Frecuencia absoluta y acumulada de la tecnología A01B.....	137
Figura 6.2 Uso de derivadas para identificación de las etapas en la tecnología A01B ...	138
Figura 6.3 Puntos crítico y etapas identificadas en la tecnología A01B	139
Figura 6.4 Funciones de pérdida rondas 1 a 7 LSTM	142
Figura 6.5 ARIMA 1	143
Figura 6.6 ARIMA 2	144
Figura 6.7 ARIMA 3	145
Figura 6.8 ARIMA 4	146
Figura 6.9 Resultados Ronda 1 Bosques Aleatorios	147
Figura 6.10 Resultados Ronda 2 Bosques Aleatorios.....	148
Figura 6.11 Resultados Ronda 3 Bosques Aleatorios.....	149
Figura 6.12 Resultados Ronda 4 Bosques Aleatorios.....	150
Figura 6.13 Optimización de LSTM	155
Figura 6.14 ARIMA 5.....	157
Figura 6.15 ARIMA 6.....	158
Figura 6.16 Mejores resultados de los métodos	160
Figura 6.17 Comportamiento de las 125 variables.....	161

Figura 6.18 Comportamiento de las 26 variables.....	162
---	-----

Lista de tablas

Tabla 1 Tecnologías de prueba para identificación de etapa del ciclo de vida.....	58
Tabla 2 Condiciones de fuentes de datos.....	73
Tabla 3 Variables iniciales del sistema.....	75
Tabla 4 Variables depuradas en el sistema	78
Tabla 5 Celdas vacías por año.....	79
Tabla 6 Variable definitiva del pilar A	84
Tabla 7 Variables definitivas del pilar B	84
Tabla 8 Variables definitivas del pilar C	86
Tabla 9 Estructura de los pilares del modelo	90
Tabla 10 Variables relacionadas entre pilares.....	91
Tabla 11 Ventanas de observación.....	99
Tabla 12 Extracto características de los datos.....	100
Tabla 13 Calculo de error absoluto del modelo ARIMA.....	118
Tabla 14 Resultado predicción Bosque Aleatorio.....	128
Tabla 15 Estructura de la clasificación de patentes (2019).....	133
Tabla 16 Grupos de descarga patentes A01B	134
Tabla 17 Rondas de optimización LSTM.....	140
Tabla 18 Ronda 1 ARIMA.....	143
Tabla 19 Ronda 2 ARIMA.....	144
Tabla 20 Ronda 3 ARIMA.....	145
Tabla 21 Ronda 4 ARIMA.....	146
Tabla 22 Resumen rondas de optimización Bosques Aleatorios	151

Tabla 23 Variables de mayor aporte a la predicción	151
Tabla 24 Optimización de red LSTM.....	153
Tabla 25 Ronda 5 ARIMA.....	156
Tabla 26 Ronda 6 ARIMA.....	157
Tabla 27 Distribución del sistema sociotécnico simplificado.....	162

Lista de ecuaciones

Ecuación 1. Derivada.....	55
Ecuación 2. Frecuencia absoluta acumulada.....	98

Resumen

La tesis presenta un modelo de predicción del cambio tecnológico incremental basado en la combinación de las teorías de ciclos de vida y evolución tecnológica, para ello se soporta en la identificación de la etapa en la que se encuentra una tecnología (introducción, crecimiento, madurez, declive) dentro de su propio ciclo de vida y la inclusión de variables del sistema sociotécnico asociadas, las que se ordenan en tres pilares: cambio tecnológico generador, interés económico y condiciones sociales circundantes, actuando como fuente de datos para el modelo de inteligencia artificial predictivo.

Los datos utilizados en el modelo provienen de fuentes oficiales con acceso abierto a nivel mundial, permitiendo la predicción independientemente de la duración del ciclo y su estado de avance, para ello se abordan patentes por métodos matemáticos y estadísticos en la fase de identificación de etapas, mientras el modelo predictivo utiliza 124 variables adicionales, procesadas por técnicas de *machine learning* supervisado.

Son comprobadas las teorías de cambio tecnológico y comportamiento de ciclos de vida expresados en curvas sigmoidales, obteniendo un método funcional de identificación de puntos críticos de cambio de etapa y un modelo de alta precisión predictiva multivariable, limitándose a identificar la etapa actual del ciclo de vida y la predicción del siguiente cambio tecnológico evolutivo, sin determinar la duración de la etapa en curso o las siguientes, aportando la técnica necesaria para estudiar cualquier tecnología y ofreciendo una herramienta útil para la toma de decisiones en la gestión de recursos y esfuerzos en el cambio tecnológico.

Abstract

The thesis presents a model for predicting incremental technological change based on the combination of the theories of life cycles and technological evolution. It is supported by the identification of the stage in which a technology is found (introduction, growth, maturity, decline) within its own life cycle and the inclusion of associated sociotechnical system variables, which are ordered into three pillars: generator technological change, economic interest, and surrounding social conditions, acting as a source of data for the predictive artificial intelligence model.

The data used in the model come from official sources with open access worldwide, allowing for prediction independently of the duration of the cycle and its advancement status. This is achieved by addressing patents using mathematical and statistical methods in the stage identification phase, while the predictive model uses 124 additional variables processed by supervised machine learning techniques.

The theories of technological change and life cycle behavior expressed in sigmoidal curves are tested, obtaining a functional method for identifying critical points of stage change and a accuracy highly predictive multivariable model. The model is limited to identifying the current stage of the life cycle and predicting the next evolutionary technological change, without determining the duration of the current stage or the next stages. It provides the necessary technique for studying any technology and offers a useful tool for decision-making in resource and effort management in technological change.

1 Introducción

El documento presenta una visión del futuro de las tecnologías, entendida desde la comprensión de la evolución alcanzada en sus propios ciclos de vida, como base del modelo predictivo se identifica la etapa alcanzada (introducción, crecimiento, madurez, declive) y desde este punto se efectúa la predicción de su comportamiento, medido en la cantidad de cambios incrementales en el siguiente periodo de tiempo.

En el capítulo dos se presentan las bases teóricas de ciclos de vida y de evolución tecnológica, las que han desarrollado en las últimas décadas diferentes profundizaciones, evaluándolas en búsqueda de comprobarlas por métodos matemáticos y estadísticos, con un procedimiento que permita la identificación de la etapa del ciclo de vida en que se encuentra la tecnología y se genere el ordenamiento que permita comprender la dinámica de predicción del cambio tecnológico aplicando un modelo de inteligencia artificial.

El tercer capítulo aborda la evaluación de los métodos de identificación del ciclo de vida, para ello se toman cinco tecnologías de prueba (VHS, Pager, Máquina de escribir, Fax y Disco Floppy), todas ellas en etapa de obsolescencia, con el objeto de evidenciar el cumplimiento de los criterios de identificación de los puntos de cambio dentro de las etapas, sobre estas tecnologías se ejecutan pruebas y depuración del método hasta lograr su funcionamiento adecuado.

En el cuarto capítulo se exponen los datos usados en el modelo, los que son recolectados de instituciones oficiales con acceso abierto a nivel mundial, organizadas sobre la teoría de sistema sociotécnico, la que está compuesta por tres pilares: el cambio tecnológico generador, el interés económico en el área y las condiciones sociales circundantes, esta construcción del

sistema admite la inclusión de variables relacionadas con cada tecnología de manera directa e indirecta, permitiendo al final determinar las de mayor aporte predictivo en el cambio tecnológico según la etapa del ciclo de vida donde se encuentre cada tecnología.

En referencia al sistema sociotécnico se abordan inicialmente 31 millones de datos, divididos en 548 variables, las cuales constituyen la base sobre la cual se ejecutan depuraciones de calidad, homogeneidad, relevancia y completitud, obteniendo un *data set* final de entrenamiento del modelo de 125 variables, en un horizonte de 50 años iniciando en 1970 y finalizando en 2019, este conjunto de variables corresponde a la organización en los tres pilares del sistema sociotécnico mencionados anteriormente, enfocados en el área de agricultura.

El planteamiento del modelo se ubica en el quinto capítulo, tomando como insumo la etapa identificada, se exploran y comparan técnicas de predicción como *Autoregressive Integrated Moving Average* - ARIMA, redes neuronales *Long Short-Term Memory* - LSTM y Bosques Aleatorios, en las cuales se insertan las variables del sistema sociotécnico, obteniendo configuraciones de hiper-parámetros que se optimizan internamente y luego por aporte cruzado entre técnicas, hasta obtener un modelo funcional con capacidad predictiva de alta precisión para el siguiente cambio incremental.

En el capítulo sexto se ejecuta el modelo sobre la tecnología seleccionada de arados, correspondiente a la subclase A01B de la Organización Mundial de la Propiedad Intelectual - OMPI, el periodo de referencia inicia en 1800 y termina en 2019 con un total de 234.842 patentes, sobre las cuales se aplica el método de identificación de etapas de ciclo de vida, encontrando que la tecnología no ha cumplido su ciclo completo, ubicándose en la etapa dos. Desde este resultado se desarrolla el modelo predictivo de *machine learning* con redes LSTM y Bosques aleatorios, obteniendo una alta precisión predictiva.

De esta manera se comprueban los planteamientos teóricos de los ciclos de vida propuestos por Geels (2005), la división en etapas propuesta por Levitt (1965), el proceso de adopción tecnológica de Bass (1969) y las trayectorias tecnológicas de Dosi (1982) y Pavitt (1984), aportando un método aplicable a cualquier tecnología sin importar la duración de su ciclo de vida y el alcance evolutivo obtenido, además de presentar el uso de técnicas de inteligencia artificial en modelos predictivos aplicados a tecnologías en constante evolución, obteniendo una mirada del futuro que permite gestionar y prepararse para los cambios esperados.

La comprobación matemática, así como la presentación del modelo de identificación de la etapa alcanzada por una tecnología dentro de su propio ciclo de vida, impulsa la teoría general de evolución tecnológica, superando la etapa teórica al lograr su comprobación, de esta manera el modelo de cambio tecnológico de Schumpeter, desarrollado por Dosi (1982), Freeman (1982), Pavitt (1984), Nelson y Winter (1982) en la década de 1980 y ampliado en la década 2000 por Geels (2005), Verspagen (2007) y Pérez (2010), cuenta con un método de aplicación que funciona tanto para ciclos de vida tecnológicos, como para ciclos de vida de producto, generando un aporte a la teoría seminal sobre la cual se soporta la gestión tecnológica, avanzando en el entendimiento de la dinámica del cambio evolutivo.

La obtención de predicciones con alto nivel de precisión permite la toma de decisiones sobre inversión de esfuerzos y recursos en las tecnologías de mayor impacto para la sociedad, así como un abanico de posibilidades sobre la administración del cambio al identificar los momentos clave de adopción tecnológica, de transferencia y de cambio paradigmático.

Pese a que el modelo se encuentra limitado a la detección de la tapa del ciclo de vida tecnológico, sin llegar a determinar la duración de esta misma etapa o al lapso necesario para el cambio a la siguiente, logra predicciones muy precisas del siguiente valor evolutivo.

Desde este punto es posible iniciar procesos de gestión de cambio tecnológico y enfocar los esfuerzos en tecnologías claves que se encuentran en fases de desarrollo y que pueden tener implicaciones positivas para el progreso de la sociedad y la industria, por otra parte amplia la comprensión del cambio tecnológico, al conocer las etapas recorridas y el camino que falta por transitar, permitiendo la identificación temprana de tecnologías que cambiarán el paradigma, la detección de tecnologías maduras que serán reemplazadas en futuros cercanos, la velocidad evolutiva de tecnologías emergentes, logrando construir un espacio en el cual el cambio tecnológico deja de ser un proceso azaroso, para convertirse en una estructura predecible y manejable que permite gestionar tanto externalidades positivas, como negativas que pueden surgir asociadas a una tecnología.

Finalmente se posibilita la creación de análisis de grupos tecnológicos, entendidos como todas las tecnologías individuales que hacen parte constitutiva de una tecnología mayor o de un artefacto, logrando identificar cuales se encuentran empujando el cambio y cuales son la barrera que impide la mejora de la propia tecnología o del artefacto, mediante la construcción de grupos tecnológicos es posible entender con mayor profundidad la evolución y optimizar la configuración de las tecnologías involucradas.

1.1 Problema

Las tecnologías evolucionan y se transforman, este comportamiento ha sido modelado como una curva sigmoideal (Bass, 1969a) (Rogers, 1976a), llamado ciclo de vida de la tecnología dentro del cual se han identificado cuatro etapas en la trayectoria: introducción, crecimiento, madurez y declive (Dean, 1950a) (Levitt, 1965a), cada una de ellas con características asociadas a la misma tecnología, su impacto y uso.

La predicción del comportamiento de una tecnología observado desde su curva de ciclo de vida, reviste una complejidad creciente por el sistema sociotécnico que la soporta (Geels, 2005c) (Geels & Raven, 2006a) (Ravena et al., 2012a) y las propias particularidades de cada tecnología. El establecer una predicción del comportamiento de la curva sigmoideal condicionada a la detección de la etapa del ciclo de vida donde se encuentra y desde allí analizar su trayectoria tecnológica futura dentro de la misma etapa o el paso a la siguiente (Pavitt, 1982a) (Pavitt, 1984a), posibilita una identificación más ajustada del cambio incremental posterior y del lapso necesario para su aparición.

El modelo de ciclo de vida tecnológico describe las etapas que lo conforman, este se puede observar con claridad en una tecnología con ciclo completo, en el caso de ciclos de vida parciales o en proceso de desarrollo, no se encuentra dentro de la literatura un método de identificación de la etapa de ciclo de vida tecnológico, por lo cual no se pueden gestionar las condiciones propias de la etapa donde se encuentra y trabajar en el desarrollo de la siguiente.

La predicción del cambio incremental permite su gestión al concentrar los esfuerzos, optimizando la asignación de recursos y obteniendo resultados de mayor impacto, puesto que el cambio tecnológico es acelerado, es necesario identificar cuál es su posible comportamiento futuro

(Huenteler et al., 2016a) y enfocarse en las áreas de cambio más prometedoras y alejarse de los estancamientos tecnológicos.

El desconocimiento del estado de avance de una tecnología dentro de su propio ciclo de vida (Freeman et al., 1982a) (Foster, 1985a) genera distorsiones entre los esfuerzos realizados por las unidades de I+D, la inversión de recursos de las empresas o los gobiernos y los posibles resultados, creando una diversificación de opciones de cambio incremental de la misma tecnología (Christensen, 1992a) y dado que de las vertientes de cambio incremental generadas solo una será dominante (Dosi, 1982), las restantes serán fracasos que configuran un proceso infértil y un desgaste de recursos.

1.2 Hipótesis

Un modelo de predicción del cambio incremental de una tecnología, basado en la identificación de su posición dentro de las etapas del ciclo de vida y en su comportamiento histórico, incluyendo el sistema sociotécnico que la afecta, responderá a una evolución temporal del tipo sigmoide, la cual posibilita la identificación de su trayectoria tecnológica reduciendo la incertidumbre de su futuro comportamiento y segmentando la probabilidad de cambio.

1.3 Objetivo General

Construir un modelo de predicción del cambio tecnológico incremental, que permita la identificación de su comportamiento futuro, basado en la posición dentro de su propio ciclo de vida tecnológico y su trayectoria histórica, usando como laboratorio de datos el sector agricultura.

1.4 Objetivos Específicos

- Establecer el método de identificación de la fase en que se encuentre la tecnología dentro de su propio ciclo de vida, por medio de las series temporales y la creación de

límites entre las fases según la pendiente de la curva.

- Caracterizar el sistema sociotécnico que afecta el sector agrícola a nivel internacional, determinando las variables que aportan al modelo.
- Explorar las metodologías estadísticas adecuadas para ejecutar el modelo de predicción y seleccionar la que mejor se adecue a la estructura de datos.
- Ejecutar el modelo de identificación de la fase y la predicción, con información de bases de datos abiertas para el sector agricultura según la clasificación internacional de patentes 2019 a nivel de subclase.
- Evaluar la efectividad del modelo, mediante la inclusión de un rezago temporal para comparar la predicción con los resultados reales obtenidos.

2 Estado del Arte

La predicción del cambio tecnológico es un proceso con múltiples facetas, por lo que es necesario separar cada una de ellas para abordar su desarrollo teórico individual, establecer los estados de la investigación relevantes, así como los últimos avances en el área, y con esta información construir un corpus general que soporte el propio desarrollo metodológico.

De esta manera se abordan los conceptos de tecnología, cambio tecnológico, evolución tecnológica, trayectoria tecnológica, curva sigmoïdal, ciclo de vida y sistema sociotécnico, presentando sus exponentes y vertientes para establecer entre ellas el constructo teórico base para el modelo de predicción.

2.1 Tecnología

Las tecnologías nos rodean y hacen parte de la cotidianidad hasta el punto de tomarlas como universales, pero a su vez son múltiples y cambiantes, de tal forma que la tecnología es entendida de diversas formas, encontrando observaciones sobre la definición de tecnología que van desde conceptos intelectualistas, utilitaristas e instrumentalistas, hasta llegar a ser presentada como un sinónimo de ciencia (Veraszto et al., 2009), teniendo en común que siempre se establece como una definición en construcción, por otra parte, el tesoro de la UNESCO presenta la tecnología como “Competencia, conocimiento, equipamiento y procedimientos para asegurar la producción de bienes y servicios”, permitiendo de esta forma la primera aproximación a la idea que la tecnología obedece a un sistema sociotécnico (Fuenfschilling & Truffer, 2014).

La tecnología es comúnmente observada desde las herramientas o los métodos según lo plantean Nawaaz et al. (1976), a lo cual es necesario aplicar un contexto para el producto o artefacto (Kaplan & Tripsas, 2008), donde la tecnología no es sólo el conocimiento con el cual se

elaboran los productos, sino que también incluye la manifestación física del conocimiento dentro de un producto.

En este sentido la tecnología no es una sola, ni es estática, por el contrario evoluciona constantemente y la frontera tecnológica, es la suma de los alcances de diversas tecnologías y se extiende como capacidad humana (Geels & Schot, 2007), pudiendo tomar forma de herramienta, técnica, proceso, equipo o método, siendo cada una diferente en capacidad y uso.

El concepto de tecnología ha cambiado con el paso del tiempo, presentando el mismo comportamiento de las propias tecnologías, con vertientes, usos técnicos y coloquiales (Taylor & Taylor, 2012). En el presente documento se toma como concepto de tecnología “la que puede entenderse, como el conjunto de conocimientos, experiencias y relaciones que sustentan el desarrollo, producción y distribución de productos y el desarrollo e implementación de procesos de transformación de materia e información” (Robledo, J, 2013).

2.2 Cambio Tecnológico

Volcando la mirada sobre la fuente original, los cambios tecnológicos planteados por Schumpeter, consisten en cualquiera de estos cinco fenómenos independientes, aunque estos pueden coexistir, requiriendo solo de uno para detonar el cambio: (1) introducción de un nuevo bien; (2) introducción de un nuevo método de producción; (3) apertura de un nuevo mercado; (4) conquista de una nueva fuente de suministro de materias primas o productos semi manufacturados; y (5) implementación de una nueva forma de organización (Schumpeter, 1935) (Schumpeter, 1939), siendo esta la base sobre la que se edifica la teoría de cambio tecnológico evolutivo.

Las siguientes características/atributos planteadas por Rogers (1983), permiten observar la innovación y su proceso de adopción, complementando los cinco fenómenos de la innovación de Schumpeter:

- Ventaja relativa - la innovación es técnicamente superior (en términos de costes, funcionalidad, imagen, etc.) a la tecnología que sustituye.
- Compatibilidad - la innovación es compatible con los valores existentes, habilidades y prácticas de trabajo de los posibles adoptantes.
- Complejidad - la innovación es relativamente fácil de entender y utilizar.
- Posibilidad de prueba (*trialability*) - la innovación se puede experimentar a modo de prueba sin un esfuerzo y gastos excesivos; se puede implementar de forma incremental y aun así proporcionar un beneficio neto positivo.
- Observabilidad - los resultados y beneficios de la utilización de la innovación pueden observarse fácilmente y se pueden comunicar a los demás.

El reconocimiento de Schumpeter como el autor seminal se evidencia en múltiples publicaciones, dentro de los aportes más relevantes se encuentran el de Godin (2008) quien incorpora y reconoce a W. Rupert Maclaurin como el autor que primero desarrolla los postulados de Schumpeter (Maclaurin, 1950) (Maclaurin, 1953) (Maclaurin, 1954). El autor argumenta que la innovación se compone de varias etapas y que los factores económicos son los responsables de la tasa de progreso tecnológico, trazando de esta forma el primer modelo lineal de innovación y cambio tecnológico neo-Schumpeteriano. Es necesario rescatar el aporte de Nelson y Winter (1982) quienes plantearon basados en Schumpeter, los conceptos de variación, selección y retención como homónimos de *Technological Life Cycle* (TLC) (Becker et al., 2012).

Para Schumpeter el tamaño de la firma y la estructura del mercado determinan la innovación, lo que Ahuja et al. (2008) adicionan indicando que son muchos más los componentes que afectan la innovación. Los mismos autores reconocen que existe innovación radical e incremental, por lo cual se requiere un modelo teórico más amplio que el planteado en 1935 por Schumpeter para entender su dinámica. En respuesta, los modelos neo-Schumpeterianos incluyen análisis de relaciones sociales, económicas y políticas, como una construcción del medio ambiente de innovación (Windrum & Garcia-Goni, 2008), al incluir en los procesos de innovación la evolución de la demanda de los usuarios.

La tecnología, vista como un proceso social donde los sistemas de innovación son los hilos conductores del cambio tecnológico, permite reconocer tres detonantes del cambio (Jamison & Hård, 2003):

- La innovación económica - centrado en la producción y los productos, siendo el mercado el juez natural del éxito de la tecnología. Este enfoque netamente comercial moviliza la tecnología en pro de los resultados financieros.
- La construcción social - centrado en los mediadores y las redes como un constante proceso de negociación y combinación de tecnologías en disputa por el uso predominante.
- La aceptación cultural - centrado en los usuarios y los sitios donde se usa la tecnología como la relación entre los medios y los fines.

En el mismo sentido, se plantea que la innovación y el cambio es transversal a todas las disciplinas y tecnologías (Block et al., 2017) y se deben incluir todas las contribuciones, en especial las sociales, por lo que deben evaluarse desde el conjunto de factores sociotécnicos que permiten a la tecnología desarrollarse, implementarse, mantenerse correctamente (de acuerdo con

las necesidades de todas las partes interesadas), atraer a los consumidores a largo plazo y crear una salida positiva o resultado de acuerdo con el propósito de la tecnología y las intenciones iniciales de sus desarrolladores, sean estas financieras, sociales, técnicas, etc. (Aizstrauta et al., 2015).

Desde la visión opuesta, la doble externalidad problemática de los cambios tecnológicos y la demanda del consumidor, sumada a la imprevisibilidad de la innovación y la existencia de múltiples equilibrios (Faber & Frenken, 2009), tiene como respuesta, los modelos evolutivos canónicos simplificados, que abarcan las relaciones no lineales fundamentales, como una interpretación de Schumpeter, en la cual la representación de los cambios tecnológicos como listas de eventos, requiere de una visualización más comprensible que encierre y simplifique el entendimiento del cambio y la innovación.

2.3 Evolución Tecnológica

Tomado de la misma forma que la biología y la teoría evolucionista de las especies planteada por Charles Darwin (1859), se entiende que la tecnología también evoluciona y con ella también evolucionan los artefactos como expresión material del conocimiento, los métodos y en general, cualquier manifestación de tecnología (Kaplan & Tripsas, 2008), por lo cual es necesario dejar claro que esta tesis versa sobre las tecnologías, no sobre los artefactos, aunque sean estos uno de sus resultados visibles.

En este sentido, la metáfora de la selección natural propuesta por Nelson y Winter (1982) va más allá de las características del entorno. Ellos proponen el cambio continuo -incremental- como un proceso de evolución tecnológica que requiere para su comprensión de la incorporación de la diversidad, la incertidumbre y la interrelación como una visión compleja pero realista del cambio tecnológico. En la misma línea el profesor Giovanni Dosi (1982), aporta desde la economía

evolutiva la relación de la innovación y los empresarios, con la dinámica del mercado y las señales que éste aporta para la definición de la ruta del cambio tecnológico. Sus documentos seminales son abordados por otros autores como fuente de soporte de sus planteamientos sobre evolución y cambio tecnológico (Adner & Kapoor, 2010; Bergek et al., 2008; Christensen & Bower, 1996; Geels, 2004; Pavitt, 1984a).

Son Freeman et al. (1982b) los principales exponentes de la relación entre el cambio tecnológico y el ciclo económico. Tomando como base la teoría de Schumpeter, Freeman establece desde la Universidad de Sussex un estudio de las conexiones entre los ciclos de onda larga, el desempleo y el desarrollo tecnológico, logrando un modelo de cambio tecnológico y la métrica asociada (Freeman, 1994) (Freeman, 2007).

Las teorías de evolución tecnológica toman fuerza a finales de la década del setenta y principios de la década del ochenta con principales exponentes como Dosi, Freeman, Pavitt, Nelson o Winter y continúan en la siguiente década con Geels, Verspagen y Pérez, todos con una misma base teórica en Schumpeter, aunque no siempre expresada abiertamente. Es Schumpeter quien da la línea base para desarrollos posteriores, como el planteamiento de paradigmas tecno-económicos de Carlota Pérez (2009) al definir un nuevo modelo de prácticas innovadoras normales (Pérez, 2012).

Aunque todos los planteamientos son válidos y aportan a la discusión, los conceptos de evolución tomados para este documento son los de Dosi, en avances científicos, factores económicos y variables institucionales, las cuales se agrupan en el concepto de sistema sociotécnico que se presenta adelante.

2.4 Trayectoria Tecnológica

El concepto de trayectoria tecnológica tiene sus orígenes en Pavitt (1982b) quien sostuvo que los cambios tecnológicos se pueden representar en una taxonomía de tres partes:

- Dominado por el proveedor - fuentes de tecnología
- Producción intensiva - requerimientos de los usuarios
- Basado en la ciencia - posibilidades de apropiación

Esta taxonomía evoluciona de tres partes en los documentos iniciales de Pavitt, a una taxonomía de cuatro partes (Pavitt, 1984b):

- Empresas dominadas por proveedores, aquellas empresas que adquieren sus conocimientos técnicos de sus proveedores.
- Proveedores especializados, principalmente en el campo de los equipos y bienes de capital, que proporcionan las innovaciones a las otras empresas.
- Firmas dominadas por escala, donde se asocia la innovación a escala.
- Empresas basadas en la ciencia, las que innovan a través de sus laboratorios propios de I+D

Dos años después de su primera publicación en el tema, el profesor Keith Pavitt usando datos de 2000 innovaciones significativas, presenta que el cambio tecnológico es en gran parte un proceso acumulativo, en el que las posibilidades de hacer cambios tecnológicos en el futuro están fuertemente condicionadas por lo que se ha podido hacer técnicamente en el pasado (Pavitt, 1984b).

Aunado a ello, la propuesta de Hummon y Dereian (1989) de un ordenamiento topológico para analizar la evolución de la comprensión científica del ADN, demuestra agrupamientos

evidentes en dendogramas basados en citas bibliográficas, sustentando que se puede realizar rastreo de la trayectoria tecnológica por métodos históricos, replicando los primeros hallazgos de Pavitt. A su vez, Verspagen (2007) realizó un ejercicio de análisis de trayectorias tecnológicas usando citas de patentes con un resultado confirmatorio para no linealidad en la trayectoria, pero con una visión macro que evidenció que su evolución sigue la estructura de ensanchamiento y contracción propuesta por Geels (Geels, 2005c, p. 472), corroborando el comportamiento sigmoideal presentado por el mismo autor y Kemps en 2007 (Geels & Kemp, 2007, p. 444).

En la vía del análisis de redes para citas y patentes que permiten determinar trayectorias tecnológicas, el documento de Mina et al. (2007), en el que aplicaron la herramienta a 11.240 artículos sobre la enfermedad arterial coronaria y a 5.136 patentes para dispositivos de angioplastia, demuestra que los mapas de trayectoria tecnológica se pueden observar mediante grafos y son consistentes con la realidad encontrada como resultado del proceso de acumulación de conocimientos.

Otras vertientes son las presentadas por Jamison y Hård, (2003), donde la trayectoria tecnológica es vista como un proceso social donde los sistemas de innovación son los hilos conductores del cambio tecnológico, reconociendo tres detonantes del cambio: la innovación económica, la construcción social y la aceptación cultural. La no linealidad y los cambios en la dirección de las trayectorias tecnológicas están relacionados con cambios en las reglas y expectativas cognitivas que guían las actividades de búsqueda y desarrollo técnico (Geels & Raven, 2006b). De tal forma que los cambios en la dirección de los desarrollos están relacionados con un cambio de las expectativas, especialmente sobre la funcionalidad de la tecnología. La adaptación tecnológica lleva a la diversificación, en la cual la tecnología existente es una

coincidencia entre el mercado y la oportunidad tecnológica, mediada por la carga histórica, así planteando la trayectoria tecnológica como un proceso causal (Kim & Kogut, 1996).

2.5 Curva Sigmoidal

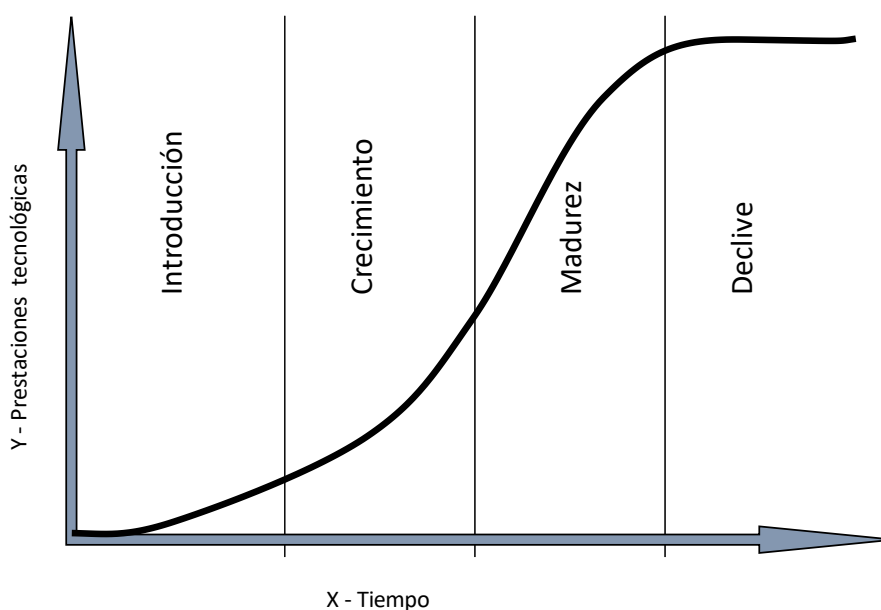
La diagramación del cambio evolutivo de la tecnología se logra durante las décadas del 1960 y 1970 con Everett Rogers y Frank Bass, (Rogers, 1976b) (Bass, 1969b) quienes presentan esta evolución como una curva sigmoidal, función logística, curva de Verhulst o también llamada curva en S, la que se puede observar en la Figura 2.1. Esta curva es una representación de dos dimensiones, en el eje de las abscisas (X) se ubica el tiempo y en el eje de las ordenadas (Y) una prestación tecnológica, siendo posible presentar en las ordenadas, ventas, usuarios, capacidades, prestaciones o cualquier otro atributo.

Este método de exposición de la evolución tecnológica también se puede usar para la evolución de productos y es desde este uso, que se importa la caracterización de cuatro etapas en la curva: introducción, crecimiento, madurez y declive, según lo planteo Theodore Levitt (1965b), la cual se toma como base en este documento, aunque se encuentran referencias anteriores (Rink & Swan, 1979), que manifiestan que en 1950 y 1951 Joel Dean lo planteo (Dean, 1950b), pero sin incluir una representación gráfica, solo una exposición de momentos de un producto en el mercado.

En general, una curva en S sirve para presentar la evolución en el tiempo de una variable, siendo usada para diversas explicaciones como el ciclo de vida o las prestaciones tecnológicas. Específicamente en el modelo de adopción tecnológica de Bass se usa en la identificación de cinco momentos de adopción: (1) Los innovadores; (2) adoptantes tempranos; (3) mayoría temprana; (4) mayoría tardía; y (5) rezagados (Bass, 2004). La conjunción de los modelos de Bass y de Levitt, constituye la base del uso de la curva sigmoidal para el análisis evolutivo. Taylor & Taylor (2012)

presentan las etapas de Levitt indicando que se adaptan al contexto específico en que se desarrolla la tecnología, dicha adaptación responde a la inclinación de la curva siempre manteniendo los cuatro momentos expuestos por Levitt (Nieto et al., 1998).

Figura 2.1 Etapas del ciclo de vida tecnológico



Nota. Curva sigmoide con cuatro etapas del ciclo de vida tecnológico. Adaptado de “Exploit the Product Life Cycle” por T. Levitt, 1965, Graduate School of Business Administration, Harvard University. Vol(43)

En detalle, la etapa de introducción representa la fase en la que el producto ha sido lanzado por primera vez al mercado, durante la cual los volúmenes de venta son bajos. Durante la fase de crecimiento, la aceptación del consumidor del producto aumenta, y con ella los volúmenes de ventas. En la madurez, el volumen de ventas se estabiliza antes de disminuir en la fase de declive, en la cual las ventas se reducen y a su vez, los consumidores (Taylor & Taylor, 2012). Esta caracterización del comportamiento según cada fase del ciclo es crucial en la definición de su conducta futura y se aplica en la literatura tanto para el ciclo de vida del producto como para el

ciclo de vida de la tecnología, puesto que tienen comportamientos equivalentes (Abernathy & Clark, 1985).

Otras denominaciones encontradas en la literatura para las etapas del ciclo, siendo todas ellas diferentes denominaciones de un mismo proceso, son:

- Pre comercialización – introducción – crecimiento – madurez – declive (Fox, 1973).
- Desarrollo de mercado – crecimiento rápido – turbulencia competitiva - saturación/madurez – declinación (Wasson, 1976).
- Introducción – crecimiento – madurez – disminución (C. R. Anderson & Zeithaml, 1984).
- Embrionario – crecimiento – sacudida - madurez – declive (Jones et al., 1998).

Desde el uso de Foster de la curva en S, para analizar la madurez tecnológica a nivel de componentes (Foster, 1985b), se inició un desarrollo continuado para incorporar las partes y materiales que interactúan dentro de un sistema de cambio tecnológico, presentando las mejoras incrementales como una evidencia de la existencia de componentes dentro de la tecnología (Christensen, 1992b). Estos elementos componentes o constitutivos producen diferentes resultados dependiendo de su orden y magnitud y por ende, la combinación o permutación de sus elementos genera múltiples posibilidades, dado que, en la medida que se incluyan más elementos en la ecuación, mayor es el abanico de posibilidades de salida, no queriendo decir con esto que todas las salidas sean exitosas para dar respuesta a un mismo problema, solamente llegando al nivel de definir que son distintas.

Los componentes se han abordado desde las patentes comunes que usan algunas tecnologías, creando grupos de productos que evolucionan simultáneamente (Klepper, 1997). La

unión entre las curvas planteadas por Foster (1985b), el modelo de Difusión de Rogers (1983) y el ciclo de vida tecnológico de Abernathy (1985), dan paso a la creación de un modelo compuesto que sirve para coordinar los esfuerzos en cambio tecnológico y de productos efectivos (Nieto et al., 1998). El ciclo de vida de la tecnología de Abernathy se plantea desde las ideas de Schumpeter (1935a), Kuhn (1962), Vernon (1966), Wells (1968), Nelson y Winter (1982) e incluye tempranamente el *Market Pull* y el *Technology Push* como elementos detonantes del cambio tecnológico.

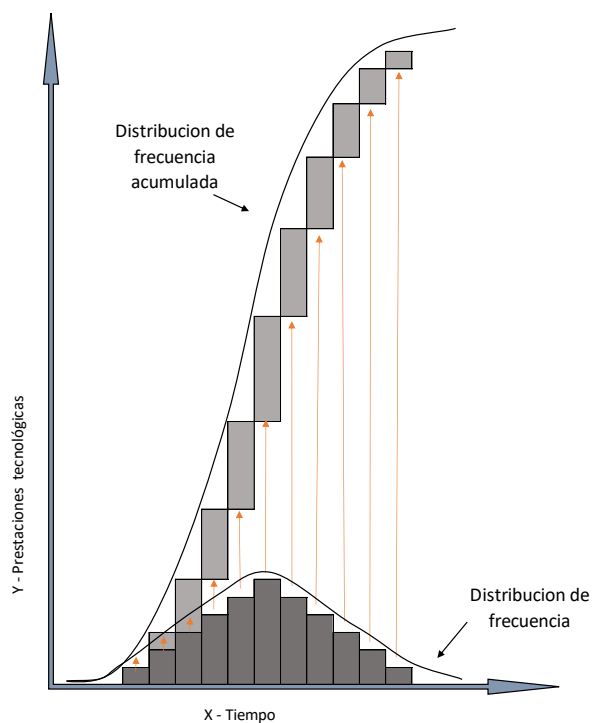
La comprensión de la evolución tecnológica desde el análisis de curvas sigmoidales avanza en 1992 con dos artículos (Christensen, 1992b) (Christensen, 1992c), en donde se analiza la evolución de los elementos componentes de la tecnología y sus resultados. En 2005 Sood y Tellis (2005a) aportan que la evolución tecnológica varía entre productos, viniendo de una tecnología antigua y creciendo hasta superar la tecnología actual, posibilitando la aparición de múltiples curvas en S. El termino progresión tecnológica se incluye en el debate mediante la diferencia entre producto, industria y tecnología, cada uno con su propio ciclo de vida que comparte etapas comunes que se entrecruzan (Aizstrauta et al., 2015).

Fujimoto (2007) presenta el modelo de elementos funcionales (prestaciones) y los elementos estructurales (componentes) que permiten determinar la evolución de un producto o tecnología, como una arquitectura integral de correspondencia entre elementos funcionales y estructurales del producto. Por su parte, la teoría neo-Schumpeteriana se usa para modelos emergentes de imitación o sustitución de componentes, encontrando que en la etapa de madurez se prefiere la especialización mientras en etapas tempranas la imitación o el reemplazo (Vermeulen et al., 2018), siguiendo el modelo de adoptadores de Bass (2004).

La forma en que la teoría evolucionista del cambio tecnológico se une con el uso de las curvas sigmoideas como representación bidimensional, logra que la interpretación de los ciclos de vida tecnológico y del producto, puedan ser analizados con mirada retrospectiva y también abre la puerta a la prospectiva, entendiéndola como un sistema, diferenciando los cambios incrementales y los diseños dominantes (Hekkert et al., 2007) (Huenteler et al., 2016b).

La presentación de la curva sigmoidea tiene una estrecha relación con la distribución normal, dado que la sigmoidea es una acumulación de las frecuencias de la otra, como se puede observar en la Figura 2.2, la curva normal, se traslada a una curva en S con solo acumular sus frecuencias.

Figura 2.2 Relación entre distribución normal y curva sigmoidea



Nota. Adaptado de “Manual de farmacología y terapéutica” por R. Hilal y L. Brunton, 2009, McGraw Hill.

2.6 Ciclos de Vida

El ciclo de vida del producto, *Product Life Cycle* (PLC) y el ciclo de vida de la tecnología, *Technology Life Cycle* (TLC), son conceptos diferentes, aunque ambos son usados para análisis macro de la evolución de la tecnología al incluir entre ellos una relación causal (P. Anderson & Tushman, 1990), estando relacionados con la curva S (Foster, 1985b), donde la progresión tecnológica se produce dentro y entre las entidades del TLC (Taylor & Taylor, 2012).

El concepto de ciclo de vida es traído de las ciencias naturales, al ubicarlo en el ámbito de la industria, el ciclo de vida se traslada al producto y a la vida de la tecnología. El TLC, es introducido por Arthur D. Little (1981) en el ámbito de la administración de tecnología pero con referencias en el área de producción y estrategia (Utterback & Abernathy, 1975) (Roussel, 1984), el TLC permite medir los cambios tecnológicos evidenciando su evolución en una curva en S con la limitante de observar a la tecnología como elemento separado, al no analizar las diferencias de las tecnologías en sí mismas (Aizstrauta et al., 2015).

Foster presenta una explicación del comportamiento de las curvas S definiéndolas como “historias gráficas de los esfuerzos humanos para resolver problemas” (Foster, 1985b). El ciclo de vida de una tecnología es un proceso gradual y se produce por cambios incrementales dentro de esta (J. Lee & Berente, 2013), los cambios incrementales tienen diferentes direcciones, pero solo uno de ellos termina siendo dominante y marcando el nuevo paso del cambio (Freeman et al., 1982b). Los cambios incrementales también pueden verse como ciclos de onda corta y la sumatoria de estas ondas cortas, permite la construcción de una onda larga (Korotayev et al., 2011).

El manual de Oslo incluye el ciclo de vida de los productos como una variable a tener en cuenta para reducir la incertidumbre de los progresos futuros de la tecnología (OCDE, 2018). En

esta vía se han usado las patentes para analizar el TLC (Haupt et al., 2007) con resultados positivos, aunque se encuentran objeciones (Järvenpää et al., 2011), al manifestar que este tipo de estudios obedece al modelo lineal de innovación y este no es necesariamente el comportamiento real, por lo cual proponen el uso de modelos alternativos para el análisis del TLC.

2.6.1 Ciclo de Vida de Onda Corta:

La difusión de sucesivos cambios en una tecnología y los productos que los reflejan, miden su avance en generaciones. Una nueva generación representa una mejora con respecto a las generaciones anteriores, impulsando a los compradores de generaciones anteriores a desplazar su compra hacia la última generación y el mercado se expanda con usos y aplicaciones crecientes a medida que la tecnología mejora (Bass, 2004). Las características del TLC de onda corta son cambios continuos o incrementales medidos en generaciones, que se basan en la generación anterior y agregan nuevas funcionalidades o características a la tecnología sin que esta se modifique profundamente.

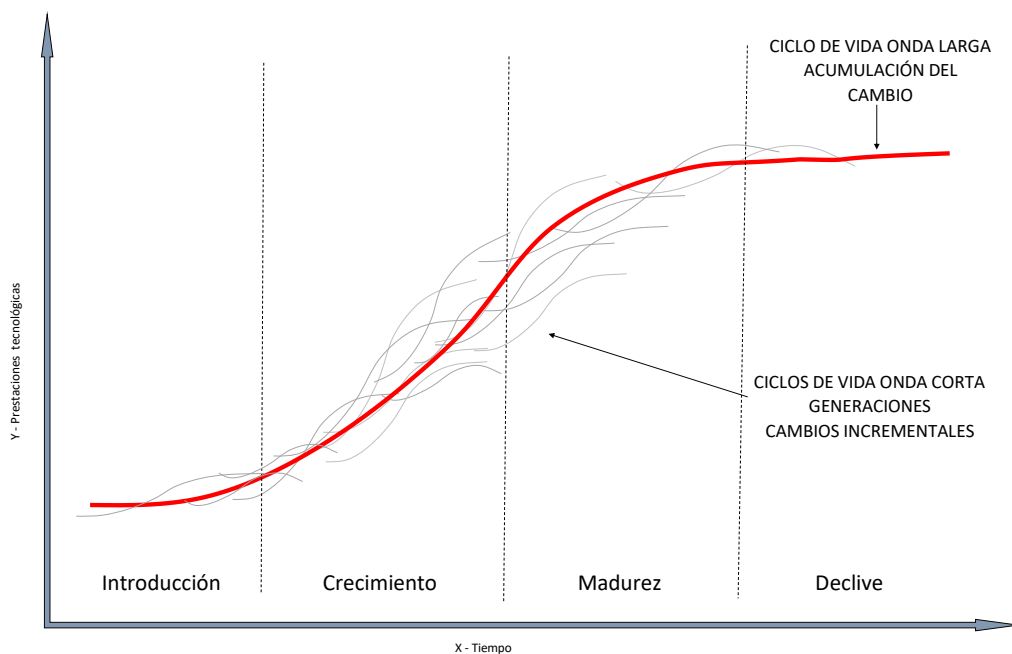
2.6.2 Ciclo de Vida de Onda Larga:

La sumatoria de los ciclos de onda corta (generaciones) dan como resultado un ciclo de mayor envergadura o ciclo de onda larga, que se caracteriza por la discontinuidad de la tecnología o finalización de las innovaciones incrementales para pasar a un cambio tecnológico disruptivo y la generación de un nuevo paradigma tecnológico (Bass, 2004). Este cambio radical de la tecnología no se basa en el paradigma anterior salvo en la finalidad central de la tecnología, pero no en su técnica.

La relación entre los ciclos de onda corta y cómo estos se trasportan a la onda larga, se puede observar en la Figura 2.3, en donde los ciclos de onda corta (generaciones) inician en la etapa de introducción con pocos cambios, que van aumentando en la etapa de crecimiento hasta la

madurez, donde nuevamente la cantidad de cambios incrementales disminuye, hasta el punto en que la tecnología inicia su declive, coincidiendo con el planteamiento Geels (Geels, 2005c, p. 450). Esta relación de cuatro etapas de los ciclos de onda corta evidenciados como generaciones, que se observan de manera agregada en las etapas del ciclo de vida de onda larga es la base para establecer predicciones de comportamiento futuro.

Figura 2.3 Ciclos de onda corta y onda larga para TLC.



Nota. Adaptado de “Exploring the limits of the technology s-curve. Part i: component technologies”, por C.M. Christensen, 1992, *Production and Operations Management*, 1. y “Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study”, por F.W. Geels, 2002, *Research Policy*, 31.

2.7 Sistema Sociotécnico

El término sistema sociotécnico fue acuñado originalmente (Emery & Trist, 1960) para describir sistemas que implican una interacción compleja entre humanos, máquinas y los aspectos

ambientales del sistema de trabajo (Baxter & Sommerville, 2011). En los primeros momentos la teoría de sistemas sociotécnicos hace poco hincapié en las relaciones humanas, después esta es incorporada como un sistema de valores de derechos y necesidades tan importante como el cambio técnico (Mumford, 2006).

Los sistemas sociotécnicos abiertos tienen cinco características (Badham et al., 2000):

- Los sistemas deben tener partes interdependientes.
- Los sistemas deben adaptarse y perseguir objetivos en entornos externos.
- Los sistemas tienen un entorno interno que comprende subsistemas técnicos y sociales separados pero interdependientes.
- Los sistemas tienen equifinalidad. Los objetivos de los sistemas se pueden lograr por más de un medio. Esto implica que hay opciones de diseño que deben realizarse durante el desarrollo del sistema.
- El rendimiento del sistema se basa en la optimización conjunta de los subsistemas técnico y social. Centrarse en uno de estos sistemas para excluir al otro, es probable que conduzca a un rendimiento y una utilidad del sistema degradados.

Los cambios en el sistema sociotécnico los presenta Geels (2005b) como las transiciones que involucran la evolución conjunta de la tecnología y la sociedad. Tomando como ejemplo el suministro de agua, muestra 5 combinaciones de co-evolución:

- Co-evolución entre tecnología y usuarios.
- Co-evolución entre tecnología, estructura industrial e instituciones políticas.
- Co-evolución de la ciencia, la tecnología y el mercado.
- Co-evolución de la ciencia y la tecnología.

- Co-evolución de la tecnología y la cultura.

Geels propone que en el nivel meso del sistema sociotécnico de cambio, se encuentran cuatro fases:

- En la primera fase, las novedades emergen en nichos en el contexto del régimen existente, todavía no hay un diseño dominante y coexisten varias formas técnicas.
- En la segunda fase, la novedad se utiliza en pequeños nichos de mercado, que proporcionan recursos para la especialización técnica.
- La tercera fase se caracteriza por un avance de la nueva tecnología.
- En la cuarta fase, la nueva tecnología sustituye al antiguo régimen.

En un segundo artículo sobre el tema Geels conceptualiza una vía de transición tecnológica, llamada "desalineación y realineación" reforzando la idea que las dinámicas no son mecánicas y lineales, sino que se producen a través de las interacciones de grupos sociales con diferentes intereses, estrategias y valores (Geels, 2005d).

Usando el ejemplo de las tecnologías de transporte, Geels presenta que el espacio para experimentar con múltiples innovaciones permite no solo la competencia entre ellas, sino que también produce interacciones positivas, tanto de forma paralela como secuencial (Geels & Schot, 2007). Las interacciones paralelas se refieren a retroalimentaciones positivas entre tecnologías coexistentes. Las interacciones secuenciales se refieren a las relaciones entre tecnologías que se suceden a lo largo del tiempo, en particular cuando una tecnología se basa en cambios que fueron activados por una tecnología anterior. Aunque el planteamiento gráfico de esta teoría sigue un patrón de ensanchamiento y estrechamiento de tecnologías existentes, continúa presentándose como una curva en S.

Dos años después de la primera publicación de Geels, este mismo presenta mejoras a su teoría (Geels & Kemp, 2007) identificando tres tipos de procesos de cambio: reproducción, transformación y transición.

- Reproducción, se refiere al cambio incremental a lo largo de las trayectorias existentes.
- Transformación, se refiere a un cambio en la dirección de las trayectorias, relacionada con un cambio en las reglas que guían la acción innovadora.
- Transición, se refiere a un cambio discontinuo o una nueva trayectoria y sistema.

También plantea que el sistema es dinámicamente estable, lo que implica que la innovación es principalmente incremental.

Una visión similar presentan Kemp et al, en la cual la gestión de la transición ayuda a las sociedades a transformarse de manera gradual y reflexiva por medio de procesos guiados de variación y selección, cuyos resultados son escalones para un cambio adicional (Kemp et al., 2007), también entendido como cambio incremental que se puede gestionar - incrementalismo dirigido.

Un aporte sobre los límites de los sistemas sociotécnicos es expuesto por Hekkert y Negro, centrándose en los factores endógenos del cambio, no en los exógenos que difícilmente se pueden predecir (Hekkert & Negro, 2009), para estos factores endógenos los autores asignan siete funciones.

- Función 1: Actividades empresariales.
- Función 2: Desarrollo del conocimiento (aprendizaje) I+D.
- Función 3: Difusión del conocimiento a través de las redes.
- Función 4: Orientación de la búsqueda, necesidades de los usuarios.

- Función 5: Formación del mercado, es importante crear espacios protegidos para las nuevas tecnologías.
- Función 6: Movilización de recursos financieros, humanos y técnicos.
- Función 7: Creación de legitimidad / contrarrestar la resistencia al cambio.

Un cambio incremental en la teoría de sistemas sociotécnicos es una segunda generación de la *Multi Level Perspective* (MLP), que incorpora explícitamente la escala espacial propuesta por Ravena et al. (2012b), que plantean como el uso de esta escala permite identificar relaciones que actúan a través del espacio y el tiempo. Esta postura es apoyada (Murphy, 2015), dado el papel que los factores espaciales y geográficos desempeñan en la evolución de la innovación sociotécnica, como determinantes de su camino. Este planteamiento de nivel espacial no será tenido en cuenta para el modelo propuesto en esta investigación, dado que su inclusión sobrepasa los objetivos del trabajo, aunque se reconoce su interesante aporte para el análisis de los datos.

Impulsando el enfoque de las funciones del sistema sociotécnico de innovación, se incluyen elementos de destrucción creativa como un aporte del sistema, para motivar la innovación por medio de la desestabilización de los elementos anteriores (Kivimaa & Kern, 2016). Una variación encontrada es el concepto de régimen sociotécnico presentado en 2014 (Fuenfschilling & Truffer, 2014) como un núcleo paradigmático de un sector, que resulta de la coevolución de instituciones y tecnologías a lo largo del tiempo. Especialmente incluye la variable institucionalización como fundamental en los efectos sobre los actores y el sistema, puesto que las tecnologías emergentes a menudo dependen de estructuras débilmente institucionalizadas y poco alineadas en un campo organizativo.

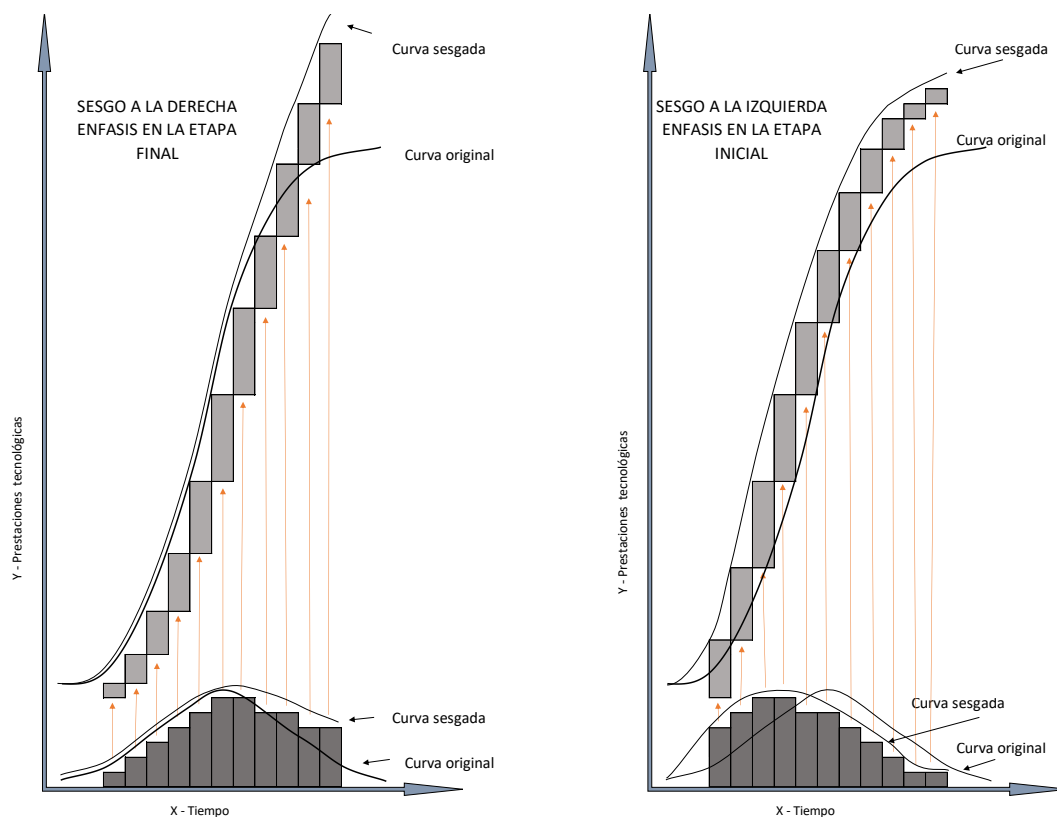
3 Identificación de Etapas en el Ciclo De Vida Tecnológico

El análisis y predicción de la evolución de las tecnologías incorpora desde su misma concepción la existencia de datos estructurados según el período de tiempo de su aparición, como se presentó en el capítulo anterior la curva sigmoideal se obtiene por la acumulación de la frecuencia absoluta de una variable, siendo necesarias las dos (absoluta y acumulada) para la representación numérica y gráfica que denote la estructura de cambio de pendiente y forma sigmoideal necesaria para la identificación de los puntos críticos que dividen las etapas del ciclo de vida de una tecnología.

Se puede evidenciar que el proceso de acumulación de frecuencia de las curvas sigmoideales tiene tres características: 1) Se puede acumular pero no des acumular, 2) La coordenada X puede ser ajustada a diferentes magnitudes (mes, año, cuatrienio), mientras que la coordenada de Y puede ser de uno entre varios factores, como: patentes, prestaciones, publicaciones, notoriedad, usuarios, valor de mercado, empresas en el sector, entre otras, y 3) Es posible aplicar medidas estadísticas que permitan la identificación de puntos críticos y la predicción de su comportamiento.

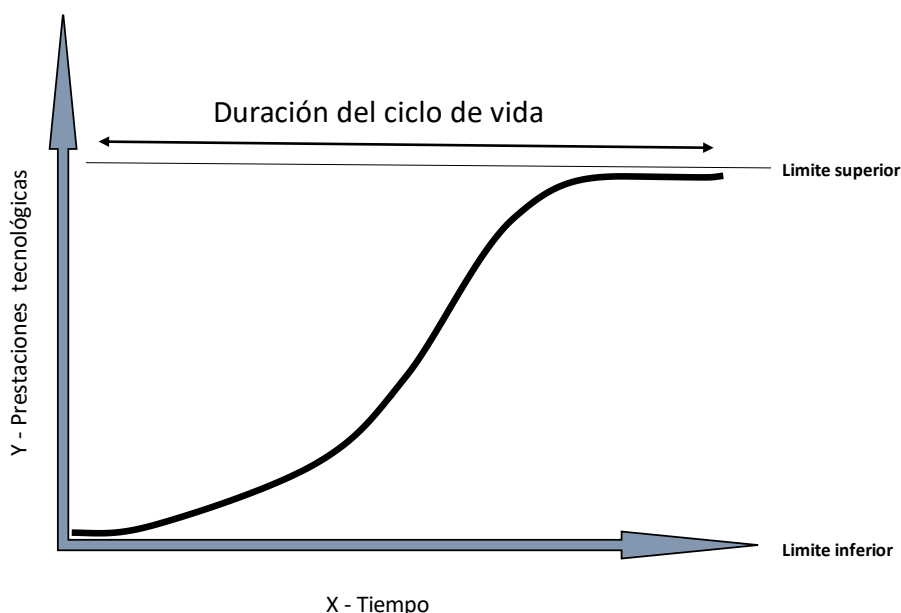
La relación entre frecuencia absoluta y acumulada, ilustradas como una curva normal y una curva sigmoideal, permite observar que, datos con sesgo a la derecha representarían mayor inclinación en una curva sigmoideal en la etapa final del ciclo de vida y una con sesgo a la izquierda, mostraría un sigmoide con mayor peso en la etapa inicial, denotando un rápido crecimiento, afectando todo el ciclo de vida, como se puede observar en la Figura 3.1, evidenciando la relación directa entre estas y su manifestación en las etapas del ciclo de vida.

Figura 3.1 Efectos de sesgos en las curvas de frecuencia absoluta y acumulada



Es necesario resaltar que las curvas del ciclo de vida de la tecnología están limitadas por dos restricciones, como se presenta en la Figura 3.2, la primera es la restricción de inicio, entendida cómo la aparición de la tecnología (introducción) y la segunda restricción es el alcance del ciclo de vida, entendido como el valor máximo obtenido por la frecuencia acumulada de esa tecnología a lo largo del tiempo (declive), el periodo de tiempo que se tarda en pasar de un punto al otro, contiene las cuatro etapas definidas previamente (introducción, crecimiento, madurez y declive) y dependiendo de la extensión de este, se pueden considerar ciclos de vida rápidos, medios o lentos, sin que se exista una escala específica para catalogarlos, caracterizándose por el contraste entre diferentes tecnologías para determinar si una tecnología es más rápida o lenta que otra.

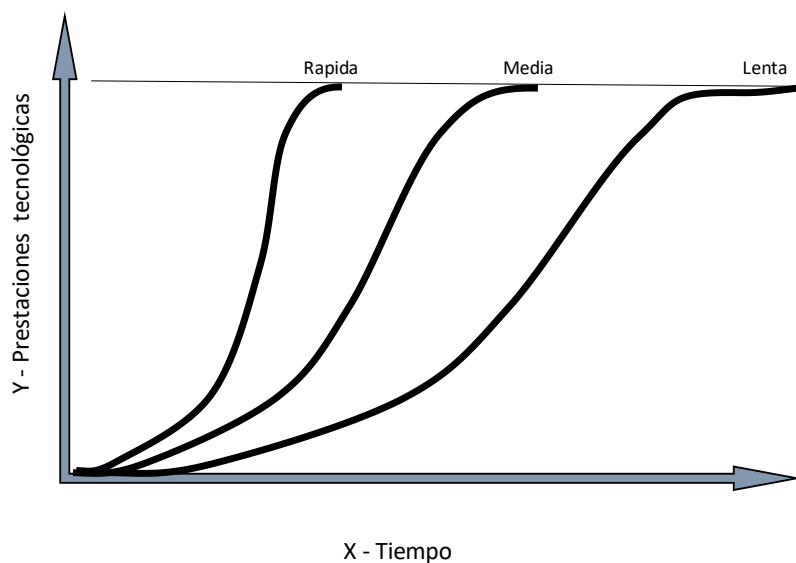
Figura 3.2 Restricciones en el ciclo de vida de la tecnología



Dentro del espacio creado por las restricciones del ciclo de vida tanto desde su inicio hasta su valor en el límite superior, se puede presentar diferentes formas de la curva, representadas a modo de ciclo rápido, de ciclo medio o de ciclo lento, como se observa en la Figura 3.3. La diferencia entre las curvas radica principalmente en el tiempo necesario para alcanzar cada una de las etapas en su ciclo de vida.

Debe destacarse que la altura máxima alcanzada, así como la duración de las curvas de ciclo de vida difieren entre tecnologías, esto representa la cantidad de cambios incrementales acumulados de la tecnología, de esta manera se entiende que cada tecnología tiene su propia dinámica y condiciones, pero respondiendo a las mismas etapas y forma de la curva. Las curvas cortadas abruptamente sin cumplir las cuatro fases obedecen a incorporación de tecnologías disruptivas que las reemplazan y eliminan del panorama tecnológico, o simplemente por ser tecnologías en proceso de evolución activa.

Figura 3.3 Diferentes curvas de ciclo de vida dependiendo de su duración

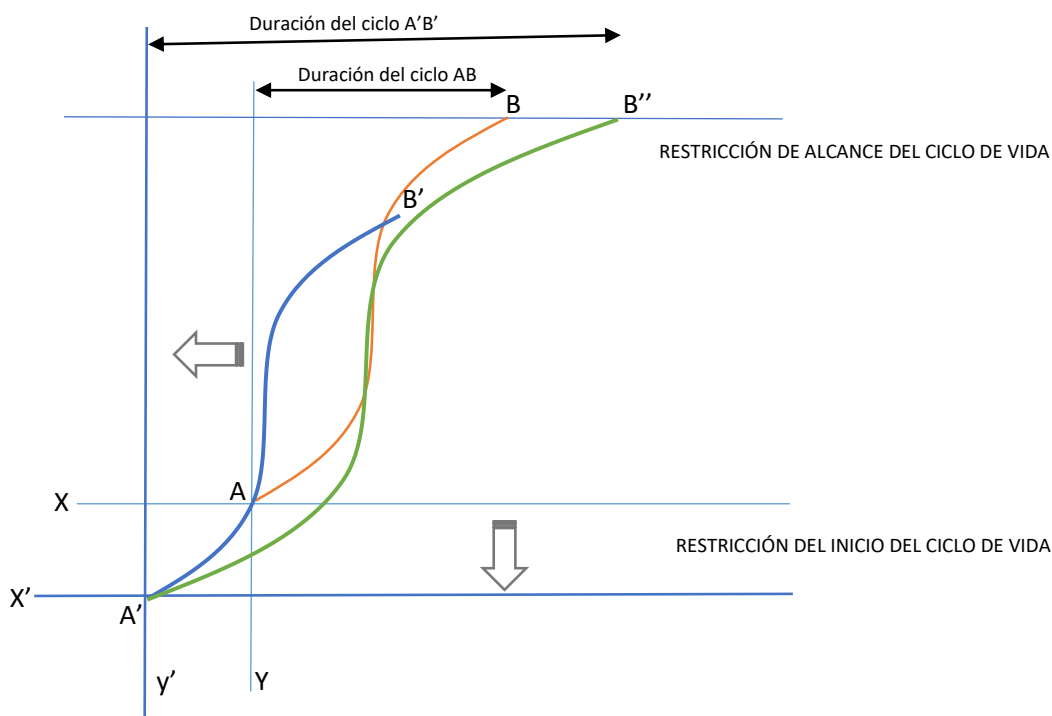


Es posible encontrar distorsiones en la forma de la curva del ciclo de vida dependiendo de la exactitud en la determinación del momento de introducción de la tecnología, siendo este fundamental en el trazado de su propia trayectoria, en el entendido que la evolución tecnológica siempre parte de un hecho anterior (Sood & Tellis, 2005b), esta identificación del momento de emergencia tecnológica debe ser tenido en cuenta con la mayor rigurosidad, puesto que un error en esta determinación acarrea un movimiento generalizado de todo el ciclo de vida, o una visión parcial de su comportamiento.

En la Figura 3.4, se presenta un caso de distorsión, en el cual se establece un eje de coordenadas XY y una curva de ciclo de vida que va del punto A al punto B (la forma de la curva es de ciclo rápido para exponer el comportamiento), al introducir una distorsión en punto de inicio de la tecnología causado por una identificación errónea del momento de introducción, pasando del punto A al punto A', y por tanto desplazando el punto B al punto B', el cual se encuentra bajo de la línea de restricción de alcance del ciclo de vida, posibilitando una extensión temporal de la curva para alcanzar el valor tope de la restricción superior.

Esta distorsión obliga a una reconfiguración de las coordenadas pasando a un eje de coordenadas $X'Y'$, tomando como punto de inicio A' para trazar una nueva curva de ciclo de vida presentando una trayectoria distinta hasta llegar a su restricción superior en el punto B'' . Se pueden observar las repercusiones de la identificación acertada del momento de introducción de una tecnología, tanto para la determinación de su duración (edad), como para las etapas de su propio ciclo de vida y el tiempo necesario para alcanzar su restricción superior, de tal forma que una curva rápida con una definición errónea del punto de introducción puede convertirse en realidad en una curva de duración media.

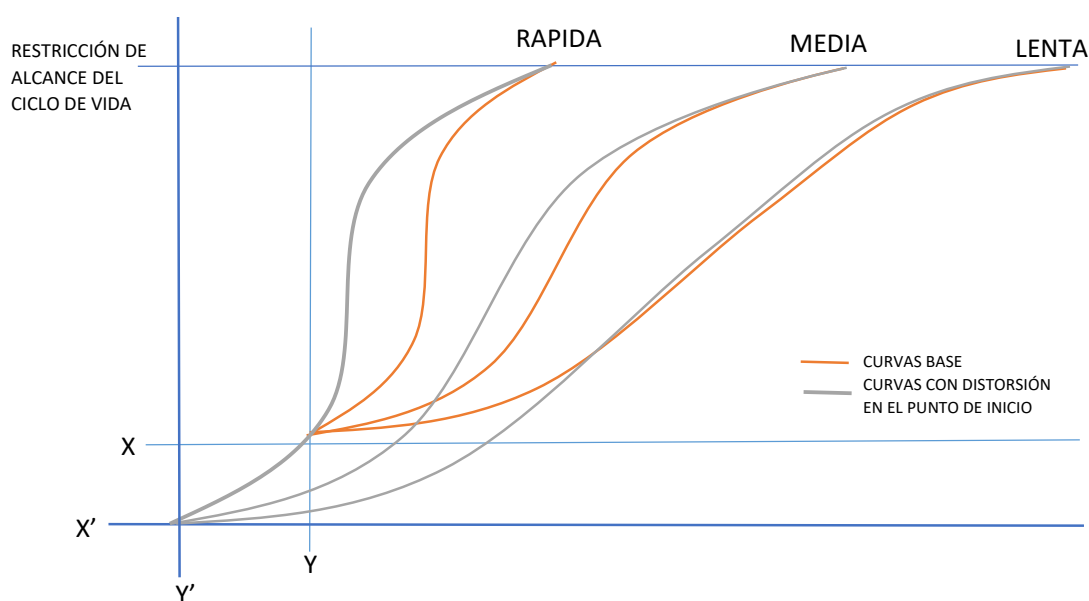
Figura 3.4 Distorsión en la curva de ciclo de vida con un punto de introducción erróneo



Dependiendo de la velocidad de la curva de ciclo de vida, los efectos de una identificación errónea del punto emergencia de una tecnología, tiene repercusiones de mayor importancia entre más corta sea la duración de su ciclo de vida, de esta forma en ciclos rápidos el cambio de trayectoria afecta las cuatro etapas del ciclo (introducción, crecimiento, madurez y declive),

evidenciado en la ausencia de cruces entre las dos curvas, en ciclos de vida medios, el cambio se presenta principalmente en las tres primeras etapas (introducción, crecimiento y madurez), en la última etapa las curvas se comportan similarmente llegando a cruzarse y en curvas lentas el cambio se produce sólo en las dos primeras etapas (introducción y crecimiento), luego las curvas se interceptan, y tienen comportamientos similares como se puede observar en la Figura 3.5.

Figura 3.5 Efectos de una distorsión en la emergencia de curvas de ciclo de vida



Para el caso de ejemplo expuesto, se ha mantenido el alcance de la curva de ciclo de vida de la tecnología en su restricción superior (declive), siendo posible que se presente una retracción de la curva como se presentó en la Figura 3.4 en la curva A'B', evidenciando que un ciclo de vida completo en sus cuatro etapas, puede pasar a ser un ciclo de vida incompleto (en evolución) con solo dos o tres etapas culminadas, por lo cual su trayectoria tecnológica aún no está completamente definida.

3.1 Etapas del ciclo de vida tecnológico

Según se plantea desde el modelo originario de Levitt (Levitt, 1965b) las 4 etapas del ciclo de vida tecnológico, así como los aportes de Nelson y Winter (Nelson & Winter, 1982), Dosi (Dosi, 1982), Pavitt (Pavitt, 1984b), Nieto et al (Nieto et al., 1998) y Pérez (Perez, 2010), tienen en común la segmentación de su comportamiento en cuatro momentos: introducción, crecimiento, madurez y declive, por lo cual se toman como base en este documento.

El límite inferior de la curva es por defecto cero (0) y se activa con la emergencia del primer registro de la tecnología, dando apertura a la etapa de introducción, desde allí es necesario establecer los tres puntos de corte siguientes en la curva, estos puntos de corte dividen las etapas así:

- a) Paso de introducción a crecimiento, la curva es cóncava, aumenta su velocidad de cambio.
- b) Paso de crecimiento a madurez (punto de inflexión), la curva pasa de cóncava a convexa.
- c) Paso de madurez a declive, la curva es convexa, disminuye su velocidad de cambio.

La determinación de los puntos de corte puede ser abordada desde diferentes perspectivas, las cuales se presentan a detalle a continuación, buscando con esta exploración metodológica la mejor configuración para la identificación, cumpliendo con los criterios de estabilidad, replicabilidad y exactitud.

3.2 Creación de límites:

Se establecen cinco límites para la curva de ciclo de vida así:

Límite inferior: es equivalente a la emergencia de la tecnología, medido como la aparición de la primera patente de la tecnología estudiada, este límite está compuesto por dos datos, el valor de la frecuencia absoluta de la emergencia, entendido como el número de patentes presentadas, y el segundo la fecha de la patente, esta información permite la graficación en dos dimensiones de los datos, siendo el inicio de la etapa de introducción.

Limite A – punto crítico a): es el punto en el cual la curva cambia la pendiente, pasando de la etapa de introducción a la de crecimiento, en la etapa de introducción se generan pocos cambios incrementales, los que se van tomando velocidad (frecuencia año a año), en esta fase se pueden presentar aceleraciones y retracciones continuadas, este proceso de estabilización tecnológica puede generar tasas de cambio de la frecuencia acumulada entre valores negativos, que luego se recuperan a positivo, cuando la tendencia de frecuencia se mantiene en positivo, se posibilita la aparición del punto crítico a).

Limite B – Punto crítico b): es el punto medio del ciclo de vida, también llamado punto de inflexión, y tiene la particularidad que en este punto la curva sigmoideal pasa de ser cóncava a convexa, denotando el cambio de la etapa de crecimiento a la de maduración, en la etapa de crecimiento se encuentra un rápido aumento de la acumulada y es cuando la tecnología se desarrolla para hacerse paradigma dominante, en esta fase no se encuentran retracciones mayores en la frecuencia absoluta. En la etapa de madurez se observan el crecimiento de la curva sigmoideal, donde la tecnología consolida su dominancia sobre otras, en la fase final de la etapa se presenta una desaceleración que da paso al siguiente punto crítico.

Limite C – punto crítico c): es el límite de crecimiento de la etapa de madurez y el paso a la etapa de declive, este punto se caracteriza por la desaceleración continuada de la tasa de cambio de la frecuencia acumulada, ralentizando sus cambios incrementales y cediendo su

posición de paradigma dominante a una tecnología sustituta, por ello se desplazan los esfuerzos de cambio tecnológico a otra, es posible que en la fase de declive también se encuentre que, el ámbito de uso de la tecnología varió y por ello su aplicabilidad ha sido desplazada, o que la tecnología ha encontrado restricciones de crecimiento con los recursos existentes en el momento (limitación intrínseca).

Límite superior: ubicado en la fase de declive, no se genera des acumulación de cambios incrementales, solo que estos son menos que antes y/o se encuentran espaciados temporalmente, en esta etapa se consolida el cambio hacía el nuevo paradigma tecnológico, aunque por ello la tecnología no deja de existir y continua siendo parte del corpus de conocimiento general, en el sentido expresado por Bass (1969b), son los adoptantes tardíos de la nueva tecnología imperante, los que mantienen el foco en la tecnología que se encuentra en fase de declive y se van alejando de ella lentamente.

En resumen, se presentan los 2 limites extremos y los 3 limites internos

Límite inferior: emergencia tecnológica

Limite a): paso de etapa de introducción a crecimiento

Limite b): paso de etapa de crecimiento a madurez

Limite c): paso de etapa de madurez a declive

Límite superior: sustitución tecnológica

3.3 Método 1 de identificación – Distribución Normal

Según se ha enunciado, la acumulación de las frecuencias absolutas de una curva normal tiene como resultado una curva sigmoïdal, razón por la cual los puntos de inflexión y puntos

críticos de una curva normal pueden ser proyectados a la curva sigmoideal, de tal manera que las características de una curva pueden ser observados en la otra.

De esta forma, tomando como base la distribución normal, se enuncian sus características:

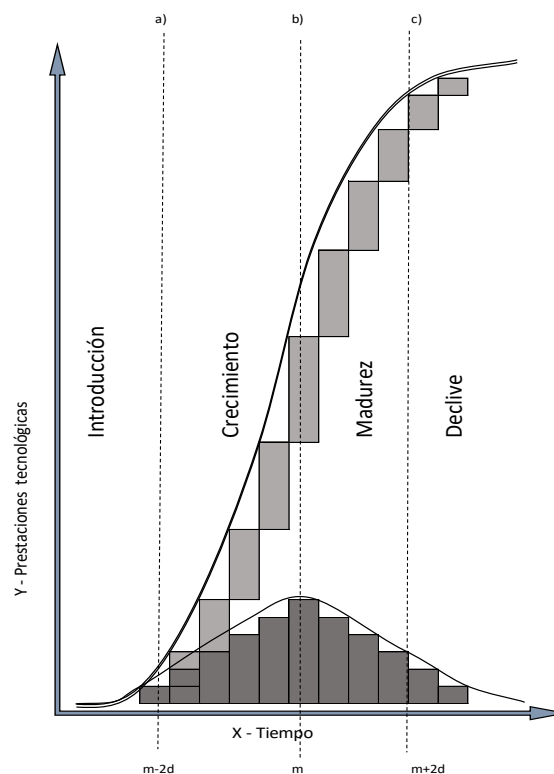
- Los valores de las mediciones tienden a agruparse alrededor de un punto central, la media
- La representación de los datos es simétrica a ambos lados de la media
- Las desviaciones estándares quedan situadas a igual distancia unas de otras
- La proporción de mediciones situada entre la media y las desviaciones es una constante en la que:
 - La media \pm 1 desviación estándar = cubre el 68,3% de los datos
 - La media \pm 2 desviaciones estándar = cubre el 95,5% de los datos
 - La media \pm 3 desviaciones estándar = cubre el 99,7% de los datos

Si este método de identificación de etapas es correcto, el punto de inflexión de la curva sigmoide es equivalente a la media de la distribución normal, y en la segunda desviación normal tanto a izquierda como derecha se encontrarán los puntos de corte a) y c) como se muestra en la Figura 3.6.

Estos puntos de corte segmentan los datos en cuatro secciones, en la primera se encontrarán el 2,28% de los datos según la tabla de valores bajo la curva de la distribución normal (Martínez, 1998, p. 175), que corresponde a la fase de introducción, en la siguiente fase de crecimiento se encuentran el 47,72% de los datos, hasta llegar al punto de inflexión que la separa de la siguiente fase de madurez, donde también se encuentran el 47,72% de los datos, la cual finaliza con la

segmentación de la segunda desviación a la derecha, para encontrar la fase de declive con el 2.28% de los datos.

Figura 3.6 Proyección de curva normal a sigmoide, con puntos críticos



Esta configuración aunque exacta en sus valores y sustentada por cálculos estadísticos que la hacen deseable, se aleja de la realidad de los datos expresados, puesto que el comportamiento de estos aunque se asemeja a una distribución normal, no se ajusta perfectamente a ella, dado que se presentan variaciones en la estructura de los datos, generando ordenamientos similares a la normalidad, con diferentes cúspides y agrupaciones de datos, sesgos a derecha o izquierda, dependiendo de la propia dinámica de la evolución tecnológica.

Este método, aunque prometedor, requiere que el ciclo de vida de la tecnología este completo para poder determinar todas sus etapas con exactitud, generando un impedimento en tecnologías cuya evolución aun no llega a la fase de declive.

El uso de la distribución normal como método de identificación de la etapa en que se encuentra la tecnología dentro de su propio ciclo de vida hace que los datos se ajusten al modelo y no que le modelo se ajuste a los datos, incumpliendo con los criterios de estabilidad, replicabilidad, exactitud, por lo cual se descarta el uso de la distribución normal para identificar las etapas del ciclo de vida de la tecnología.

3.4 Método 2 de identificación – Derivadas

El uso de las derivadas, como herramienta para la identificación de los puntos críticos dentro de la curva de ciclo de vida de una tecnología, es evaluada en su efectividad (Parvin et al., 2022), para ello se toma la monotonía de la sigmoide generada por la acumulada de la frecuencia, como base de cálculo de la primera derivada y sus curvas de tendencia lineal y polinomial.

La derivada se define como la pendiente de la recta tangente al punto donde se ubica x , la fórmula matemática se presenta en la Ecuación 1 a continuación.

Ecuación 1. Derivada.

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

Siendo, para este caso:

f : función

$f'(x)$: la pendiente de la recta tangente a la gráfica de f en un punto

El proceso de identificación inicia con el cálculo de la derivada usando la ecuación 1, tomando el segundo dato de la frecuencia acumulada y restándole el primer dato, este valor es dividido entre la resta obtenida del segundo año del periodo, menos el primero y se realiza para todos los datos del *set* de frecuencia acumulada, el resultado es un nuevo grupo de datos que

contienen los valores de la derivada en cada punto de x (año). Usando el programa Microsoft Excel, se grafica el *set* de datos recientemente calculado y se agregan dentro del área de la gráfica, la recta de tendencia lineal y la curva de tendencia polinomial de orden dos (2), con sus respectivas ecuaciones, las que se calculan sobre los datos base y se agregan a la matriz de datos.

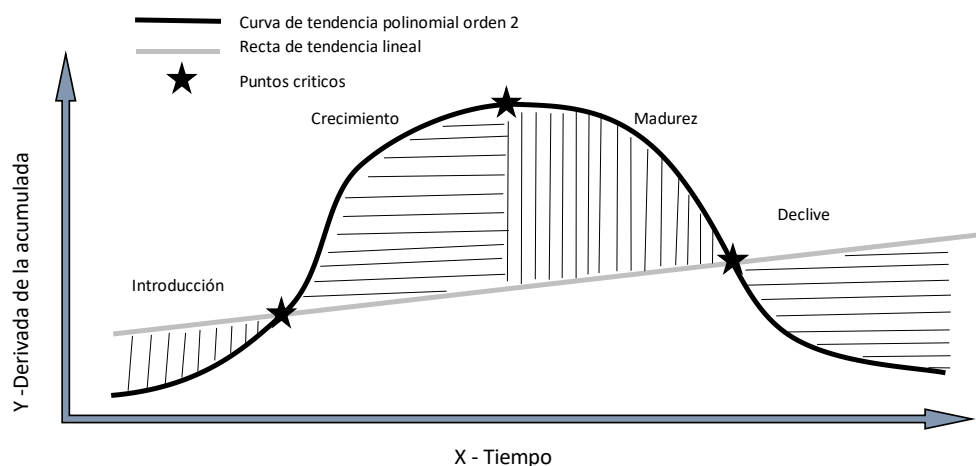
La identificación del punto crítico a), se obtiene en el primer punto de cruce de la recta de tendencia lineal, con la curva de polinomio orden dos (2), en el cuadrante positivo, estableciendo el momento en el cual la tecnología pasa de la etapa de introducción a la de crecimiento.

La identificación del punto crítico b) es el punto más alto (mayor valor) de la línea de tendencia polinomial de orden dos (2), este es el punto de inflexión, donde la forma de la curva sigmoideal pasa de cóncava a convexa, estableciendo el momento en el cual la tecnología cambia de etapa de crecimiento a etapa de madurez.

La identificación del punto crítico c) se obtiene en el segundo punto de corte de la recta de tendencia lineal, con la curva polinomial de orden dos (2), estableciendo el momento en el cual la tecnología pasa de etapa de madurez a etapa de declive.

Este proceso de identificación gráfico y numérico permite la determinación exacta de los puntos críticos que dan paso a las etapas del ciclo de vida de una tecnología, según se presenta en la Figura 3.7, las líneas de tendencia de la primera derivada ofrecen una división clara de las cuatro etapas y permiten ser proyectadas a la curva sigmoideal, usando como identificar el valor de X (año de ocurrencia del punto crítico).

Figura 3.7 Modelo de identificación de etapa en el ciclo de vida



Dado que este método matemático es viable para cualquier estructura de datos de una tecnología y es aplicable a ciclos rápidos, medios o lentos (Li et al., 2021), así como para tecnologías en evolución, se procede con su validación con pruebas sobre datos reales de tecnologías con ciclo de vida completo.

3.5 Prueba de la metodología de identificación de la etapa

Sobre la metodología de derivadas, se procede a realizar pruebas de funcionamiento usando datos reales, para lo cual se seleccionaron cinco tecnologías que han cumplido sus cuatro etapas del ciclo de vida o que se presume se encuentran en la etapa de declive, lo que permite una observación completa del ciclo y por ende la posibilidad de calcular los límites internos de la curva.

Las tecnologías de prueba seleccionadas son:

1. Buscador de personas – *Beeper - Pager*
2. *Video Home System - VHS*
3. Facsímil – *FAX*.
4. Máquina de escribir - *Typewriter*
5. Disco de almacenamiento flexible - *Disquete – Floppy disk*

Los datos son obtenidos de la página oficial de la Organización Mundial de la Propiedad Intelectual OMPI, en su aplicación PATENTSCOPE, el horizonte de tiempo de los datos de las tecnologías no tiene establecido un límite inicial, según se indicó líneas arriba en lo referente a la importancia de la exactitud de la emergencia tecnológica, el corte final de datos se ubica en el año 2020. Referente al origen de las patentes, no se asignan restricciones de lugar y se realiza sobre el reporte mundial, para ello se aplica ecuación de búsqueda de palabras clave, con lo cual se obtienen las bases de datos necesarias en la prueba.

El resumen de los datos obtenidos para estas tecnologías es presentado en la Tabla 1, donde se especifica la duración de su ciclo de vida (periodo) y el número de patentes en el conjunto de datos de cada tecnología.

Tabla 1 Tecnologías de prueba para identificación de etapa del ciclo de vida

Tecnología	Fecha inicio	Fecha fin	Período (años)	Patentes
Pager	1957	2020	63	6.022
VHS	1974	2020	46	10.000
Fax	1960	2020	60	9.921
Typewriter	1829	2020	191	11.026
Floppy	1966	2020	54	6.039

Para cada una de las tecnologías se aplicó la metodología de cálculo de derivada y líneas de tendencia lineal y polinomial, presentando para cada una de ellas tres figuras, la primera incluye las curvas de los datos de la frecuencia absoluta de las patentes y su frecuencia acumulada, la segunda es el resultado gráfico del cálculo de la derivada con sus curvas de tendencia lineal y polinómica, donde se identifican los puntos críticos a), b) y c) y la tercera es la ubicación de los puntos hallados en la curva sigmoideal, construyendo las cuatro etapas del ciclo de vida de la tecnología.

3.5.1 Frecuencias absolutas y acumuladas

Para la obtención de datos se usaron las siguientes ecuaciones de búsqueda en Patentscope, FP:(PAGER), FP:(VHS), FP:(FAX), FP:(TYPEWRITER AND DP: [1906 TO 1980]), FP:(TYPEWRITER AND DP: [1981 TO 2020]), FP:(FLOPPY), de las cuales se obtienen la serie de datos a usar, en las Figura 3.8, Figura 3.9, Figura 3.10, Figura 3.11 y Figura 3.12, se presentan las frecuencias absolutas, trazadas en color azul y las frecuencias acumuladas trazadas en color naranja.

Se puede observar que las líneas de frecuencia absoluta se asemejan a comportamientos normales, dentro de los cuales se evidencian sesgos, como el presentado por la tecnología de Floppy a la izquierda y comportamientos diferenciados en extensión y forma, como la curva platicúrtica de la tecnología de VHS, o mesocúrtica como en la curva de la tecnología de *Pager*.

El resultado de la acumulación de frecuencias que construye las curvas sigmoidales de cada tecnología corrobora la existencia de múltiples tipos de velocidad de ciclo de vida, así como las diferentes inclinaciones que éstas presentan en su recorrido, como un reflejo de la dinámica de su propia evolución.

Figura 3.8 Frecuencia absoluta y acumulada de la tecnología de prueba *Pager*

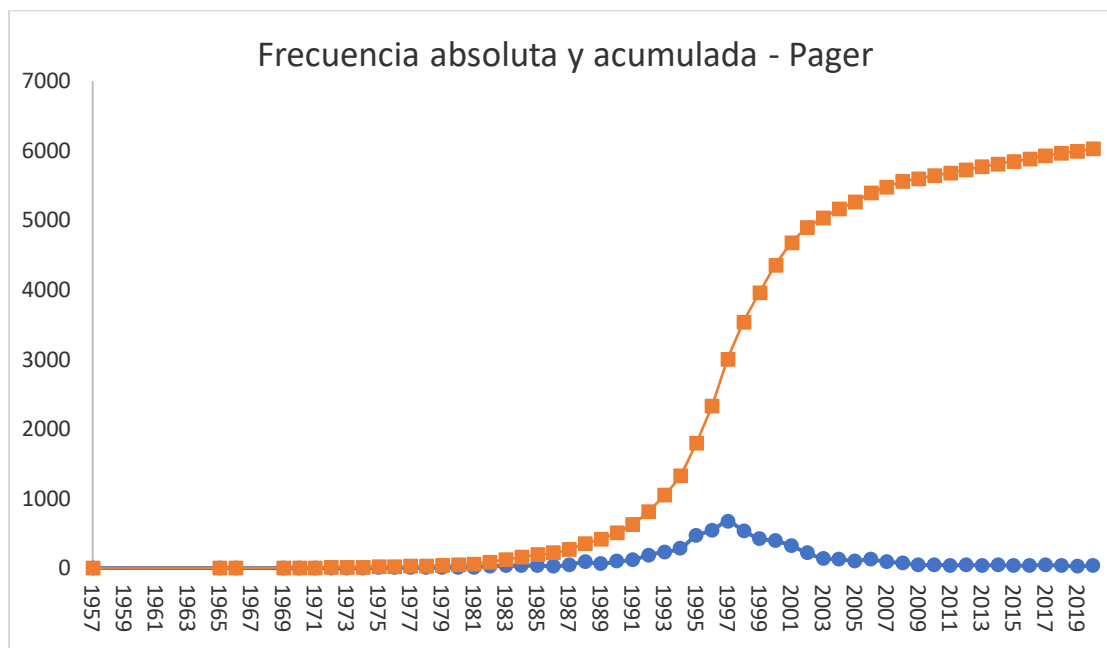


Figura 3.9 Frecuencia absoluta y acumulada de la tecnología de prueba VHS

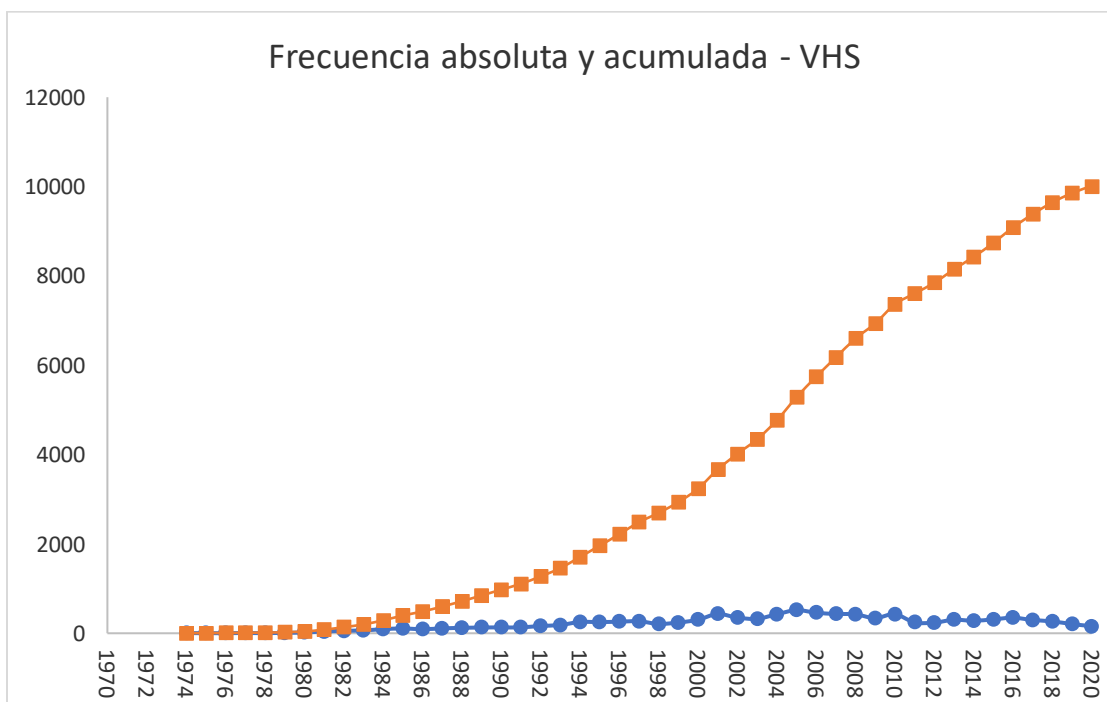


Figura 3.10 Frecuencia absoluta y acumulada de la tecnología de prueba FAX

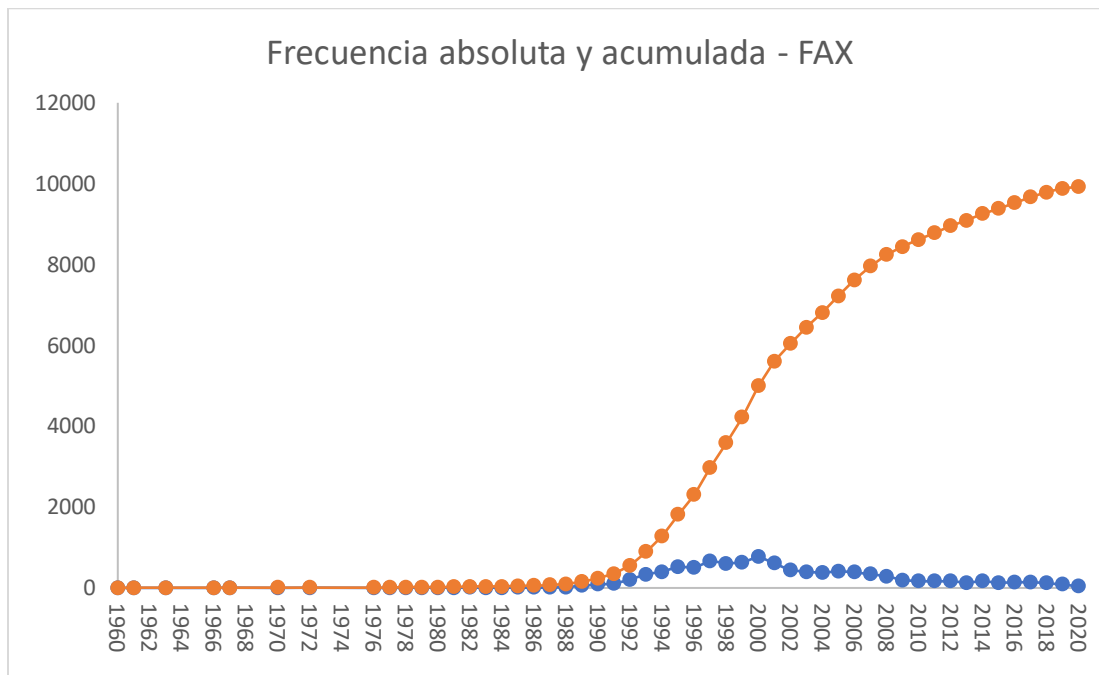


Figura 3.11 Frecuencia absoluta y acumulada de la tecnología de prueba Máquina de escribir

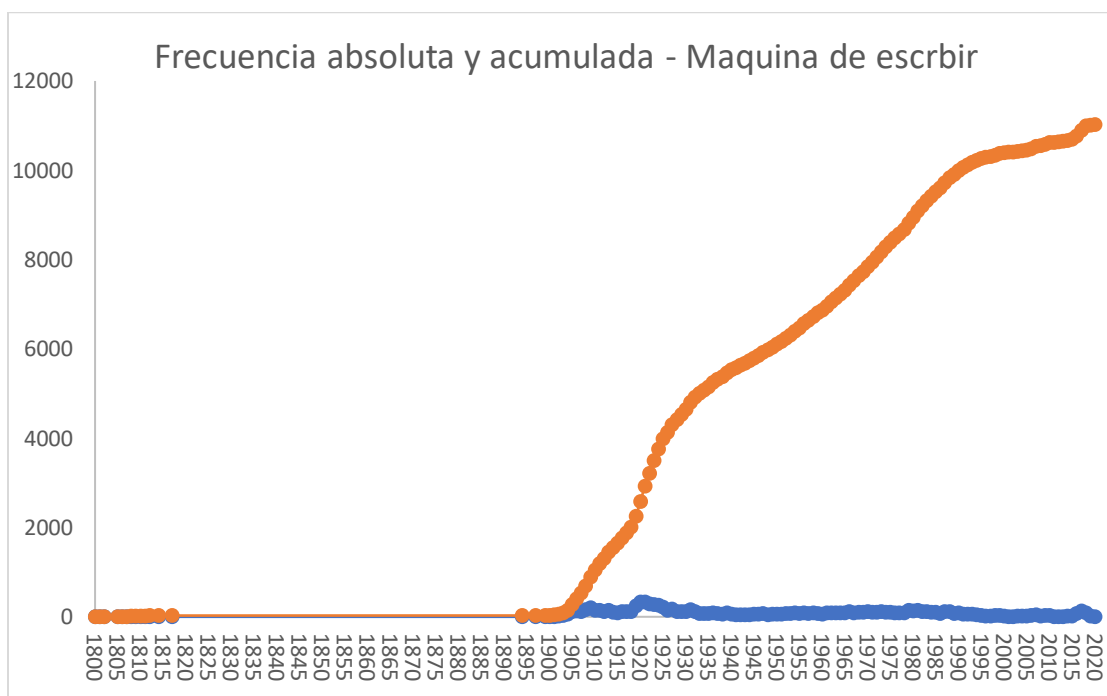
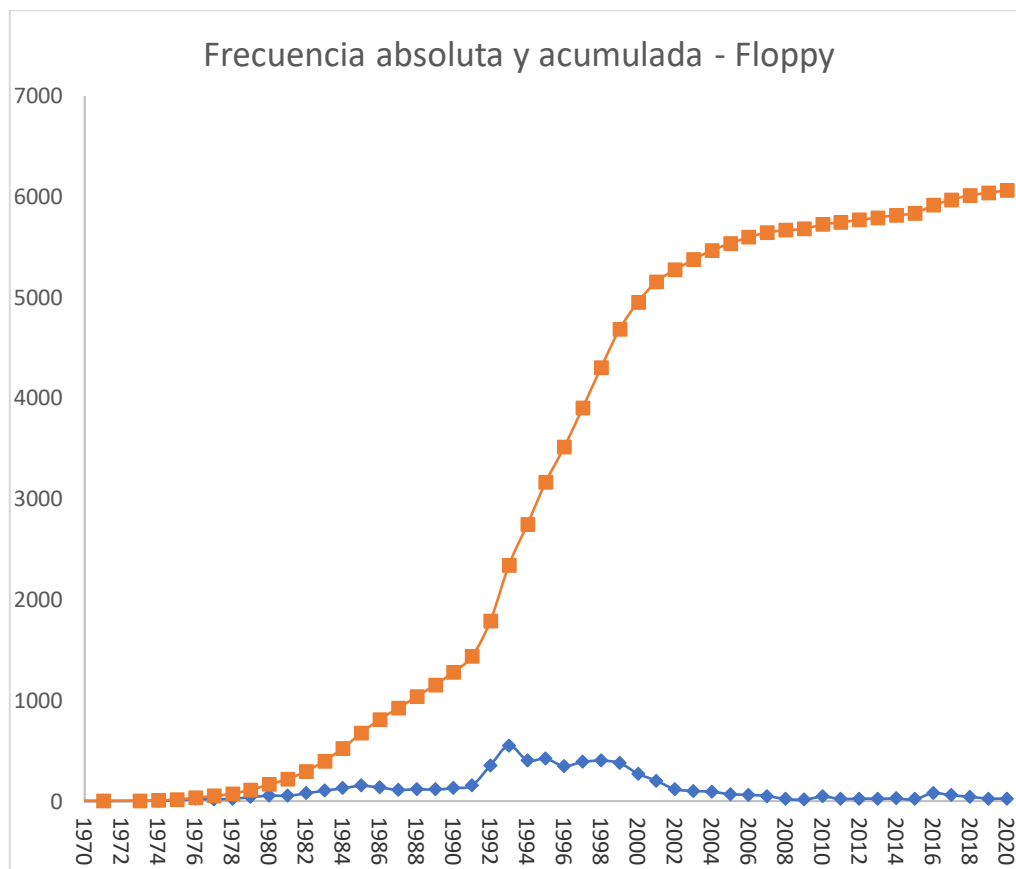


Figura 3.12 Frecuencia absoluta y acumulada de la tecnología de prueba *Floppy*



3.5.2 Cálculo de la derivada y sus curvas de tendencia

Las pruebas del método de identificación de puntos críticos, se presenta en las Figura 3.13, Figura 3.14, Figura 3.15, Figura 3.16 y Figura 3.17, en las cuales se grafican con líneas color azul las curvas de las derivadas de las frecuencias acumuladas y se traza en color verde las tendencias lineales y en color rojo las tendencias polinomiales, identificando los puntos críticos a), b) y c), según se indicó previamente.

Dentro del área de cada figura se encuentran las ecuaciones de las tendencias lineal y polinomial, que permiten la construcción del *set* de datos de las tendencias.

Figura 3.13 Uso de derivadas para identificación de las etapas en la tecnología *Pager*

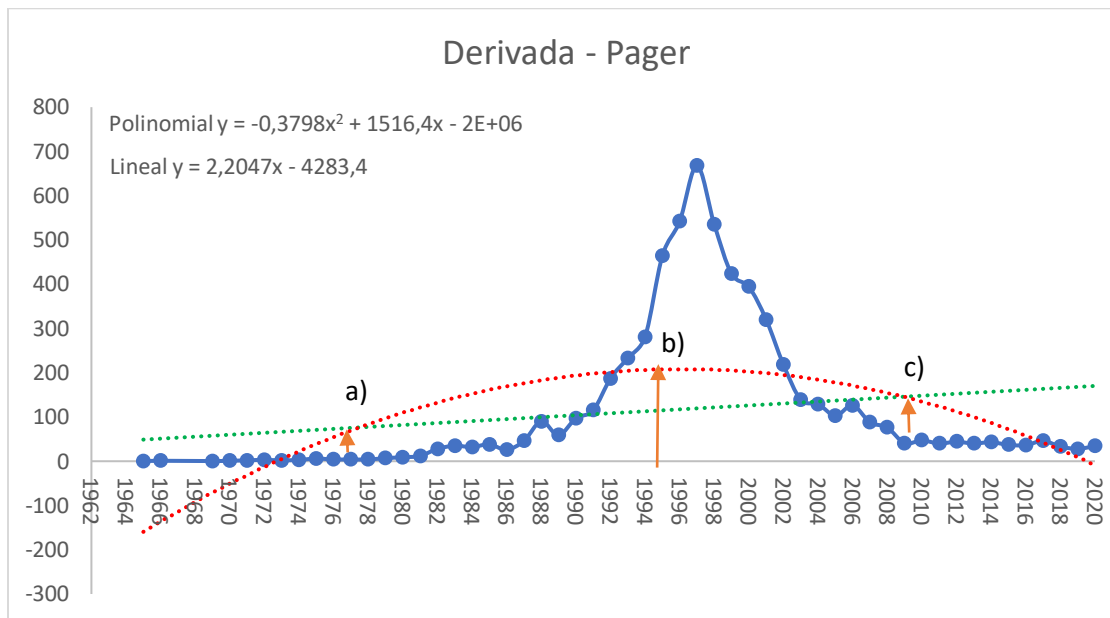


Figura 3.14 Uso de derivadas para identificación de las etapas en la tecnología VHS

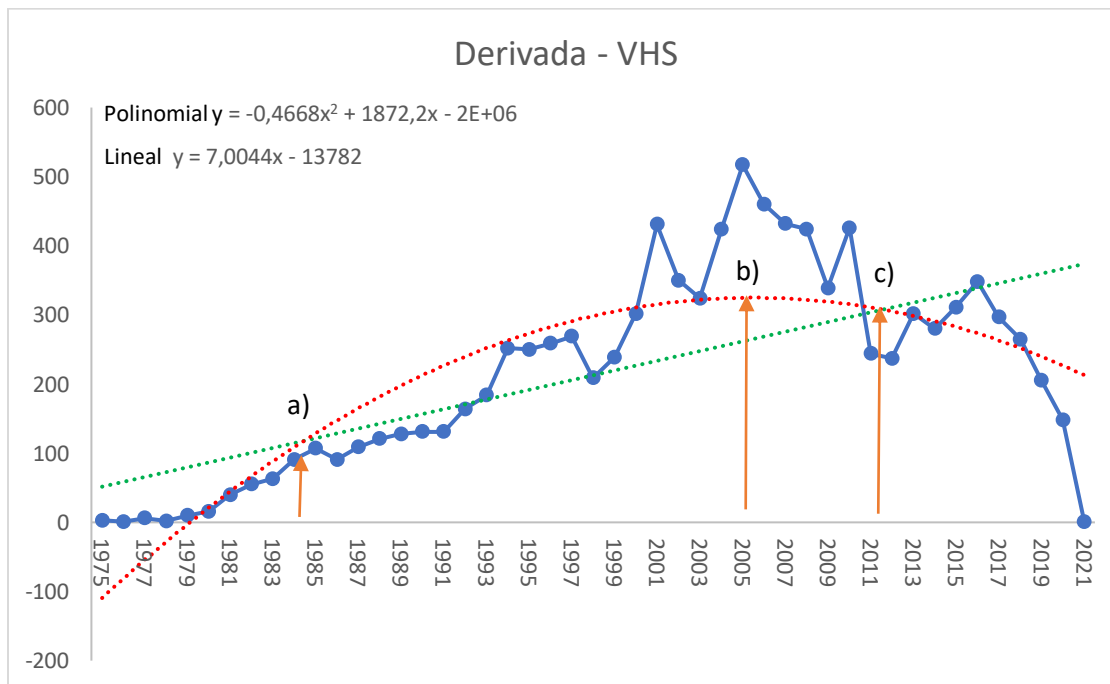


Figura 3.15 Uso de derivadas para identificación de las etapas en la tecnología FAX

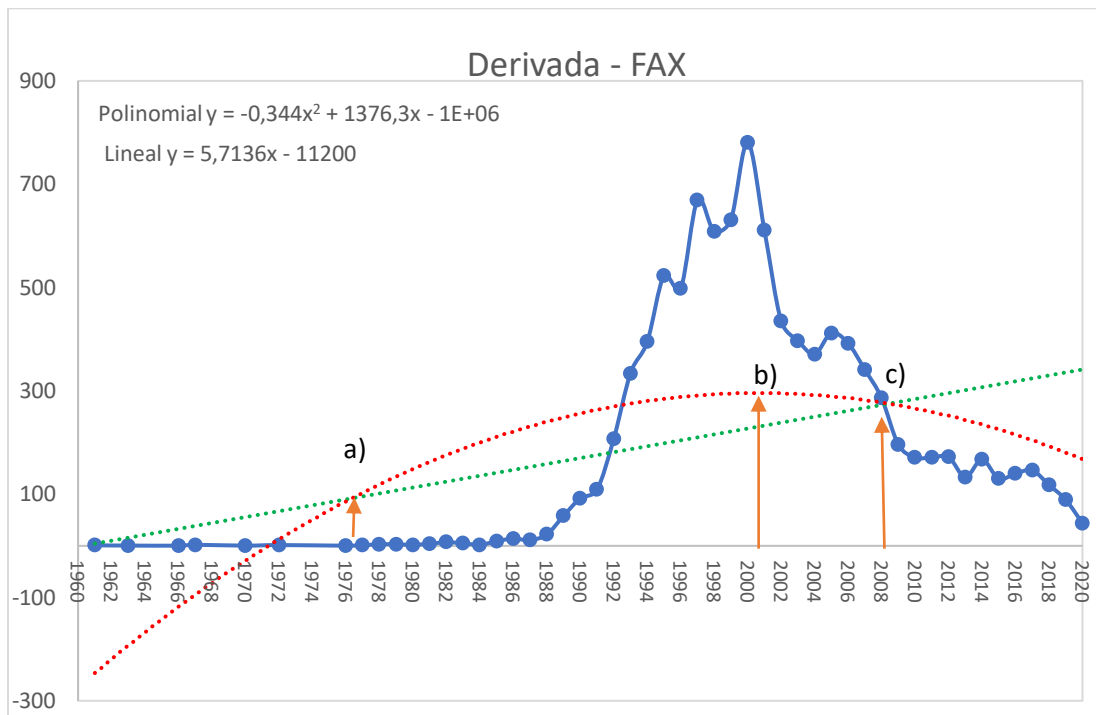


Figura 3.16 Uso de derivadas para identificación de las etapas en la tecnología Máquina de escribir

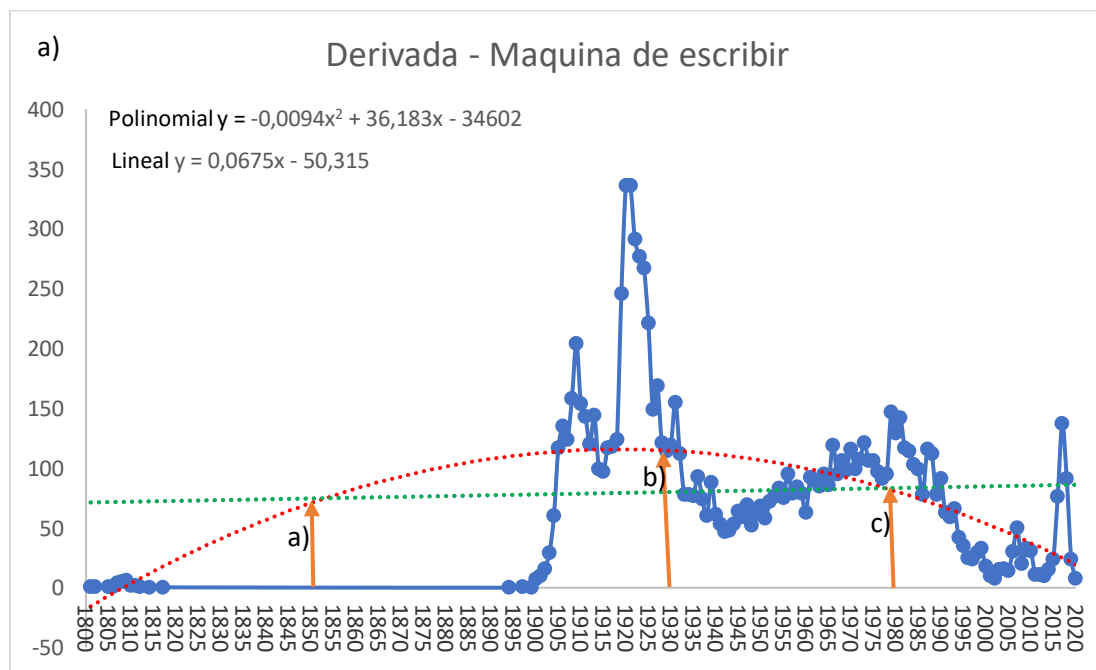
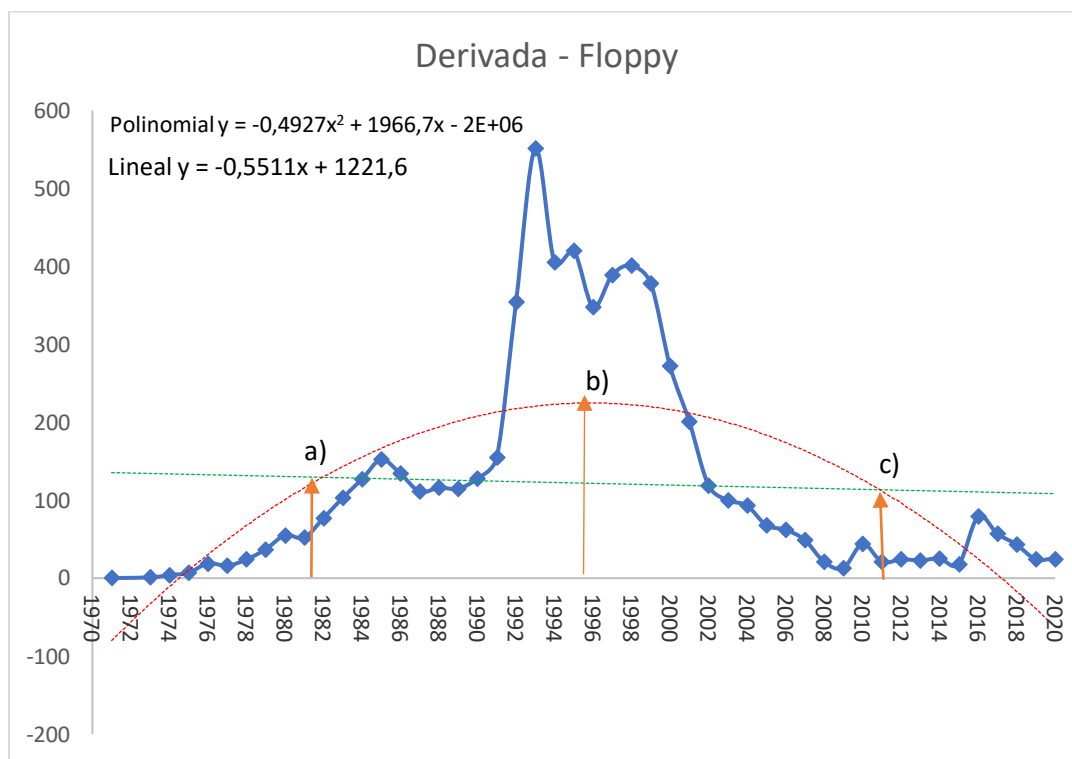


Figura 3.17 Uso de derivadas para identificación de las etapas en la tecnología *Floppy*



3.5.3 Puntos críticos y etapas en la sigmoide

Una vez identificados los puntos críticos a), b) y c), estos son trasladados a la curva de frecuencia acumulada, donde se resaltan en color rojo los años correspondientes, según se observa en las Figura 3.18, Figura 3.19, Figura 3.20, Figura 3.21 y Figura 3.22, en estos mismos puntos se trazan rectas de separación, creando la segmentación de 4 las etapas: introducción, crecimiento, madurez y declive.

Se puede observar con claridad cómo independientemente de la duración de la curva de ciclo de vida y sus sesgos, es posible generar una identificación exacta de los puntos críticos y como éstos permiten la segmentación de la sigmoide, comprobando los postulados teóricos de la de evolución tecnológica y su ciclo de vida, según presentaron Theodore Levitt (1965b), Frank Bass (1969b) y Everett Rogers y (1976b).

Figura 3.18 Puntos críticos y etapas identificadas en la tecnología de prueba *Pager*

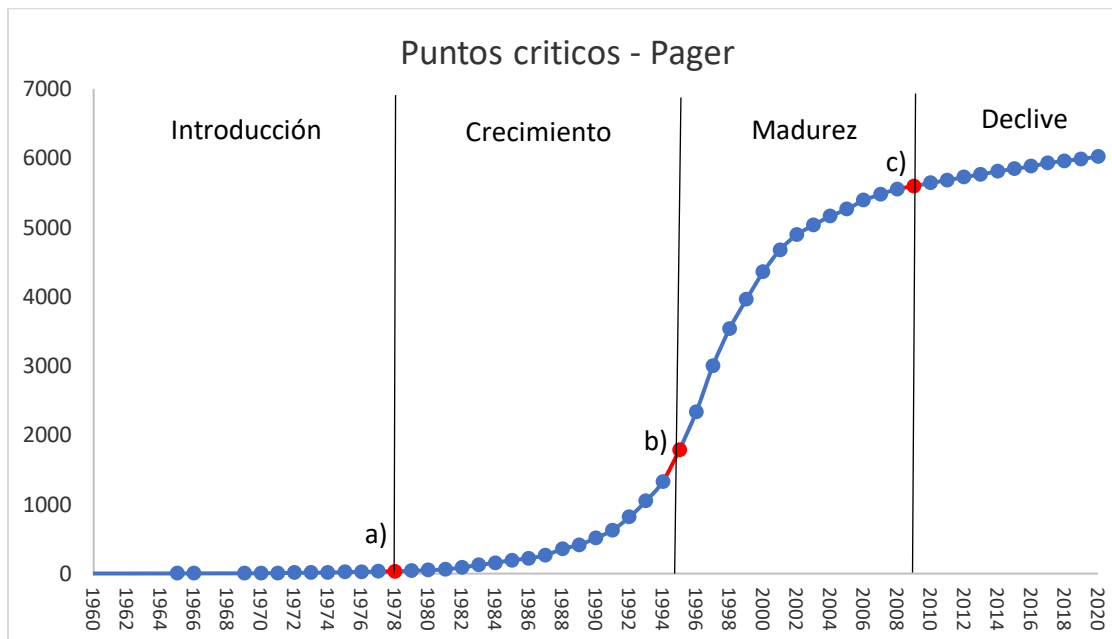


Figura 3.19 Puntos críticos y etapas identificadas en la tecnología de prueba VHS

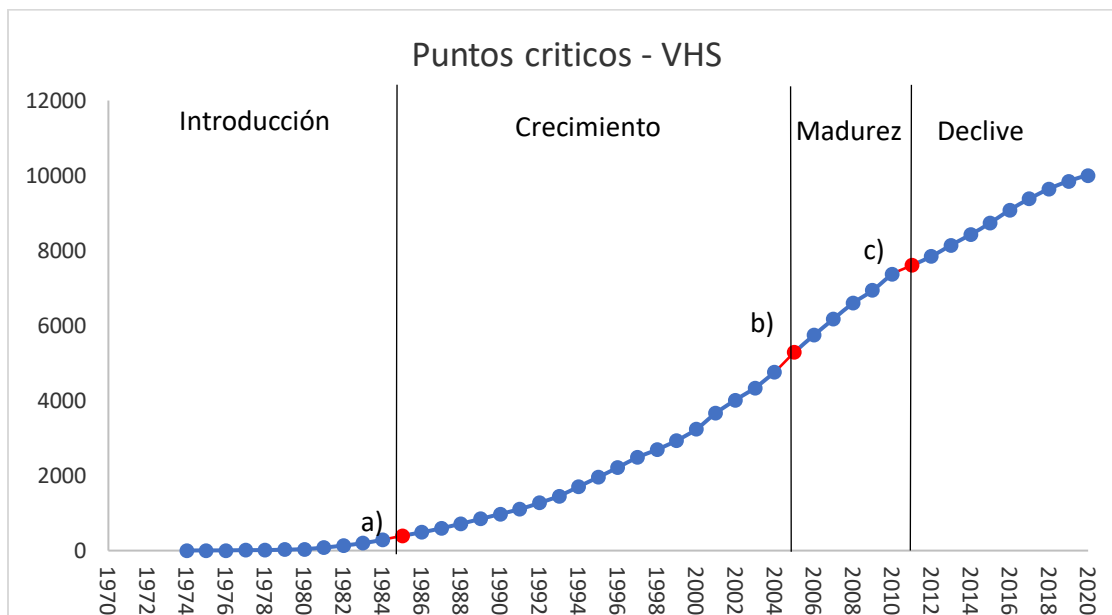


Figura 3.20 Puntos críticos y etapas identificadas en la tecnología de prueba FAX

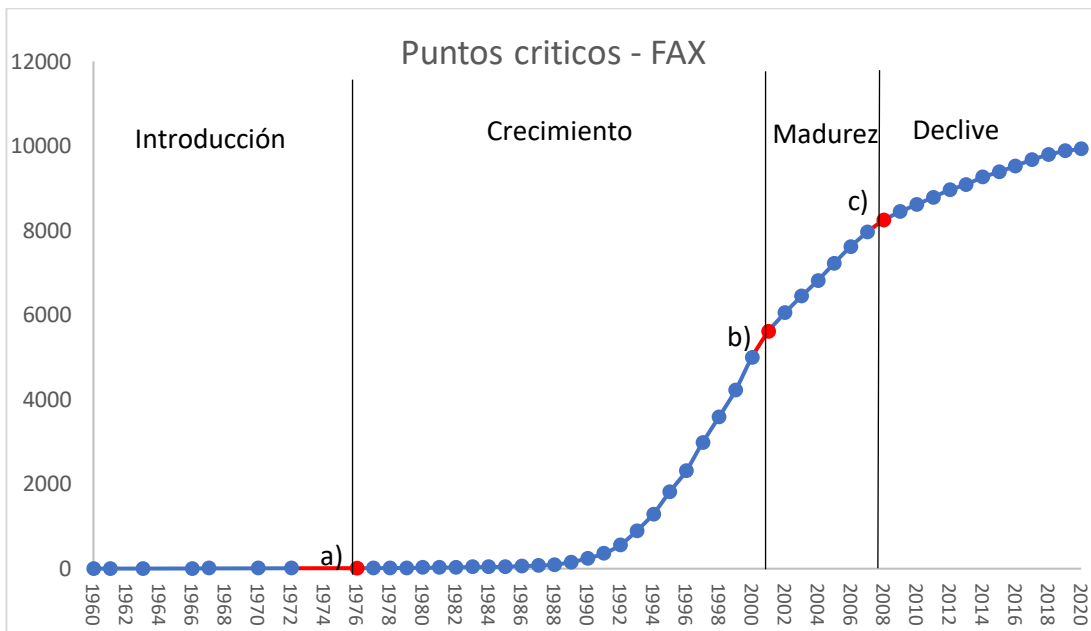


Figura 3.21 Puntos críticos y etapas identificadas en la tecnología de prueba Máquina de escribir

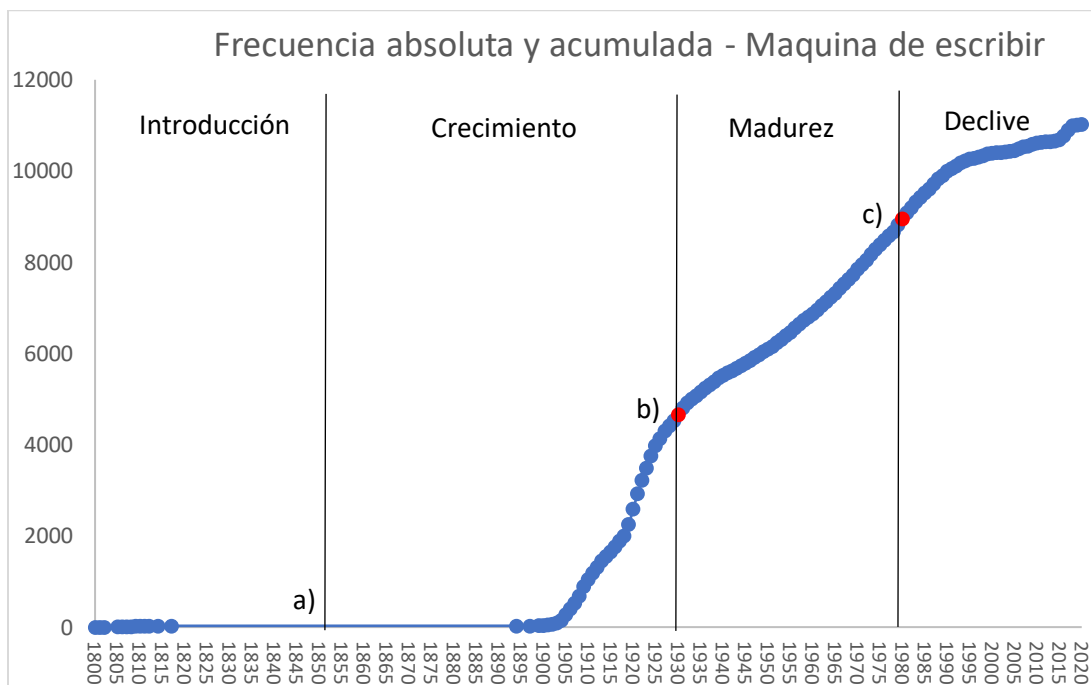
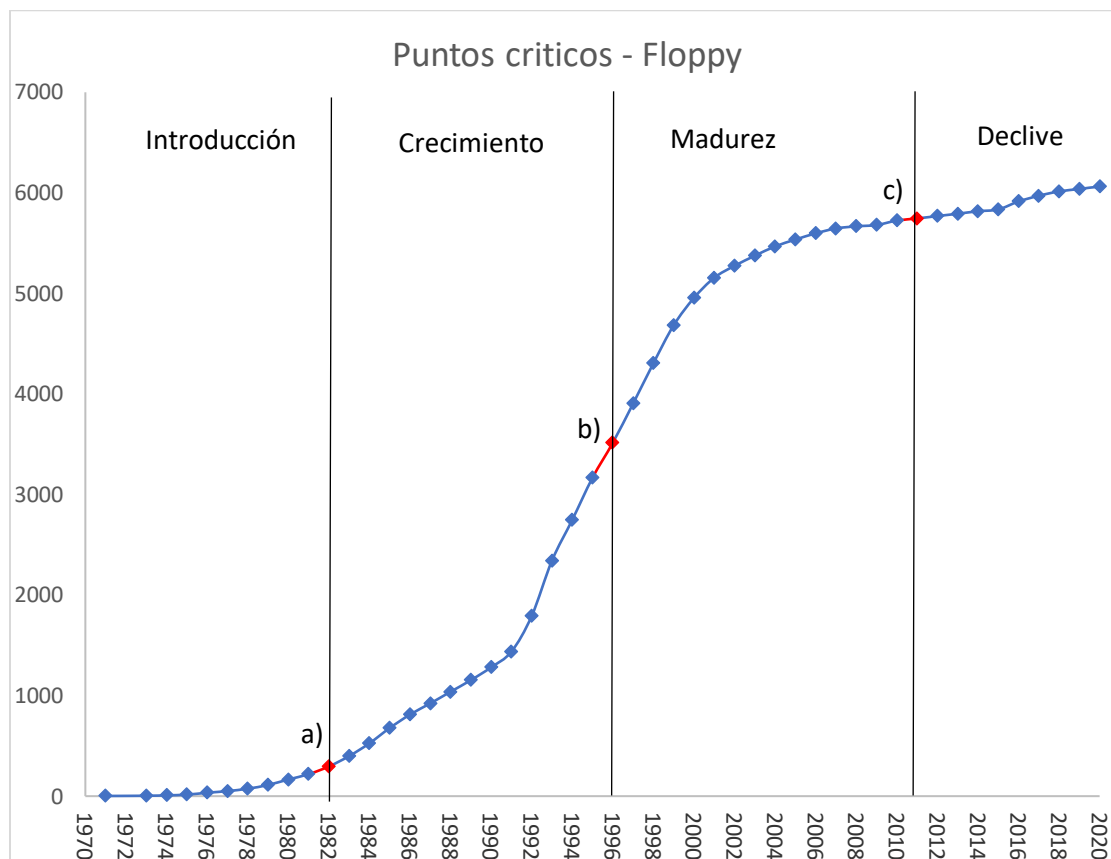


Figura 3.22 Puntos críticos y etapas identificadas en la tecnología de prueba *Floppy*



Como se puede observar, el comportamiento de las curvas responde a la forma sigmoide esperada, aunque los horizontes de estudio de las tecnologías de prueba varían en el rango de los 54 a los 191 años, las inclinaciones son diferentes, por ende la duración y alcance de las etapas, se demuestra el funcionamiento del método de identificación basado en derivadas, puesto que funciona para todas ellas, aun cuando la forma sigmoide no es evidente a primera vista como en el caso de la tecnología de máquina de escribir.

4 Caracterización del Sistema Sociotécnico

Las tecnologías al igual que muchos otros factores no evolucionan de manera independiente, por el contrario, hacen parte de sistemas complejos, en los cuales diferentes elementos económicos, sociales y medioambientales hacen parte constitutiva y afectan Interrelacionalmente su comportamiento.

Este proceso de coexistencia de factores y relaciones obedece al denominado sistema sociotécnico, presentado como una relación entre la tecnología y los elementos presentes en el contexto de la misma, encontrando variables que posibilitan, ralentizan y aceleran, así como variables que identifican usuarios, mercados, condiciones del sistema político, económico, social y medio ambiental.

El sistema sociotécnico se concibe como los elementos y condiciones presentes en el entorno de la evolución tecnológica, en este sentido el sistema sobrepasa los límites políticos de los países y continentes, refiriéndose a variables medidas mundialmente que dan cuenta de condiciones generales en un momento determinado.

La superación del referente territorial del sistema sociotécnico se genera en diferentes niveles: i) el nivel interno de la generación de cambios tecnológicos, concernientes a las condiciones propias de la investigación y desarrollo, incluyendo los recursos disponibles para que esto ocurra, ii) el nivel local, corresponde a las condiciones existentes en la sociedad, área geográfica y economía en la cual se materializa el desarrollo tecnológico incremental, dentro de ellos son especialmente relevante las condiciones políticas, de incentivos y financiación para la realización del desarrollo tecnológico, iii) el tercer nivel es regional, que agrupa condiciones de límites político administrativos y disponibilidad de recursos, se incluyen factores culturales y medio ambientales que permiten dicho desarrollo, iv) el ultimo nivel es mundial, en el cual todas

las condiciones de las diferentes, regiones, se agrupan y ofrecen un panorama amplio de información que representa el contexto general para el cambio tecnológico.

También existen condiciones transversales en el nivel mundial, tales como la disponibilidad de recursos y materias primas, la facilidad de transporte, el intercambio comercial, la condición de paz o guerra, la propiedad intelectual, el tamaño del área productiva donde se aplica una tecnología, su importancia estratégica, y la velocidad de cambio de otras tecnologías relacionadas (especialmente relevante en el sentido de la interconexión de tecnologías), todas ellas aportando en la construcción de un sistema sociotécnico exhaustivo en su alcance.

En esta tesis se usa el sistema sociotécnico a nivel mundial, tomando datos abiertos de entidades multilaterales de referencia mundial, siempre obtenidas desde sus fuentes oficiales disponibles en internet.

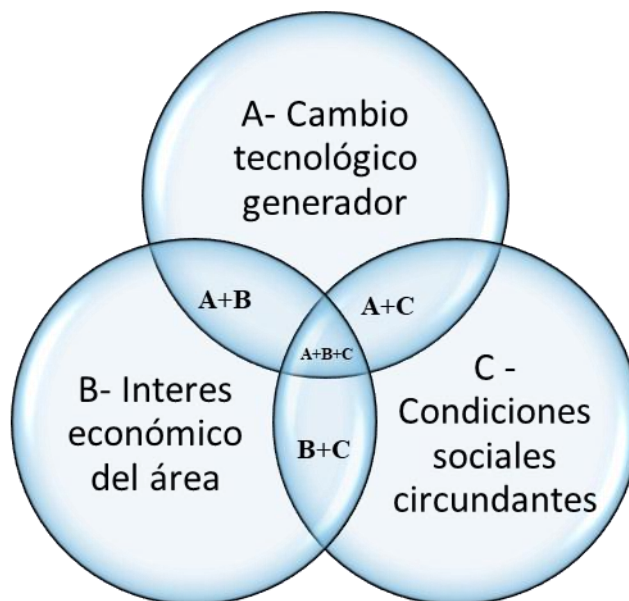
La estructura del sistema propuesto, establece la coexistencia de tres pilares que soportan el cambio tecnológico, cada uno de los pilares tiene asociadas áreas de interés y dentro de cada área se establecen variables que los representan, en el entendido que la configuración del contexto general de las condiciones de coexistencia de la tecnología es muy amplio, se busca la inclusión del mayor número de variables, siempre que éstas sean relevantes, homogéneas y de alcance mundial, como criterios mínimos para ser tenidas en cuenta dentro de la configuración del sistema.

4.1 Pilares del sistema sociotécnico

La coexistencia de factores en el modelo de predicción responde a un ordenamiento de tres pilares según se presenta en la Figura 4.1 en la cual el pilar A, representa el cambio tecnológico generador, específicamente hace referencia a las patentes como mecanismo de evidencia de cambio tecnológico, definidas al nivel de sub clase de la OMPI, el pilar B, denominado interés

económico en el área, refiere la importancia del sector donde la tecnología es aplicable, así como el valor estratégico de esta, siendo específico al ámbito de cada tecnología, el pilar C, llamado condiciones sociales circundantes, incorpora las variables transversales y es aplicable para cualquier tecnología.

Figura 4.1 Pilares del sistema sociotécnico



La combinación de los tres pilares establece un área de trabajo fértil para el análisis de los ciclos de vida de la tecnología, las uniones de los pilares posibilitan distintas salidas del sistema sociotécnico, de esta forma los resultados de las variables de los pilares A y B generan condiciones distintas a las generadas por la unión de resultados en las variables de los pilares B y C, la estructura de las uniones de pilares con resultados agregados positivos planteada es la siguiente:

- $A(+)$ + $B(+)$ = Posibilidad de adopción tecnológica – Se cuenta con cambios tecnológicos en el área (*Technology push*) y el interés económico también es positivo, haciendo posible la financiación de la materialización comercial de los cambios,

aunque las condiciones sociales no están definidas, es posible que se presente la innovación en nichos de mercado específicos.

- $B(+) + C(+) =$ Existe un sector económico fuerte, sumado a las condiciones de mercado y sociales en disposición del cambio, pero sin generación de nuevas tecnologías, denotando una estructura de cambio motivado por el mercado (*Market pull*).
- $A(+) + C(+) =$ Se generan nuevas tecnologías (*Technology push*), la sociedad está en disposición y condiciones para el cambio, pero se carece de industria financiadora, es posible que el sector económico este estancado, se ubique en una zona de confort en la producción o este saturado.
- $A(+) + B(+) + C(+) =$ Es el sistema sociotécnico óptimo para la evolución, se tienen todas las condiciones propicias para el cambio tecnológico incremental.

Cada una de las posibles combinaciones genera análisis más profundos, los cuales teóricamente no serán abordados, puesto que la investigación se realiza solo sobre las variables del sistema sociotécnico del sector agricultura, para una tecnología específica, pero dejando el espacio abierto para desarrollos más profundos en la configuración y opciones del modelo de pilares para el cambio tecnológico.

4.2 Variables del sistema sociotécnico

Con el objeto de establecer condiciones mínimas viables para la selección e inclusión de variables dentro del sistema sociotécnico, se instituyen las siguientes reglas sobre los datos:

- Los datos deben ser de libre acceso
- Los datos deben provenir de fuentes oficiales
- Los datos deben presentarse a nivel mundial, no son aceptables grupos de países,

bloques económicos o regiones

- Las variables deben tener un método de medición estable que garantice la homogeneidad en el tiempo
- Los datos no deben ser proyecciones o extrapolaciones, sino datos reales
- Los datos deben tener un periodo de medición superior a 25 años continuos, para ser tenidos en cuenta.

Establecidas las condiciones, se procede a identificar y explorar las fuentes candidatas por medio de descarga de variables y generación de una matriz inicial de características de las variables, sobre esta información se realiza calificación de las fuentes identificadas, según se presenta en la Tabla 2, estableciendo en las columnas las condiciones y en las filas las fuentes calificadas, obteniendo siete fuentes viables y once fuentes no viables.

Tabla 2 Condiciones de fuentes de datos

Fuente	Acceso	Origen	Nivel	Homogeneidad	Tipo	Periodo	Resultado
OMPI	Libre	Oficial	Mundial	Homogéneas	Real	>100	Viable
Banco Mundial BM	Libre	Oficial	Mundial	Homogéneas	Real	>60	Viable
Naciones unidas ONU	Libre	Oficial	Mundial	Homogéneas	Real	>60	Viable
World federation of exchanges - WFE	Libre	Oficial	Mundial	Homogéneas	Real	>60	Viable
FAO	Libre	Oficial	Mundial	Homogéneas	Real	>60	Viable
Unesco	Libre	Oficial	Mundial	Homogéneas	Real	>60	Viable
Doing business	Libre	Oficial	Mundial	Homogéneas	Real	>25	Viable
OCDE	Libre	Oficial	Grupal	Homogéneas	Real	>50	No viable
EMIS University,	Restringida	Secundaria	Mundial	Homogéneas	Real y proyección	>25	No viable
Fitch Solutions,	Restringida	Secundaria	Mundial	Homogéneas	Real y proyección	>25	No viable
Passport, Factiva	Restringida	Secundaria	Mundial	Homogéneas	Real y proyección	>25	No viable

Fuente	Acceso	Origen	Nivel	Homogeneidad	Tipo	Periodo	Resultado
Thomson Reuters	Restringida	Secundaria	Mundial	Homogéneas	Real y proyección	>25	No viable
Knoema	Libre	Secundaria	Mundial	No homogénea	Real y proyección	>40	No viable
Gapminder	Libre	Secundaria	Mundial	Homogéneas	Real y proyección	>25	No viable
Our world in data	Libre	Secundaria	Mundial	Homogéneas	Real y proyección	>100	No viable
OEA	Libre	Oficial	Regional	Homogéneas	Real	>60	No viable
Unión europea - UE	Libre	Oficial	Regional	Homogéneas	Real	>25	No viable
Unión Africana - UA	Libre	Oficial	Regional	Homogéneas	Real y proyección	>25	No viable

Del grupo de fuentes no viables, se obtuvieron 277 variables, que fueron eliminadas, algunas de ellas reemplazadas por su fuente original, dado que presentaban duplicidad, las restantes no fueron reemplazadas por incumplir las condiciones mínimas requeridas.

Tomando las fuentes viables, se procede con la descarga de bases de datos de variables, organizando estas según su fuente de información, como se presenta en la Tabla 3, para el caso del Banco Mundial se requirió de segmentación de descarga usando series temáticas y la primera letra del nombre de las variables como identificador de grupo, generando así catorce grupos de descargas de indicadores de desarrollo mundial. Similar situación presentan las variables de Naciones Unidas, las que se dividieron por su oficina generadora, de tal forma que de la FAO se descargaron once grupos de variables, divididos según el área de información contenida. Finalmente, de la UNESCO se descargaron cuatro grupos de variables siguiendo su propia estructura división temática.

Tabla 3 Variables iniciales del sistema

Grupo variables	Cantidad Original	Volumen de datos
BM-Global Economic Monitor	19	2.408.256
BM-Human Capital Index	25	9.390
BM-Acciones negociadas	3	51.931
Doing Business	32	149.207
Indicadores desarrollo mundial AB	20	345.827
Indicadores desarrollo mundial C	38	657.011
Indicadores desarrollo mundial E	36	662.435
Indicadores desarrollo mundial FGH	26	449.555
Indicadores desarrollo mundial IL	35	605.147
Indicadores desarrollo mundial M	15	295.387
Indicadores desarrollo mundial NO	15	295.387
Indicadores desarrollo mundial P	38	657.011
Indicadores desarrollo mundial R	18	311.251
Indicadores desarrollo mundial S	20	345.827
Indicadores desarrollo mundial T1	22	380.403
Indicadores desarrollo mundial T2	19	328.539
Indicadores desarrollo mundial T3	13	224.811
Indicadores desarrollo mundial UY	7	121.083
BM -Acciones % PIB	1	14.418
FAO-Valor de producción	5	3.770.672
FAO-Precios al productor	4	1.856.852
FAO-Uso de la tierra	10	8.932
FAO-Población	5	169.830
FAO-Medio ambiente	13	173.619
FAO-Inversión extranjera directa	6	42.022
FAO-Gasto publico	6	128.104
FAO-Inversión crédito agricultura	6	65184
FAO-Indicadores de empleo	47	439.319

Grupo variables	Cantidad Original	Volumen de datos
FAO-Índices de producción	2	4.046.336
FAO-Inversión reserva de capital	12	166.535
WFE -Bolsa de valores	1	2.263.824
UNESCO-Gasto en RD	1	5.310
UNESCO-Investigadores en agricultura	1	5.313
UNESCO-Investigadores en ingeniería	1	5.312
UNESCO-Investigadores total	1	10.112
UNIDO-Estadísticas industriales	24	9.338.461
OMPI-A01B	1	230.617
TOTALES	548	31.039.230

El total de variables incluidas en la selección de descarga inicial es de 548 las que incluyen cerca de 31 millones de datos, sobre este primer conjunto de datos se procedió a organizarlos por pilares para ser depurados, según se establece en el procedimiento presentado a continuación.

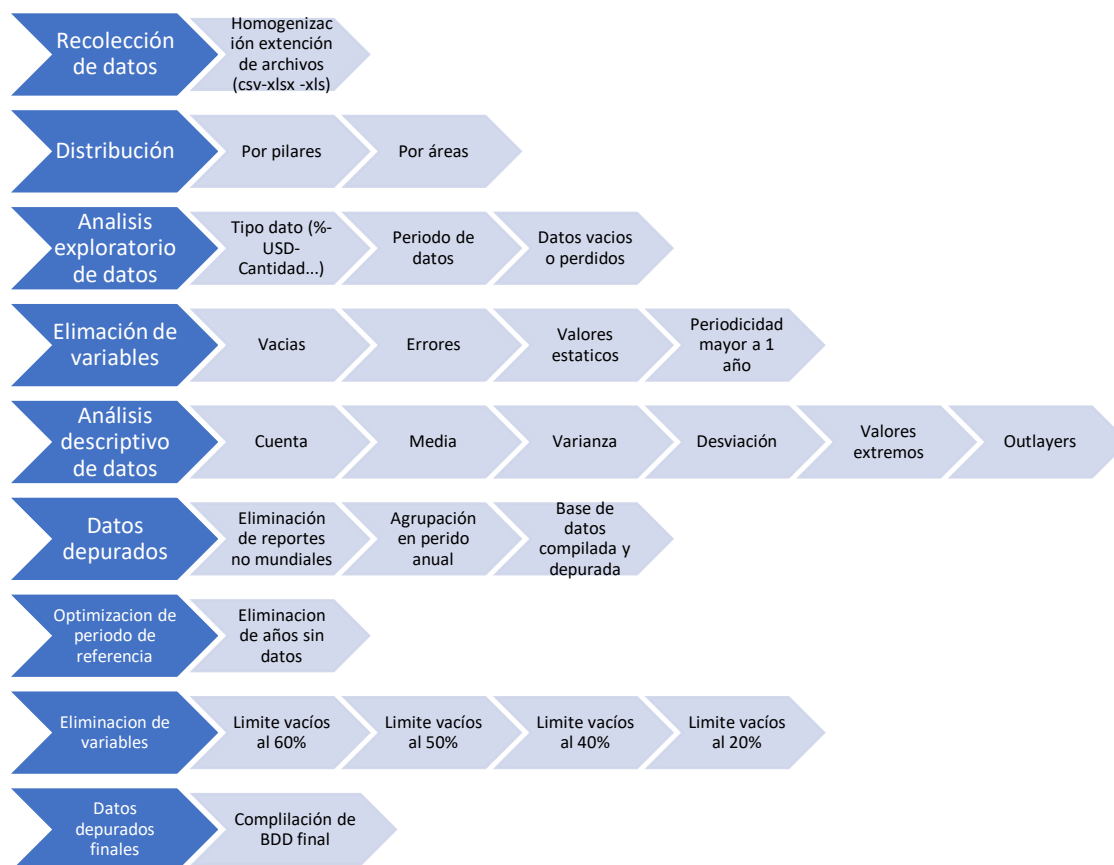
4.3 Depuración de variables

Sobre el conjunto de variables y sus datos, se efectúa limpieza y depuración, el proceso y sus reglas, se presenta en la Figura 4.2, el objeto de cumplir con estos pasos es mantener en el sistema solo las variables que cumplan criterios de calidad y pertinencia necesarios para la creación del modelo de predicción.

En este proceso es fundamental la homogeneidad en el periodo de reporte de los datos, puesto que se encuentran variables con datos de frecuencia diaria, mensual y anual, toda vez que solo 5 variables presentan reportes en periodos de días y meses, aunado a la inviabilidad de realizar la división de los datos anuales en frecuencias mensuales o diarias sin que se altere su dinámica, al alejarse de la realidad, posibilitando la ocurrencia de errores futuros, se procede a establecer la

periodicidad en reportes anuales, usando adición simple para los cinco casos encontrados.

Figura 4.2 Proceso depuración de variables



Fruto de la primera depuración se reduce el sistema sociotécnico a 298 variables y 2.6 millones de datos, el resultado se presenta en la Tabla 4, la reducción de volumen obedece a la aplicación de la cuarta y sexta etapas del proceso de depuración, donde se eliminan las variables que no cumplen criterios y se eliminan los datos que no corresponden con reportes mundiales.

Tabla 4 Variables depuradas en el sistema

Grupo variables	Cantidad Depurada	Volumen de datos
BM-Global Economic Monitor	11	7.489
BM-Human Capital Index	0	0
BM-Acciones negociadas	3	260
Doing Business	4	84
Indicadores desarrollo mundial AB	14	1.105
Indicadores desarrollo mundial C	23	2.535
Indicadores desarrollo mundial E	21	2.405
Indicadores desarrollo mundial FGH	11	1.755
Indicadores desarrollo mundial IL	17	2.340
Indicadores desarrollo mundial M	5	1.040
Indicadores desarrollo mundial NO	4	845
Indicadores desarrollo mundial P	20	1.820
Indicadores desarrollo mundial R	14	1.235
Indicadores desarrollo mundial S	15	1.365
Indicadores desarrollo mundial T1	14	1.495
Indicadores desarrollo mundial T2	13	1.300
Indicadores desarrollo mundial T3	7	910
Indicadores desarrollo mundial UY	7	520
BM -Acciones % PIB	1	108
FAO-Valor de producción	3	34.891
FAO-Precios al productor	1	222.788
FAO-Uso de la tierra	9	8.932
FAO-Población	5	848
FAO-Medio ambiente	10	1.718
FAO-Inversión extranjera directa	4	16.339
FAO-Gasto publico	2	81
FAO-Inversión crédito agricultura	5	249
FAO-Indicadores de empleo	6	313
FAO-Índices de producción	2	49.377
FAO-Inversión reserva de capital	4	429
WFE -Bolsa de valores	1	434
UNESCO-Gasto en RD	1	52
UNESCO-Investigadores en agricultura	1	50
UNESCO-Investigadores en ingeniería	1	50
UNESCO-Investigadores total	1	100
UNIDO-Estadísticas industriales	7	2.138.176
OMPI-A01B	1	114.277
TOTALES	268	2.617.715

Sobre la matriz resumen de las 268 variables previamente depuradas, se seleccionan todas las celdas sin reporte de datos, asignado el valor de cero y marcado con color naranja la celda, para todos los años reportados se cuentan los vacíos y se calcula el porcentaje de vacíos por año, un extracto del porcentaje de vacíos es presentado en la Tabla 5, donde se evidencia que el periodo entre 1960-1969, presenta valores superiores al 70% de vacíos, lo que obliga a la eliminación de la década como fuente de datos, según la regla 7, similar situación acontece con el año 2020, que presenta 158 vacíos, por lo cual este año también es eliminado del periodo de referencia lo que se comprueba con una revisión general de la matriz, donde la ausencia de datos en los periodos iniciales es evidente visualmente, como lo muestra la

Figura 4.3 en la cual se observa la densidad de vacíos (color naranja) en la periodo referenciado, como marca divisora del comportamiento se asigna color rojo a los datos del año 1970, año desde el cual se continua el proceso de depuración.

Tabla 5 Celdas vacías por año

Año	Porcentaje de celdas vacías
1960	86,1
1961	74,5
1962	72,3
1963	71,9
1964	71,9
1965	71,9
1966	71,5
1967	71,2
1968	70,8
1969	71,9
1970	57,3
1971	55,8

Figura 4.3 Identificación visual de celdas vacías

Con el nuevo periodo de datos establecido entre 1970 a 2019, se calcula nuevamente la incidencia de vacíos en columnas y filas, el resultado permite identificar variables con bajo reporte de datos en el periodo de referencia (Daim et al., 2016), por lo cual se acude a las reglas del octavo nivel del proceso de depuración (limite vacíos al 60%, limite vacíos al 50%, limite vacíos al 40% y limite vacíos al 20%), eliminando las variables que no cuenten con datos completos en cada uno de los cuatro escenarios límite.

El grupo de variables que presentan vacíos en sus datos, igual o superior al 60% del periodo

de referencia, son eliminadas de la matriz, del primer grupo se eliminan 37 variables, la referencia de la eliminación se presenta en la Figura 4.4, el segundo grupo de variables que incumplen la regla de límite de vacíos del 50% del periodo de referencia contiene 27 variables, que se presentan en la Figura 4.5, el tercer grupo de eliminación es de 63 variables, las cuales superan el límite de reportes vacíos del 40% mostrado en la Figura 4.6, el grupo final de eliminación donde la regla es más restrictiva, corresponde a 16 variables, las que presentan vacíos en las del 20% del periodo de referencia, eliminando en el proceso un total de 143 variables por ausencia de datos en el periodo de referencia.

Figura 4.4 Eliminación de variables límite de vacíos 60%

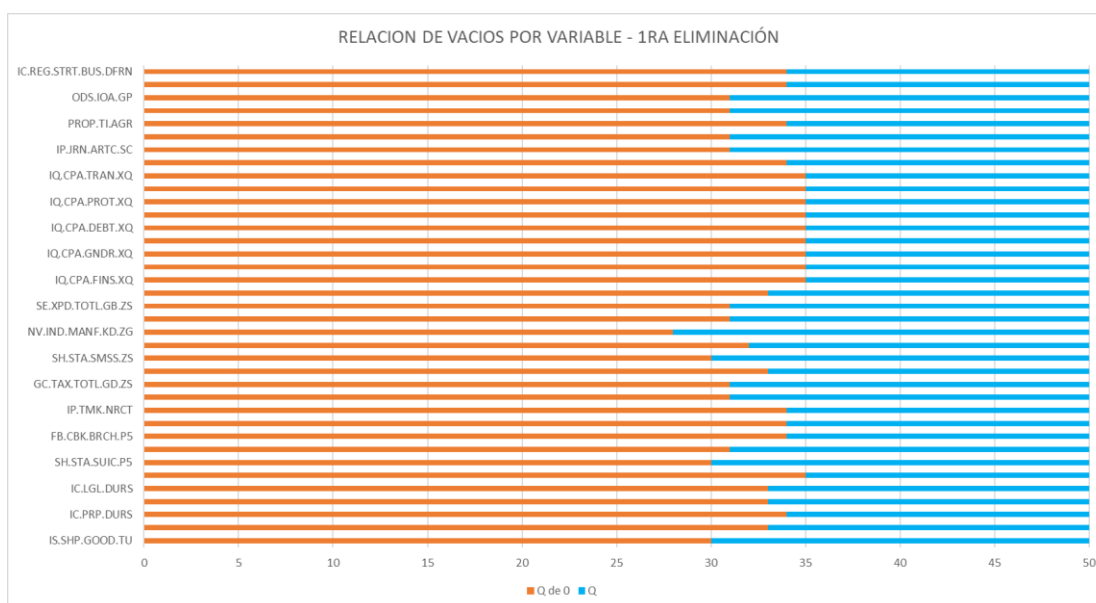


Figura 4.5 Eliminación de variables límite de vacíos 50%

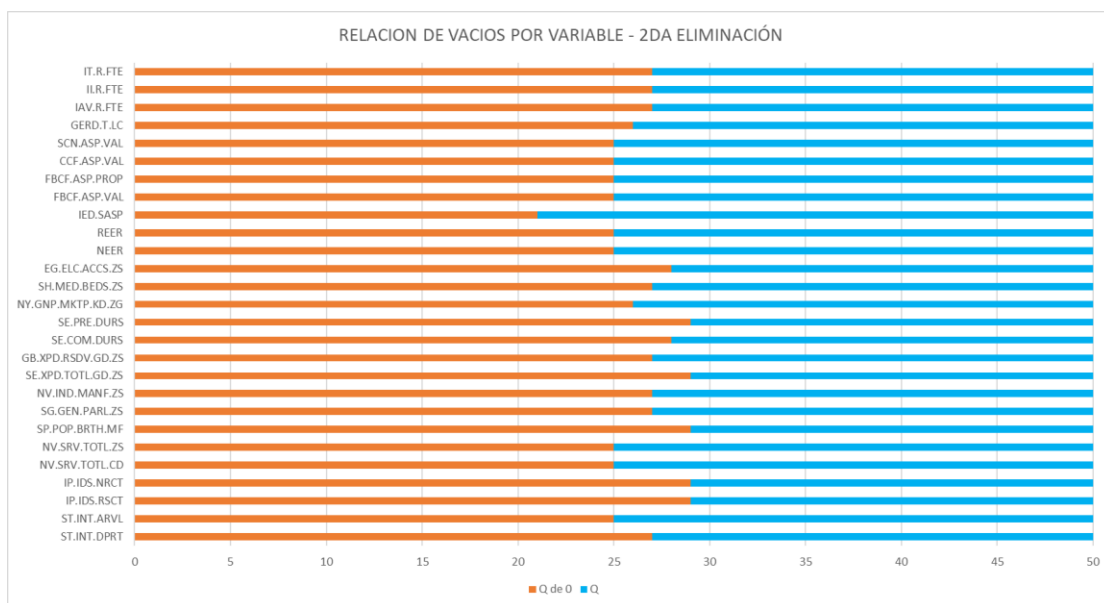


Figura 4.6 Eliminación de variables límite de vacíos 40%

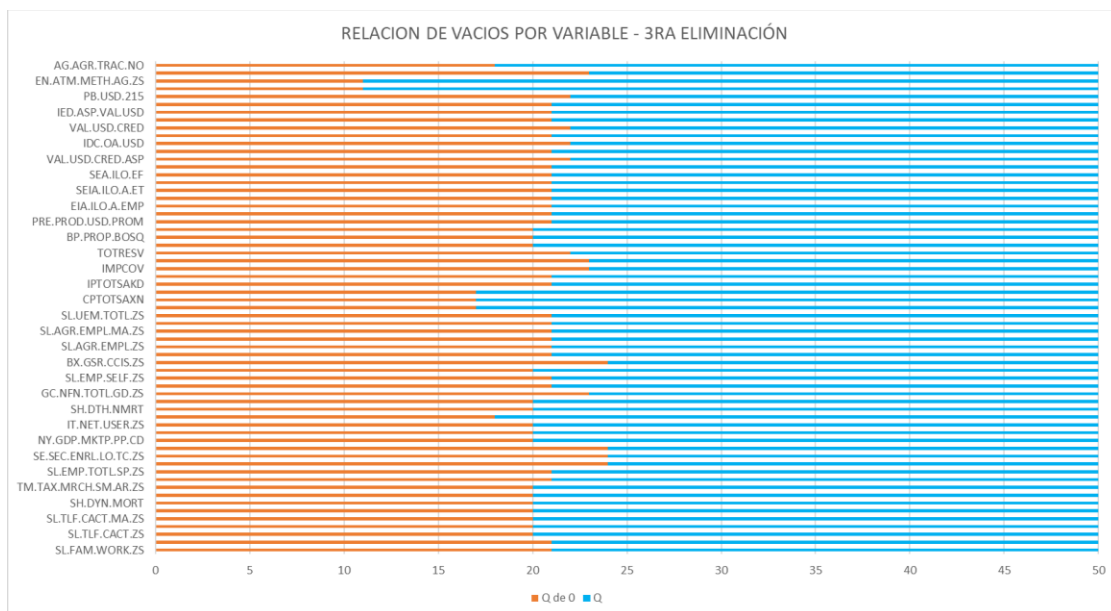
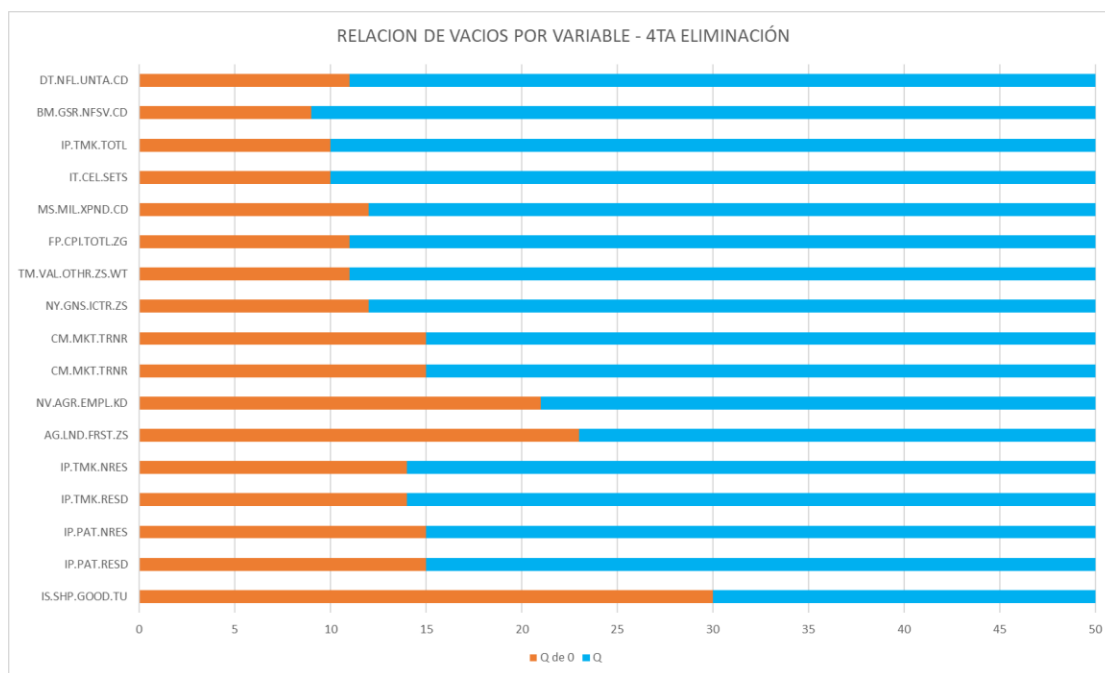


Figura 4.7 Eliminación de variables límite de vacíos 20%

La matriz final resultante, que contiene 125 variables, en el rango temporal 1970-2019, se verifica nuevamente la incidencia de vacíos, siendo estos tratados por imputación, el procedimiento inicial es la verificación en la fuente original para comprobar la existencia del dato o su pérdida en el proceso, una vez comprobado el vacío, se imputa el dato faltante por análisis de casos disponibles, usando método de regresión simple (Velásquez & Martínez, 2009) y se comprueba por medias móviles, buscando no afectar la varianza y la correlación (Medina & Galván, 2007, p. 27), como resultado se imputan 139 datos equivalentes al 2.15% de la matriz final.

Una vez comprobado que la matriz de variables no presenta vacíos, se asignan códigos temporales a las variables, que operan como identificadores en el modelo, y facilitan su gestión en el software a usar, estos códigos corresponden a la secuencia D1 a D124 para las variables del pilar B y el pilar C, mientras para el pilar A, se usa el código WIPO1.

Con el fin de facilitar los análisis posteriores de las variables del sistema sociotécnico se

traspone la matriz, obteniendo un tamaño final de 6.426 datos divididos en 126 columnas (una de fecha, una de WIPO1 y 124 de variables del sistema) y 51 filas (una de título y 50 del rango temporal 1970-2019).

Las variables definitivas del sistema sociotécnico a usar en el modelo de predicción se presentan en la Tabla 6, correspondiente a la variable del pilar A, en la Tabla 7 las variables correspondientes al pilar B y en la Tabla 8, las variables correspondientes al pilar C, para todas las tablas se mantiene la estructura de identificación de la fuente, el nombre completo de la variable el código interno y el código a usar como identificador único dentro del modelo de predicción.

Tabla 6 Variable definitiva del pilar A

Fuente	Nombre de la variable	Código Variable	Cod Modelo
WIPO	A01B patentes presentadas por año	A01B.P.Y	WIPO1

Tabla 7 Variables definitivas del pilar B

Fuente	Nombre de la variable	Código Variable	Cod Modelo
Banco Mundial	Tierras destinadas al cultivo de manera permanente (% del área de tierra)	AG.LND.CROP.ZS	D7
Banco Mundial	Tierras cultivables (hectáreas por persona)	AG.LND.ARBL.HA.PC	D8
Banco Mundial	Tierras cultivables (% del área de tierra)	AG.LND.ARBL.ZS	D9
Banco Mundial	Tierras agrícolas (kilómetros cuadrados)	AG.LND.AGRI.K2	D10
Banco Mundial	Tierras agrícolas (% del área de tierra)	AG.LND.AGRI.ZS	D11
Banco Mundial	Tierra utilizada para la producción de cereales (hectáreas)	AG.LND.CREL.HA	D12
Banco Mundial	Rendimiento de los cereales (kg por hectárea)	AG.YLD.CREL.KG	D28
Banco Mundial	Producción de cereales (toneladas métricas)	AG.PRD.CREL.MT	D36
Banco Mundial	Población rural	SP.RUR.TOTL	D41

Fuente	Nombre de la variable	Código Variable	Cod Modelo
Banco Mundial	Índice de producción de alimentos (2004-2006 = 100)	AG.PRD.FOOD.XD	D57
Banco Mundial	Índice de cosecha (2004-2006 = 100)	AG.PRD.CROP.XD	D58
Banco Mundial	Importaciones de materias primas para la actividad agrícola (% de mercaderías importadas)	TM.VAL.AGRI.ZS.UN	D61
Banco Mundial	Exportaciones de materias primas para la actividad agrícola (% de mercaderías importadas)	TX.VAL.AGRI.ZS.UN	D71
Banco Mundial	Crecimiento de la población rural (% anual)	SP.RUR.TOTL.ZG	D84
Banco Mundial	Consumo de fertilizantes (kilogramos por hectárea de tierras cultivables)	AG.CON.FERT.ZS	D87
Banco Mundial	Área de tierra (kilómetros cuadrados)	AG.LND.TOTL.K2	D96
Banco Mundial	Agricultura, valor agregado (US\$ a precios constantes de 2010)	NV.AGR.TOTL.KD	D98
Banco Mundial	Agricultura, valor agregado (US\$ a precios actuales)	NV.AGR.TOTL.CD	D99
Banco Mundial	Agricultura, valor agregado (% del PIB)	NV.AGR.TOTL.ZS	D100
Banco Mundial	Agricultura, valor agregado (% del crecimiento anual)	NV.AGR.TOTL.KD.ZG	D101
FAO	Tierras agrícolas (Proporción de superficie de tierra %)	AGR.TRR.PROP	D105
FAO	Tierras arables (Proporción de tierras agrícolas %)	ARA.TRR.PROP	D106
FAO	Tierras destinadas a cultivos permanentes (Proporción de tierras agrícolas %)	DES.TRR.PERM	D107
FAO	Tierras de cultivo (Proporción de tierras agrícolas %)	CUL.TRR.AGR	D108
FAO	Tierras de cultivo (Proporción de superficie de tierra %)	CUL.TRR.SUPE	D109
FAO	Tierras dedicadas a praderas y pastizales permanentes (Proporción de tierras agrícolas %)	PERM.PAST.AGR	D110
FAO	Tierras dedicadas a praderas y pastizales permanentes (Proporción de superficie de tierra %)	PERM.PAST.SUPE	D111
FAO	Superficie de tierras equipada para riego (Proporción de tierras agrícolas %)	RIEG.TRR.AGR	D112
FAO	Superficie de tierras equipada para riego (Proporción de tierras de cultivo %)	RIEG.TRR.CUL	D113
FAO	Índice de producción Bruto (base 2014-2016) (<i>Vegetables and Fruit Primary índice</i> PROMEDIO)	GR.PROD.VG	D114
FAO	Índice de producción bruto per cápita (base 2014-2016) (<i>Vegetables and Fruit Primary índice</i> PROMEDIO)	GR.PROD.PCT.VG	D115

Fuente	Nombre de la variable	Código Variable	Cod Modelo
FAO	Producción Bruta 2014-2016 (miles US\$) (217 productos _1000 US\$)	PROD.VAL.217	D116
FAO	Producción Bruta 2014-2016 (miles US\$) (198 productos _1000 US\$)	PROD.USD.198	D117
FAO	Población rural (1000 personas)	RUR.POB.	D121

Tabla 8 Variables definitivas del pilar C

Fuente	Nombre de la variable	Código Variable	Cod Modelo
Banco Mundial	Acciones negociadas, valor total (% del PIB)	CM.MKT.TRAD.GD.ZS	D1
Banco Mundial	Uso de energía (kg de equivalente de petróleo per cápita)	EG.USE.PCAP.KG.OE	D2
Banco Mundial	Transporte aéreo, partidas de vuelos en todo el mundo de compañías registradas en el país	IS.AIR.DPRT	D3
Banco Mundial	Transporte aéreo, pasajeros transportados	IS.AIR.PSGR	D4
Banco Mundial	Transporte aéreo, carga (millones de toneladas-kilómetros)	IS.AIR.GOOD.MT.K1	D5
Banco Mundial	Transferencias de armamento, importaciones (valores del indicador de tendencia del Instituto Internacional de Investigación para la Paz de Estocolmo) (SIPRI)	MS.MIL.MPRT.KD	D6
Banco Mundial	Tasa de natalidad, nacidos vivos en un año (por cada 1.000 personas)	SP.DYN.CBRT.IN	D13
Banco Mundial	Tasa de mortalidad, adultos, mujeres (por cada 1.000 mujeres adultas)	SP.DYN.AMRT.FE	D14
Banco Mundial	Tasa de mortalidad, adultos, varones (por cada 1.000 varones adultos)	SP.DYN.AMRT.MA	D15
Banco Mundial	Tasa de inactividad por edades (% de la población en edad de trabajar)	SP.POP.DPND	D16
Banco Mundial	Tasa de finalización de la educación de nivel primario, total (% del grupo etario correspondiente)	SE.PRM.CMPT.ZS	D17
Banco Mundial	Tasa de fertilidad, total (nacimientos por cada mujer)	SP.DYN.TFRT.IN	D18
Banco Mundial	Tasa de alfabetización, total de jóvenes (% de personas entre 15 y 24 años)	SE.ADT.1524.LT.ZS	D19
Banco Mundial	Tasa de alfabetización, total de adultos (% de personas de 15 años o más)	SE.ADT.LITR.ZS	D20
Banco Mundial	Suscripciones a telefonía fija	IT.MLT.MAIN	D21

Fuente	Nombre de la variable	Código Variable	Cod Modelo
Banco Mundial	Rentas totales de los recursos naturales (% del PIB)	NY.GDP.TOTL.RT.ZS	D22
Banco Mundial	Rentas mineras (% del PIB)	NY.GDP.MINR.RT.ZS	D23
Banco Mundial	Rentas forestales (% del PIB)	NY.GDP.FRST.RT.ZS	D24
Banco Mundial	Rentas del petróleo (% del PIB)	NY.GDP.PETR.RT.ZS	D25
Banco Mundial	Rentas del carbón (% del PIB)	NY.GDP.COAL.RT.ZS	D26
Banco Mundial	Renta del gas natural (% del PIB)	NY.GDP.NGAS.RT.ZS	D27
Banco Mundial	Remesas de trabajadores y compensación de empleados, recibidas (% del PIB)	BX.TRF.PWKR.DT.GD.ZS	D29
Banco Mundial	Relación alumno-maestro, preescolar	SE.PRE.ENRL.TC.ZS	D30
Banco Mundial	Proporción de inscripciones de mujeres con respecto a varones en la educación terciaria (%)	SE.ENR.TERT.FM.ZS	D31
Banco Mundial	Proporción alumnos-maestro, nivel secundario	SE.SEC.ENRL.TC.ZS	D32
Banco Mundial	Proporción alumnos-maestro, nivel primario	SE.PRM.ENRL.TC.ZS	D33
Banco Mundial	Producción pesquera total (toneladas métricas)	ER.FSH.PROD.MT	D34
Banco Mundial	Producción de electricidad a partir del petróleo (% del total)	EG.ELC.PETR.ZS	D35
Banco Mundial	Población, total	SP.POP.TOTL	D37
Banco Mundial	Población, hombres	SP.POP.TOTL.MA.IN	D38
Banco Mundial	Población, mujeres	SP.POP.TOTL.FE.IN	D39
Banco Mundial	Población urbana	SP.URB.TOTL	D40
Banco Mundial	Población entre 15 y 64 años de edad, total	SP.POP.1564.TO	D42
Banco Mundial	Población entre 0 y 14 años de edad, total	SP.POP.0014.TO	D43
Banco Mundial	Población de 65 años de edad y más, total	SP.POP.65UP.TO	D44
Banco Mundial	PIB per cápita (US\$ a precios actuales)	NY.GDP.PCAP.CD	D45
Banco Mundial	PIB (US\$ a precios actuales)	NY.GDP.MKTP.CD	D46
Banco Mundial	Niños que no asisten a la escuela, nivel primario	SE.PRM.UNER	D47
Banco Mundial	Mercaderías exportadas por la economía informante (US\$ actuales)	TX.VAL.MRCH.WL.CD	D48
Banco Mundial	Masa monetaria (% del PIB)	FM.LBL.BMNY.GD.ZS	D49

Fuente	Nombre de la variable	Código Variable	Cod Modelo
Banco Mundial	Inversión extranjera directa, salida neta de capital (% del PIB)	BM.KLT.DINV.WD.GD.ZS	D50
Banco Mundial	Inversión extranjera directa, entrada neta de capital (% del PIB)	BX.KLT.DINV.WD.GD.ZS	D51
Banco Mundial	Inscripción escolar, nivel terciario (% bruto)	SE.TER.ENRR	D52
Banco Mundial	Inscripción escolar, nivel secundario (% bruto)	SE.SEC.ENRR	D53
Banco Mundial	Inscripción escolar, nivel primario (% bruto)	SE.PRM.ENRR	D54
Banco Mundial	Ingreso nacional neto ajustado (% de crecimiento anual)	NY.ADJ.NNTY.KD.ZG	D55
Banco Mundial	Ingreso nacional bruto (ING) (US\$)	NY.GNP.MKTP.CD	D56
Banco Mundial	Importaciones de minerales y metales (% de importaciones de mercaderías)	TM.VAL.MMTL.ZS.UN	D59
Banco Mundial	Importaciones de productos manufacturados (% de importaciones de mercaderías)	TM.VAL.MANF.ZS.UN	D60
Banco Mundial	Importaciones de combustibles (% de importaciones de mercaderías)	TM.VAL.FUEL.ZS.UN	D62
Banco Mundial	Importaciones de bienes y servicios (% del crecimiento anual)	NE.IMP.GNFS.KD.ZG	D63
Banco Mundial	Gasto nacional bruto (US\$ a precios actuales)	NE.DAB.TOTL.CD	D64
Banco Mundial	Gasto militar (% del PIB)	MS.MIL.XPND.GD.ZS	D65
Banco Mundial	Gasto de consumo final de los hogares (US\$ a precios constantes de 2010)	NE.CON.PRVT.KD	D66
Banco Mundial	Gasto de consumo final (US\$ a precios actuales)	NE.CON.TOTL.CD	D67
Banco Mundial	Formación bruta de capital (% del PIB)	NE.GDI.TOTL.ZS	D68
Banco Mundial	Flujos oficiales netos procedentes de organismos de las Naciones Unidas, PNUD (US\$ a precios actuales)	DT.NFL.UNDP.CD	D69
Banco Mundial	Exportaciones de metales y minerales (% de las exportaciones de mercaderías)	TX.VAL.MMTL.ZS.UN	D70
Banco Mundial	Esperanza de vida al nacer, total (años)	SP.DYN.LE00.IN	D72
Banco Mundial	Emisiones de gases de efecto invernadero totales (kt de equivalente de CO2)	EN.ATM.GHGT.KT.CE	D73
Banco Mundial	Emisiones de CO2 procedentes del consumo de combustibles sólidos (% del total)	EN.ATM.CO2E.SF.ZS	D74
Banco Mundial	Emisiones de CO2 (kt)	EN.ATM.CO2E.KT	D75
Banco Mundial	Emisiones agrícolas de óxido nitroso (miles de toneladas métricas de equivalente de CO2)	EN.ATM.NOXE.AG.KT.CE	D76

Fuente	Nombre de la variable	Código Variable	Cod Modelo
Banco Mundial	Emisiones agrícolas de gas metano (miles de toneladas métricas de equivalente de CO2)	EN.ATM.METH.AG.KT.CE	D77
Banco Mundial	Educación de nivel secundario, duración (años)	SE.SEC.DURS	D78
Banco Mundial	Educación de nivel secundario, alumnos	SE.SEC.ENRL	D79
Banco Mundial	Educación de nivel primario, maestros	SE.PRM.TCHR	D80
Banco Mundial	Densidad de población (personas por kilómetro)	EN.POP.DNST	D81
Banco Mundial	Crecimiento del PIB per cápita (% anual)	NY.GDP.PCAP.KD.ZG	D82
Banco Mundial	Crecimiento del PIB (% anual)	NY.GDP.MKTP.KD.ZG	D83
Banco Mundial	Crecimiento de la población urbana (% anual)	SP.URB.GROW	D85
Banco Mundial	Crecimiento de la población (% anual)	SP.POP.GROW	D86
Banco Mundial	Compañías nacionales que cotizan en Bolsa, total	CM.MKT.LDOM.NO	D88
Banco Mundial	Comercio (% del PIB)	NE.TRD.GNFS.ZS	D89
Banco Mundial	Cargos por el uso de propiedad intelectual, recibos (balanza de pagos, US\$ a precios actuales)	BX.GSR.ROYL.CD	D90
Banco Mundial	Cargos por el uso de propiedad intelectual, pagos (balanza de pagos, US\$ a precios actuales)	BM.GSR.ROYL.CD	D91
Banco Mundial	Capitalización en el mercado de empresas nacionales que cotizan en bolsa (USD a precios actuales)	CM.MKT.LCAP.CD	D92
Banco Mundial	Capitalización en el mercado de empresas nacionales que cotizan en bolsa (% del PIB)	CM.MKT.LCAP.GD.ZS	D93
Banco Mundial	Balanza comercial de bienes y servicios (% del PIB)	NE.RSB.GNFS.ZS	D94
Banco Mundial	Ayuda Oficial al Desarrollo (ODA) neta recibida (% del INB)	DT.ODA.ODAT.GN.ZS	D95
Banco Mundial	Ahorro interno bruto (% del PIB)	NY.GDS.TOTL.ZS	D97
Banco Mundial	Acciones negociadas, valor total (US\$ a precios actuales)	CM.MKT.TRAD.CD	D102
Banco Mundial	Acciones negociadas, valor total (% del PIB)	CM.MKT.TRAD.GD.ZS	D103
Bolsa valores	Acciones negociadas, valor total (% del PIB)	CM.MKT.TRAD.GD.ZS	D104
FAO	Población total (1000 personas)	TT.POB.	D118
FAO	Hombres (1000 personas)	TT.POB.HM	D119
FAO	Mujeres (1000 personas)	TT.POB.MJ	D120
FAO	Población urbana (1000 personas)	TT.POB.URB	D122
Acciones B.M.	Acciones negociadas, total	CM.MKT.TRAD.GD.T	D123

Fuente	Nombre de la variable	Código Variable	Cod Modelo
Acciones B.M.	Acciones negociadas, valor total (US\$ a precios actuales)	CM.MKT.TRAD.CD	D124

Se pueden observar en la Tabla 9 la estructura de los resultados para los tres pilares del sistema sociotécnico aplicables al modelo, que muestran el resumen del agrupamiento según las variables que se asocian a cada uno.

Tabla 9 Estructura de los pilares del modelo

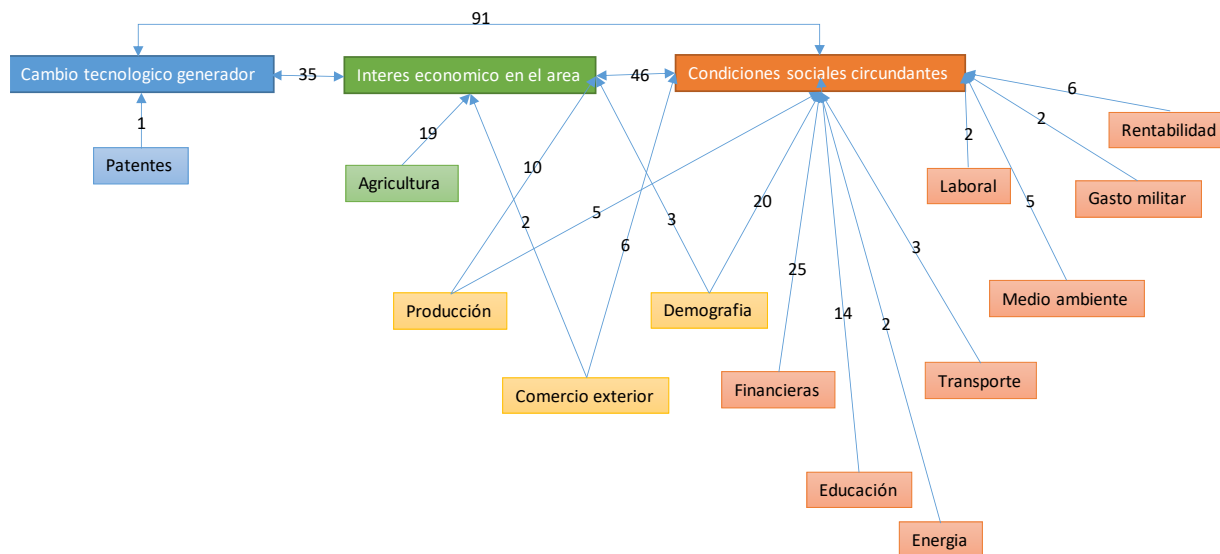
Pilares del sistema	Nombre del Pilar	Variables incluidas	Nivel de aplicación	Fuentes
Pilar A	Cambio tecnológico generador	1	Solo para tecnologías específicas	Organización Mundial de la Propiedad Intelectual - OMPI
Pilar B	Interés económico del área	34	Solo para sectores específicos (puede incluir diferentes tecnologías del mismo sector)	Banco Mundial - BM / Organización de las Naciones Unidas para la Alimentación y la Agricultura - FAO / Federación Mundial Bolsa de valores WFE
Pilar C	Condiciones sociales circundantes	90	Útiles para todas las industrias y tecnologías	Banco Mundial - BM / Organización de las Naciones Unidas para la Alimentación y la Agricultura - FAO / Federación Mundial Bolsa de valores WFE

Retomando la estructura de pilares, sus respectivas áreas de interés y las variables incluidas en cada una, se presenta en la Figura 4.8 la visualización del sistema sociotécnico desarrollado para el modelo de predicción, en el cual cada pilar tiene asociadas áreas y las líneas de conexión muestran la cantidad de variables que soportan esa asociación, para el caso específico del pilar A (en color azul) se presenta una sola área y una sola variable, mientras para el pilar B, se presentan cuatro áreas, una de ellas específica (agricultura) con 19 variables y tres áreas compartidas con el pilar C, estas áreas (producción, comercio exterior y demografía), tienen dentro de su grupo de variables diferentes pilares de atención, por ejemplo, demografía participa con tres variables en el

pilar B y con 20 variables en el pilar C, finalmente tributarias del interés específico aportante al pilar C.

Figura 4.8

Sistema sociotécnico de las categorías de variables incluidas en el modelo



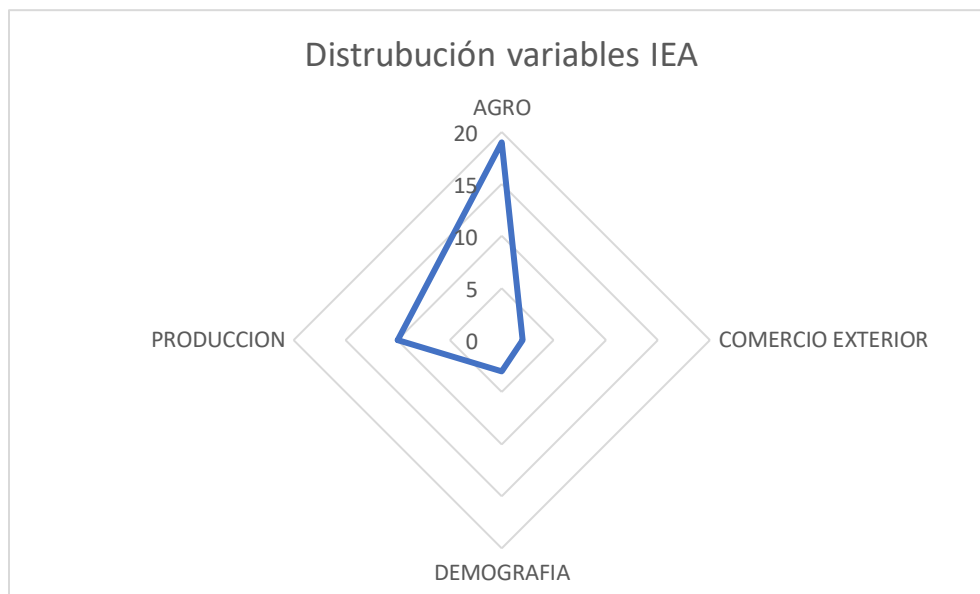
Una interesante relación se presenta entre las uniones de los pilares, donde las variables de uno tienen relaciones directas con las variables de otro, cómo se presenta en la Tabla 10, y también se puede visualizar en las líneas de conexión superiores de la Figura 4.8, en las cuales se evidencian las relaciones internas de las variables dentro del sistema sociotécnico.

Tabla 10 Variables relacionadas entre pilares

Relación	Variables
A+B	35
B+C	46
C+A	91
A+B+C	125

En específico sobre el pilar C y sus cuatro áreas de interés, se presenta la Figura 4.9, en la cual se observa la potencia relativa del peso de las variables en cada una de las áreas, no presentando sorpresa que el área de agro contenga la mayor presencia de variables, puesto que el foco de la investigación se centra en ese mismo sector económico.

Figura 4.9 Radar de áreas en el pilar B



Frente a las áreas de interés y las variables relacionadas a estas, de las condiciones sociales circundantes del pilar C, dónde se ubican el 72% de las variables totales del sistema sociotécnico, las cuales según se mencionó son el contexto general, aplicable a todas las tecnologías y por ende presentan una distribución de radar con múltiples picos, como se observa en la Figura 4.10 presentando énfasis en las áreas financiera, demografía y educación.

El resumen del segundo y tercer nivel de análisis de la estructura del sistema sociotécnico se presenta en la Figura 4.11, por medio de una distribución de radar general, donde claramente se ven los énfasis temáticos de las variables, especialmente para el área de demografía, financiera producción y agro, esta última explicada por la propia área de investigación y el desarrollo del

modelo de predicción, como se mencionó anteriormente.

Figura 4.10 Radar de áreas en el pilar C

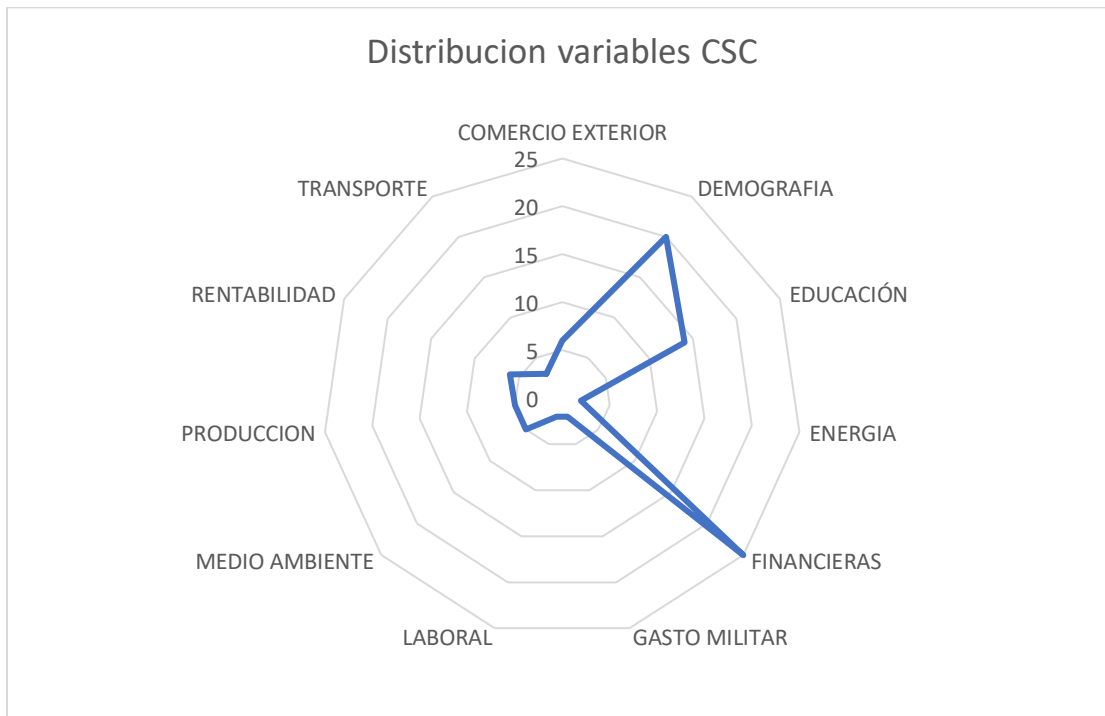
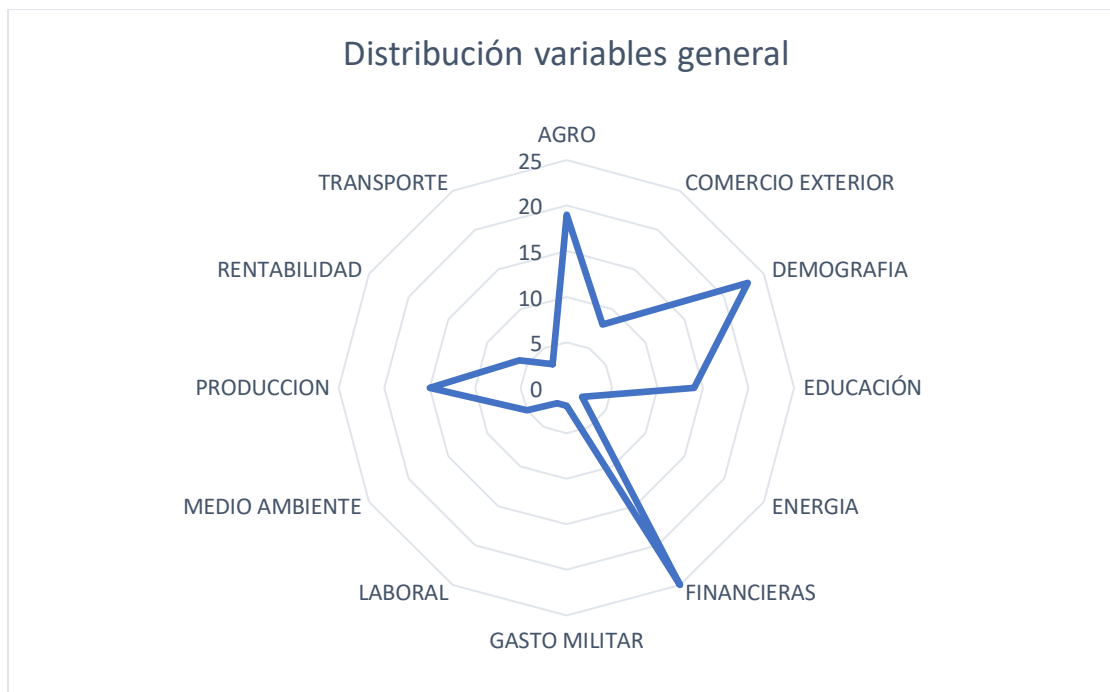


Figura 4.11 Radar general de áreas y variables en el modelo



La configuración final del sistema sociotécnico responde a la disponibilidad de datos según las reglas y condiciones establecidas para el tratamiento de estos, existe la probabilidad de extender a un número mayor de las 547 variables identificadas inicialmente y se logre obtener una matriz final con más de 125 variables y un rango temporal mayor a 50 años, en la medida que la calidad de los datos mundiales disponibles mejore y los repositorios de información permitan el acceso.

Algunas variables muy interesantes fue necesario dejarlas afuera, toda vez que su medición está atada a condiciones regionales que no permiten su extrapolación a nivel mundial, otras variables de interés, aunque deseables no se encuentran estandarizadas ni cuentan datos para todo el periodo de tiempo, por otra parte la ausencia de mediciones sobre morbilidad, alimentación, salud mental, criminalidad entre otras, que cumplan los criterios establecidos, genera un espacio de mejora futura que permita ampliar la base de información en cada pilar, esta condición ideal supera las posibilidades de esta tesis, pero abre la puerta para posteriores desarrollos.

5 Métodos Para el Modelo de Predicción

Basado en la estructura del sistema sociotécnico presentada, en especial la configuración de patentes como variable a predecir, la que actúa como indicador de la evolución tecnológica y soportada sobre un conjunto multivariado de 124 arreglos de datos, se presentan tres métodos de predicción, todos ellos ejecutados sobre lenguaje de programación Python versión 3.7.15, en la plataforma Google Colab, correspondientes a aplicaciones de aprendizaje automático de máquina, siendo estos: red LSTM, ARIMA y Bosques Aleatorios.

La teoría de cambio tecnológico de Schumpeter (1935b) ha sido evaluada usando patentes para medir la evolución tecnológica, encontrando una contradicción con los ciclos de Kondratieff (Korotayev et al., 2011), se evidenció que existe un rezago en la función del ciclo propuesto por Kondratieff sobre el modelo de Schumpeter, pues la mayor cantidad de invenciones (patentes) se producen en la época de recesión y no se ven aprovechadas hasta la recuperación y bonanza. La evidencia presentada por Korotayev demuestra que la reducción de patentes en la bonanza se debe a que los cambios son abruptos y no incrementales como en la recesión, evidenciando relaciones entre variables.

La aplicación de redes neuronales para mejorar el proceso de predicción ha obtenido resultados positivos al tomar como base de información las patentes (Altuntas et al., 2015; Chang & Fan, 2015, 2016; Gao et al., 2013; Jun et al., 2012; C. Lee et al., 2017; Leydesdorff, 2015; Ramadhan et al., 2018; Madvar et al., 2016), en este sentido se puede establecer que los modelos de predicción actúan basados en información anterior, con entrada principal en las patentes y posibilitando una visión de futuro cercana a la realidad.

Profundizando en el cambio incremental, se evidencia un rezago en los incrementos tecnológicos, ya que las empresas mantienen la atención en los componentes clave aún después de que una tecnología dominante se imponga (J. Lee & Berente, 2013), dando pie al uso de desarrollos acumulativos o cambios incrementales como una buena medida de predicción, para el cambio tecnológico (Huenteler, Schmidt, Ossenbrink, & Hoffmann, 2016).

Fundado en las investigaciones y desarrollos anteriores, que incluyen el uso de patentes para estudiar el cambio tecnológico, instrumentalizados por medio inteligencia artificial, específicamente aprendizaje supervisado de máquina (Oliveira et al., 2021; Smith & Agrawal, 2015), se procede a explorar los métodos propuestos, incluyendo adaptaciones y ajustes tanto a los modelos como a los datos, aplicados individualmente y luego en configuración multietapa (Chen & Klasky, 2022), buscando la mejor combinación para la predicción del cambio incremental.

El primer modelo explorado es la red LSTM (*Long Sort Term Memory* -Memoria de Corto y Largo Plazo), seguido se explora el modelo ARIMA (*Autoregressive Integrated Moving Average* - Modelo Autorregresivo Integrado de Promedio Móvil) y finalmente el Modelo de Bosques Aleatorios, sobre los resultados de la primera ronda de exploración se efectúan ajustes a los modelos, realizando rondas sucesivas de optimización en especial en la red LSTM, una vez optimizado el modelo se prueban diferentes configuraciones del sistema sociotécnico con el fin de comprobar su precisión predictiva (Jha et al., 2021) y seleccionar la configuración que mejores resultados genera según la estructura de datos.

5.1 Modelo 1 – red LSTM

Dentro de las ramas de la inteligencia artificial se encuentran las redes neuronales, a su vez dentro de ellas se ubican algoritmos de aprendizaje supervisado, destacando las redes LSTM usadas ampliamente en modelos de predicción en series de tiempo. Las características más notables son la identificación de datos relevantes para predicción y desecho de otros que no generan aportes (Vora et al., 2022). Ante la estructura del sistema sociotécnico donde se encuentran 125 variables, es pertinente el uso de redes LSTM con el fin de evaluar su desempeño y capacidad predictiva.

Más específicamente las redes LSTM son un tipo especial de red neuronal recurrente, dotadas de memoria interna, con capacidad de aprender a filtrar y propagar información importante a través del tiempo. Siendo muy útiles para la predicción de series temporales, ya que pueden tener en cuenta la información relevante del pasado para hacer predicciones precisas por medio de optimizaciones (Liu & Zhao, 2022).

Una red LSTM está compuesta por una serie de celdas LSTM, cada una de las cuales tiene tres puertas que permiten a la información entrar o salir de la celda, estas puertas son:

- Una puerta de olvido, que controla qué información se olvida o se mantiene en la memoria de la celda.
- Una puerta de entrada, que controla qué información se almacena en la memoria de la celda.
- Una puerta de salida, que controla qué información se propaga a las celdas LSTM siguientes.

Las celdas LSTM se conectan entre sí en la serie temporal, lo que permite que la red recuerde y utilice información de entradas anteriores cuando procesa nuevas entradas.

La red LSTM explorada es de tipo multivariable, basada en la red clásica LSTM pero entrenada para aprender a tomar en cuenta múltiples variables relevantes al mismo tiempo, esto se logra conectando varias entradas a la red LSTM, cada una correspondiente a una de las 125 variables de la serie temporal, y entrenando la red para que aprenda a procesar todas las entradas juntas, olvidar las variables que no aportan en la predicción y mantener las que si lo hacen.

5.1.1 Definición de parámetros en el modelo LSTM

Según se presentó en los capítulos anteriores, la variable base de predicción son las patentes, las que se presentan en la matriz de trabajo en su forma de frecuencia absoluta, siguiendo el método de identificación de la etapa del ciclo de vida de la tecnología, donde se usa la frecuencia acumulada para la generación de la curva sigmoideal de ciclo de vida, es necesario calcular esta frecuencia para aplicarla en los métodos de predicción explorados, e incluir el resultado en la matriz de trabajo.

El cálculo de la frecuencia absoluta acumulada- FAA se realiza por la aplicación de la Ecuación 2, en la cual (N_i) representa la FAA de un valor X_i del conjunto de datos (X_1, X_2, \dots, X_N), para este caso cada X_i representa las patentes de un año específico, siendo la suma de las frecuencias absolutas de los valores menores o iguales a cada X_i ,

Ecuación 2. Frecuencia absoluta acumulada.

$$N_i = n_1 + n_2 + \dots + n_i$$

Las ventanas de observación corresponden a grupos de datos en un rango de cinco años, que se inician en 1970 y se mueven en pasos de un año a la vez, desechando el año más bajo e

incluyendo los datos de un nuevo año superior, como se presenta en la Tabla 11, por lo cual se obtienen 45 ventanas dentro del periodo establecido, limitándose por el último año conocido (2019) dado que la ventana siguiente (46) implica una predicción para 2020, la cual aún no se ha calculado.

Tabla 11 Ventanas de observación

Rangos	Año 1	Año 2	Año 3	Año 4	Año 5
Ventana 1	1970	1971	1972	1973	1974
Ventana 2	1971	1972	1973	1974	1975
Ventana 3	1972	1973	1974	1975	1976
Ventana n
Ventana 45	2015	2016	2017	2018	2019

5.1.2 Pasos del modelo LSTM

Paso 1: Importación de librerías necesarias para el desarrollo del modelo y carga de la matriz base de trabajo, esta se encuentra en archivo tipo hoja de cálculo, cargado en Google Drive.

```
import tensorflow as tf
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
np.random.seed(50)

from google.colab import drive
drive.mount('/content/drive')
```

Paso 2: Una vez cargados los datos se realiza transformación de la primera columna de fecha al formato *datetime*, para ser leída correctamente por el programa y se eliminan las columnas iniciales que contenían los datos de fecha y de frecuencia absoluta de la variable a predecir (*WIPO1*).

```
data.index = pd.to_datetime(data['Fecha'], format='%d.%m.%Y')
data.drop(columns=['Fecha', 'WIP01'], inplace=True)
```

Paso 3: Verificación de la forma de los datos y comprobación de existencia de datos faltantes en la matriz, permitiendo la verificación de la correcta carga de datos.

```
data.head()
data.shape
print(f'Cantidad de datos faltantes: {data.isna().sum().sum()}' )
```

Paso 4: Revisión de las principales características de los datos: cuenta, media, desviación estándar, mínimos y máximos, un extracto de los datos de salida se presenta Tabla 12, con esta información se verifica la consistencia de los datos.

```
data.describe()
```

Tabla 12 Extracto características de los datos

	acumu	D1	D2	D3	...	D124
Count	50	50	50	50	...	50
Mean	67.086,48	59,0022	1.619,3720	1,9230E+13	...	2,7993E+19
Std	43.847,83	47,8553	193,0724	8,9102E+12	...	3,1265E+19
Min	914,00	6,0616	1.337,6120	9,2447E+12	...	2,9807E+17
0,25	32.753,25	10,8912	1.426,8840	1,0718E+13	...	1,0187E+18
0,5	64.176,50	42,8905	1.622,5883	1,7614E+13	...	9,3085E+18
0,75	94.113,75	99,9965	1.796,5374	2,5331E+13	...	5,8143E+19
Max	173.874E+11	161,1660	1.919,9918	3,9204E+13	...	1,0117E+20

Paso 5: Asignación como variable indexadora el año, eliminando información de mes y día por no ser relevantes para el modelo.

```
data.index = data.index.year
```

Paso 6: Partición de los datos en dos grupos, también llamados *sub-sets* de datos:

- Entrenamiento: 45 primeros años (*data_train*)
- Prueba: 5 últimos años (*data_test*)

Para volúmenes de datos mayores, es posible dividir el *set* en tres partes, entrenamiento, validación y prueba, pero ante un horizonte de 50 años, es necesario mantener el modelo en su mínima complejidad.

```
N = data.shape[0]
data_train = data[0:45]
data_test = data[45:]
```

Paso 7: Estandarización de los datos por medio de normalización, para ello se calcula, el promedio (*train_mean*) y la desviación estándar (*train_std*) de cada una de las variables del *set* de entrenamiento.

```
train_mean = data_train.mean()
train_std = data_train.std()
```

Paso 8: Verificación de variables con desviación igual a cero, necesaria para evitar divisiones entre cero (0) dentro de las celdas LSTM.

```
train_std[train_std==0]
```

La variable denominada D78 presenta desviación estándar igual a cero (0). por lo cual se elimina del modelo, quedando este con 124 variables.

```
data_train = data_train.drop(columns=['D78'])
data_test = data_test.drop(columns=['D78'])
```

Paso 9: Una vez depuradas las variables a usar, se procede con la normalización, en cada uno de los sub-sets o grupos de datos en el modelo.

```
data_train = (data_train - train_mean) / train_std
data_test = (data_test - train_mean) / train_std
```

Paso 10: Presentación grafica de los datos normalizados en diagramas de violín, para ellos se subdividen las variables en grupos que faciliten su visualización, evidenciando la distribución y la densidad de probabilidad (McKinney, 2017).

```
data_std = (data - train_mean) / train_std

columns = data_std.columns
cols_group1 = columns[0:25]
cols_group2 = columns[25:50]
cols_group3 = columns[50:75]
cols_group4 = columns[75:100]
cols_group5 = columns[100:]

fig, ax1 = plt.subplots(figsize=(15,3))
sns.violinplot(ax=ax1, data=data_std[cols_group1]);
```

Los diagramas generados se presentan en la Figura 5.1 incluyendo la variable *acumu* (frecuencia absoluta acumulada) y las variables D1 a D24, en Figura 5.2 se presentan las variables D25 a D49, la Figura 5.3 contiene las variables D50 a D74 , la Figura 5.4 muestras las variables D75 a D100, exceptuando la variable D78 eliminada anteriormente, la Figura 5.5 completa la visualización con las variables D10 a D124.

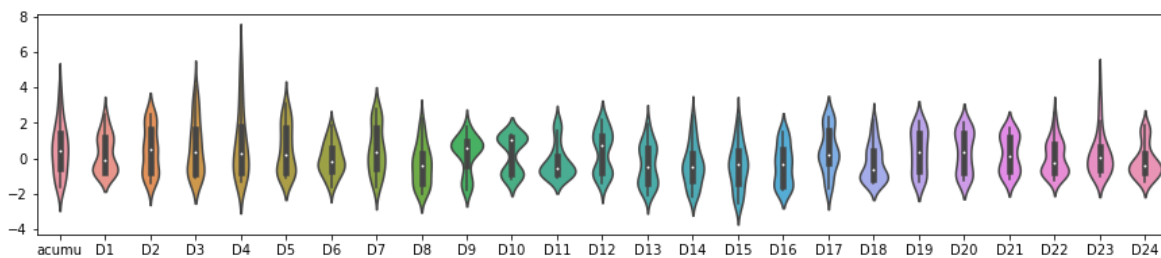
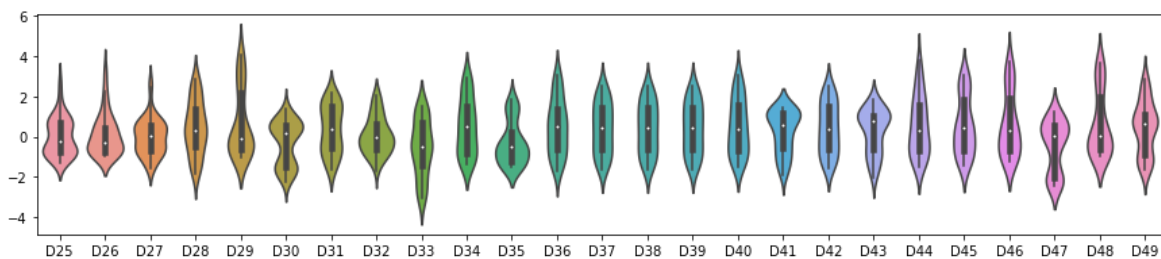
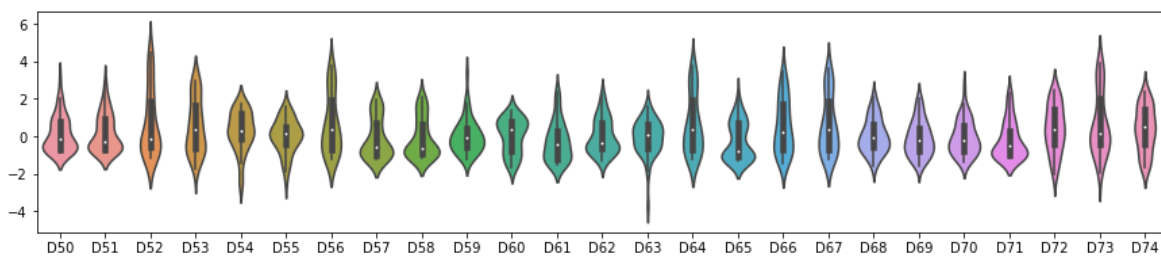
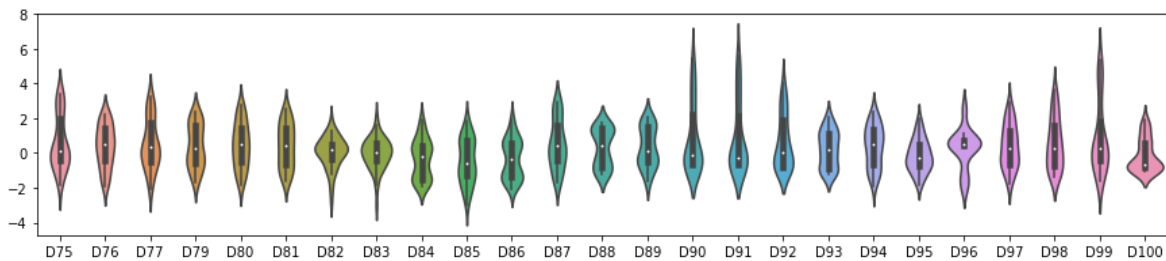
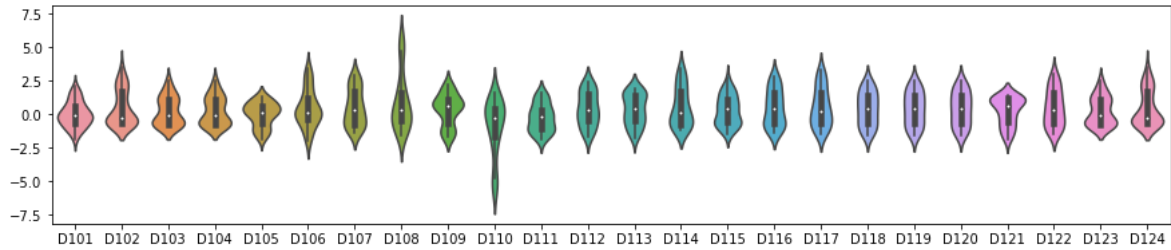
Figura 5.1 Distribución de las variables D1 A D24**Figura 5.2** Distribución de las variables D25 A D49**Figura 5.3** Distribución de las variables D50 A D74**Figura 5.4** Distribución de las variables D75 A D100

Figura 5.5 Distribución de las variables D101 A D124

Paso 11: Creación de ventanas, según se indicó, el modelo de predicción es del tipo t+1 (predice el siguiente periodo de tiempo), usando pasos simples para la misma frecuencia temporal (anual), el tamaño asignado a las ventanas es de cinco años, tomando los primeros cuatro años para predecir el quinto, luego desplazando la ventana un paso en el tiempo y repitiendo el cálculo nuevamente.

```
input_width: 4 (cantidad de años consecutivos requeridos)
label_width: 1 (duración en años de la predicción)
Shift: 1      (separación entre dos ventajas consecutivas)
```

```
class WindowGenerator:
    def __init__(self, train_df, val_df, test_df, input_width=4, label_width=1, shift=1, label_columns=['acumu']):
```

```
self.input_width = input_width
self.label_width = label_width
self.shift = shift
self.total_window_size = input_width + shift

self.train_df = train_df
self.test_df = test_df

self.label_columns = label_columns
self.label_columns_indices = {name: i for i, name in
    enumerate(self.label_columns)}
self.column_indices = {name: i for i, name in
    enumerate(self.train_df.columns)}
self.input_slice = slice(0, self.input_width)
self.input_indices =
    np.arange(self.total_window_size)[self.input_slice]
```

```

self.label_start = self.total_window_size -
self.label_width
self.labels_slice = slice(self.label_start, None)
self.label_indices =
np.arange(self.total_window_size)[self.labels_slice]

```

La estructura de las ventanas tiene dos contenedores temporales de datos (entrenamiento y prueba) que son usados en cada iteración para predecir y ajustar el modelo, estos contenedores mantienen la estructura de cuatro datos previos para predecir el quinto y compararlo con el valor real, esta forma se confirma mediante la verificación de creación de la estructura de datos.

```

{self.total_window_size}
{self.input_indices}
{self.label_indices}
{self.label_columns}
w1 = WindowGenerator(data_train, data_test, label_columns=['
acumu'])
w1

```

La configuración de la creación de ventanas presenta el siguiente arreglo:

- Tamaño total de la ventana: 5
- Ventana - índices de entrada: [0 1 2 3]
- Ventana - índices de salida: [4]
- Nombre de la columna a predecir: ['acumu']

Paso 12: Uso del método *Split* para lograr particiones dinámicas del grupo de datos dentro de cada contenedor, desplazando la ventana, un periodo a la vez (Garcia et al., 2018).

```

def split_window(self, features):
    inputs = features[:, self.input_slice, :]
    labels = features[:, self.labels_slice, :]
    if self.label_columns is not None:
        labels = tf.stack([labels[:, :, self.column_index
s[name]] for name in self.label_columns], axis=-1)

```

Con esta configuración, el modelo toma como entrada para cada año de la ventana las 124 variables: una a predecir (*acumu*) y 123 relacionadas, como salida predice el valor de la variable *acumu* en el año cinco, la forma general del modelo tiene esta disposición:

$t = 0$ (1970): [*acumu*, D1, D2, ..., D124] (1x124)

$t = 1$ (1971): [*acumu*, D1, D2, ..., D124] (1x124)

$t = 2$ (1972): [*acumu*, D1, D2, ..., D124] (1x124)

$t = 3$ (1973): [*acumu*, D1, D2, ..., D124] (1x124)

Predicción de salida:

$t = 4$ (1974): [*acumu*] (1x1)

En resumen, la ventana contiene cinco años, cada uno de ellos incluye a su vez 124 variables, los primeros cuatro años se usan para determinar el arreglo en que 123 variables se relacionan con una de ellas, la 124 (*acumu*), este proceso se realiza en cuatro rondas (4 años) ajustando el arreglo en cada ronda, la salida es la predicción de una única variable (*acumu*) en el año cinco.

Paso 13: Comprobación gráfica del comportamiento de las ventanas y su predicción, tomado como ejemplo la ventana inicial, se presentan los valores de entrada de la variable *acumu*, de los primeros cuatro años (puntos y líneas color azul) y la predicción del quinto año, marcado con un punto color verde, como se observa en la Figura 5.6, evidenciando el funcionamiento de las ventanas y su predicción.

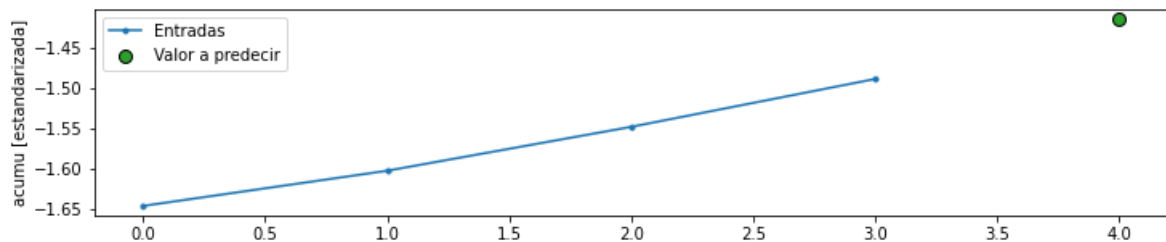
```

w1.example = example_inputs, example_labels
def plot(self, model=None, plot_col='acumu',
max_subplots=3):
    inputs, labels = self.example
    plt.figure(figsize=(12, 8))
    plot_col_index = self.column_indices[plot_col]
    max_n = min(max_subplots, len(inputs))
    for n in range(max_n):
        plt.subplot(max_n, 1, n+1)
        plt.ylabel(f'{plot_col} [estandarizada]')
        plt.plot(self.input_indices, inputs[n, :,
plot_col_index], label='Entradas', marker='.', zorder=-10)
        if self.label_columns:
            label_col_index =
self.label_columns_indices.get(plot_col, None)
        else:
            label_col_index = plot_col_index
        if label_col_index is None:
            continue
        plt.scatter(self.label_indices, labels[n, :,
label_col_index], edgecolors='k', label='Valor a predecir',
c='#2ca02c', s=64)
        if model is not None:
            predictions = model(inputs)
            plt.scatter(self.label_indices, predictions[n, :,
label_col_index], marker='X', edgecolors='k',
label='Predictions',
                        c='#ff7f0e', s=64)

        if n == 0:
            plt.legend()
            plt.xlabel('Año')
WindowGenerator.plot = plot

```

Figura 5.6 Comportamiento de las ventas de predicción



Paso 14: Transformación de los datos al formato *dataset*, para ser leídos por la biblioteca *Tensorflow* de Python, tanto para los datos de entrenamiento, como para los de prueba (Gerón, 2019).

```
def make_dataset(self, data):
    data = np.array(data, dtype=np.float32)
    ds = tf.keras.utils.timeseries_dataset_from_array(
        data=data,
        targets=None,
        sequence_length=self.total_window_size,
        sequence_stride=1,
        shuffle=True,
        batch_size=32,)

    ds = ds.map(self.split_window)

WindowGenerator.make_dataset = make_dataset

@property
def train(self):
    return self.make_dataset(self.train_df)

@property
def test(self):
    return self.make_dataset(self.test_df)

@property
def example(self):
    result = getattr(self, '_example', None)
    if result is None:
        result = next(iter(self.train))
        self._example = result
    return result

WindowGenerator.train = train
WindowGenerator.test = test
WindowGenerator.example = example
```

Paso 15: Creación del modelo LSTM, para ello inicialmente se seleccionan los valores para los hiper-parámetros, entendidos como los puntos de ajuste de las condiciones del modelo:

- Número de neuronas (*nunits*)
- Numero de rondas de entrenamiento (*epoch*)

Los valores iniciales seleccionados son 16 neuronas y 20 rondas de entrenamiento, usando como función de perdida el error cuadrático medio y el optimizador Adam.

```
NUNITS = 16
lstm_model = tf.keras.models.Sequential([
    tf.keras.layers.LSTM(NUNITS, return_sequences=False),
    tf.keras.layers.Dense(units=1)])
MAX_EPOCHS = 20
```

```
def compile_and_fit(model, window):
    model.compile(loss=tf.keras.losses.MeanSquaredError(reduction=tf.keras.losses.Reduction.SUM), optimizer=tf.keras.optimizers.Adam(), metrics=[tf.keras.metrics.MeanAbsoluteError()])

    history = model.fit(window.train, epochs=MAX_EPOCHS, validation_data = window.val)
    return history

ventana = WindowGenerator(data_train, data_test,
    input_width=4, label_width=1, shift=1,
    label_columns=['acumu'])

history = compile_and_fit(lstm_model, ventana)
```

Paso 16: Ejecutar el modelo, en este punto el modelo realiza tantas rondas de entrenamiento como se asignaron y en cada una de ellas calcula: la función de perdida, el error absoluto medio, la perdida de la validación cruzada y el error absoluto medio de la validación cruzada. Este proceso de entrenamiento busca ajustar el modelo, evidenciado con la reducción constante de las funciones arriba nombradas.

```
Epoch 1/50
```

```

2/2 [=====] - 3s 570ms/step -
loss: 27.7570 - mean_absolute_error: 0.7647 - val_loss:
38.5391 - val_mean_absolute_error: 3.0949
Epoch 2/50
2/2 [=====] - 0s 39ms/step - loss:
17.0113 - mean_absolute_error: 0.6059 - val_loss: 29.5854 -
val_mean_absolute_error: 2.7107
Epoch 3/50
2/2 [=====] - 0s 44ms/step - loss:
12.1935 - mean_absolute_error: 0.4819 - val_loss: 22.8902 -
val_mean_absolute_error: 2.3813
Epoch 4/50
2/2 [=====] - 0s 39ms/step - loss:
6.9741 - mean_absolute_error: 0.3648 - val_loss: 18.0458 -
val_mean_absolute_error: 2.1114
Epoch 5/50
2/2 [=====] - 0s 57ms/step - loss:
4.9852 - mean_absolute_error: 0.2912 - val_loss: 14.1831 -
val_mean_absolute_error: 1.8690
Epoch 6/50
2/2 [=====] - 0s 41ms/step - loss:
2.9217 - mean_absolute_error: 0.2312 - val_loss: 11.5096 -
val_mean_absolute_error: 1.6800
Epoch 7/50
2/2 [=====] - 0s 37ms/step - loss:
1.7038 - mean_absolute_error: 0.1999 - val_loss: 9.8170 -
val_mean_absolute_error: 1.5466
Epoch 8/50
2/2 [=====] - 0s 39ms/step - loss:
1.3297 - mean_absolute_error: 0.1791 - val_loss: 8.9274 -
val_mean_absolute_error: 1.4695
Epoch 9/50
2/2 [=====] - 0s 59ms/step - loss:
0.8437 - mean_absolute_error: 0.1633 - val_loss: 8.5230 -
val_mean_absolute_error: 1.4309

```

Paso 17: Verificar el desempeño del modelo, se ejecuta prueba basada en el error cuadrático medio y se imprimen por pantalla los resultados.

```

train_performance = lstm_model.evaluate(ventana.train)
test_performance = lstm_model.evaluate(ventana.test,
verbose=1)
print(f'Desempeño entrenamiento: {train_performance[1]}')
print(f'Desempeño prueba: {test_performance[1]}')

```

- Desempeño entrenamiento: 0.06602973490953445
- Desempeño prueba: 2.815837860107422

Encontrando que el desempeño en el entrenamiento es bueno (inferior a 1), pero se reduce su efectividad en la prueba, denotando sobre ajuste (*overfitting*) por lo cual el modelo aún no tiene la configuración óptima de los parámetros e hiper-parámetros.

Paso 18: Predicción del valor de la variable *acumu* para el siguiente año (2020), como parte del proceso es necesario invertir la normalización para obtener el valor de la predicción en el mismo rango de la variable original (Muller & Guido, 2016).

```
def predecir(model, window):
    pred = model.predict(ventana.test, verbose=0)[0][0]

    pred = pred*train_std['acumu'] + train_mean['acumu']
    print(f'Resultado de la predicción (acumu): {pred}')

predecir(lstm_model, ventana)
```

Como resultado del modelo predijo un valor de la variable *acumu* para el año 51 (2020) de 111.868,20 en el entendido que el ultimo valor conocido es el año 50 (2019) de 173.874, se presenta una diferencia negativa de 62.006.

Dado que uno de los principios de la curva sigmoideal del ciclo de vida es la no existencia de desacumulación, se concluye que el modelo funciona, pero no tiene el ajuste necesario para obtener predicciones certeras aun, por lo cual se requieren nuevas rondas de optimización y el uso de la frecuencia absoluta como variable a predecir.

5.2 Modelo 2 – ARIMA

El modelo ARIMA es presentado en 1970 por los estadísticos George Edward Pelham Box, Gwilym Meirion Jenkins y Gregory C Reinsel, también conocido como el modelo Box-Jenkins, ARIMA es un acrónimo de tres componentes AR (Autoregresivo), I(Integrado) y MA (Medias Móviles), en español: Promedio Móvil Integrado Autoregresivo, o en su idioma original *Autoregressive Integrated Moving Average* (Box & Jenkins, 1970).

El modelo se usa para comprender el comportamiento de series temporales y pronosticar su evolución futura (You et al., 2017), el uso de ARIMA requiere de pasos específicos:

1. Determinar si los datos son estacionarios o no (estacionarios son aquellos con media, varianza y autocorrelación constantes), si son estacionarios se continua, en caso contrario se debe aplicar transformación o suavizado.
2. Determinar la tendencia por grafico o correlación
3. Elegir los parámetros del modelo: P (número de términos autorregresivos), D (número de diferencias necesarias para convertir los datos en estacionarios) y Q (número de términos de la media móvil)
4. Entrenar el modelo utilizando los datos y los parámetros seleccionados.
5. Hacer el pronóstico de los datos según el modelo entrenado.

Este modelo reviste complejidad y un proceso de tratamiento de datos en varias etapas, por lo cual se usa la librería *pmarima* para Python que automatiza parte del proceso, permitiendo centrarse en el ajuste de hiper-parámetros y la optimización del modelo (Hingley & Dikta, 2022).

La implementación del modelo continua con la estructura de pasos usado anteriormente.

Paso 1: Importación de librerías necesarias y carga de datos en Google Colab desde Google Drive, de la misma manera indicada en el modelo LSTM, manteniendo la estructura de 123 variables predictoras, obtenida después de eliminar la variable D78 según se indicó y una variable a predecir: *acumu* (frecuencia absoluta acumulada de las patentes).

```
import pandas as pd
from google.colab import drive
drive.mount('/content/drive')
```

Paso 2: Eliminación de variables no necesarias (frecuencia absoluta – *WIPO1* y variable D78) y transformación de la fecha en formato *datetime*.

```
data.drop(columns=['WIPO1'], ['D78'] inplace=True)

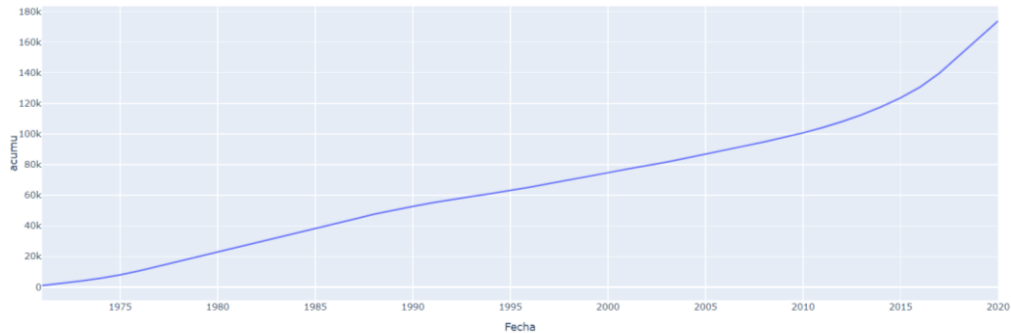
data['Fecha'] = pd.to_datetime(data['Fecha'],
format='%d.%m.%Y')
```

Paso 3: Graficar la variable a predecir (*acumu*) que servirá de referencia para el desarrollo del modelo, el resultado se presenta en la

Figura 5.7, donde se puede observar que la serie no presenta estacionalidad, al no tener repeticiones periódicas, pero si presenta tendencia positiva, cumpliendo con los primeros requisitos del modelo ARIMA.

```
import plotly.express as px
fig = px.line(data, x='Fecha', y='acumu')
fig.show()
```

Figura 5.7 Comportamiento de la variable a predecir



Paso 4: Asignación de la variable fecha como índice del modelo.

```
df = data.set_index('Fecha')
```

Paso 5: Instalación e importación de la librería *pmdarima*, automatizadora del proceso de búsqueda de la mejor combinación posible de hiper-parámetros,

```
!pip install pmdarima
import pmdarima as pm
```

Paso 6: Inicialización del modelo en busca de los hiper-parámetros P, D y Q, en este mismo paso se confirma la inexistencia de estacionalidad, así como la periodicidad de los datos en uno ($M=1$), haciendo referencia a lapsos anuales.

```
M = 1
SEASONAL = False
START_P = 0
START_Q = 0
STATIONARY = False
```

```

modelo = pm.auto_arima(df['acumu'], m=M, seasonal=SEASONAL,
                       start_p=START_P, start_q=START_Q, max_order=4,
                       stationary = STATIONARY, test='adf', error_action='
                       ignore', suppress_warnings=True, stepwise=True, tr
                       ace=True)

```

Paso 7: Ejecutar la búsqueda automática de los hiper-parámetros.

```

Performing stepwise search to minimize aic
ARIMA(0,2,0)(0,0,0)[0] intercept : AIC=739.671, Time=0.04 sec
ARIMA(1,2,0)(0,0,0)[0] intercept : AIC=722.388, Time=0.02 sec
ARIMA(0,2,1)(0,0,0)[0] intercept : AIC=726.827, Time=0.07 sec
ARIMA(0,2,0)(0,0,0)[0]          : AIC=745.025, Time=0.01 sec
ARIMA(2,2,0)(0,0,0)[0] intercept : AIC=724.363, Time=0.06 sec
ARIMA(1,2,1)(0,0,0)[0] intercept : AIC=724.359, Time=0.07 sec
ARIMA(2,2,1)(0,0,0)[0] intercept : AIC=726.329, Time=0.10 sec
ARIMA(1,2,0)(0,0,0)[0]          : AIC=722.123, Time=0.02 sec
ARIMA(2,2,0)(0,0,0)[0]          : AIC=723.982, Time=0.02 sec
ARIMA(1,2,1)(0,0,0)[0]          : AIC=723.891, Time=0.04 sec
ARIMA(0,2,1)(0,0,0)[0]          : AIC=729.611, Time=0.04 sec

```

```

Best model: ARIMA(1,2,0)(0,0,0)[0]
Total fit time: 0.645 seconds

```

Como resultado se obtiene el mejor *set* de hiper-parámetros:

- P = 1
- D = 2 (la serie es no estacionaria ni presenta estacionalidad)
- Q = 0

Paso 7: División de los datos en *set* en entrenamiento, los primeros 45 años y *set* de prueba, los últimos cinco años, ambos conteniendo las 124 variables del sistema sociotécnico.

```

df_train = df[:45]
df_test = df[45:]
print(df_train.shape)
print(df_test.shape)

```

Paso 8: Entrenar el modelo ARIMA, según los hiper-parámetros definidos, el tamaño de los *sets* de datos y predecir el comportamiento de la variable, asignado el tiempo de predicción $n_anos = 5$ (predecir 5 años) que incluye los intervalos de confianza de la predicción.

```
modelo.fit(df_train['acumu'])
n_anos = 5
predicciones = modelo.predict(n_periods=n_anos, return_conf_int=True)
```

El primer grupo de datos corresponde al valor de la variable *acumu* predicho para los cinco años (2015-2019), el segundo grupo de datos son los intervalos de confianza para el mismo periodo.

```
(2015-12-31 129799.414986
 2016-12-31 136278.843699
 2017-12-31 142908.475361
 2018-12-31 149630.564631
 2019-12-31 156409.5662944

Freq: A-DEC, dtype: float64, array

([[129340.9420519, 130257.88791949],
 [134995.02876080, 137562.65863815],
 [140435.55628586, 145381.39443669],
 [145637.70678311, 153623.42247886],
 [150598.48537043, 162220.64721765]]))
```

Paso 9: Almacenamiento de las predicciones como un *dataframe*, para lograr la graficación posterior.

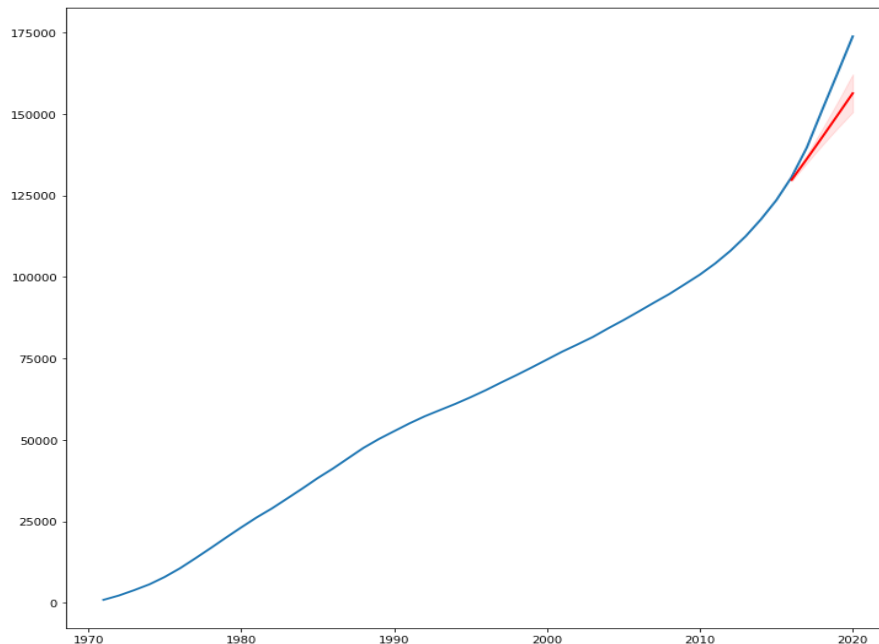
```
df_pred = pd.DataFrame(predicciones[0], index = df_test.index, columns=['Predicciones'])
df_pred = predicciones[0]
df_pred
```

Paso 10: Generación de gráfico de la curva original y la predicción, incluyendo sus intervalos de confianza.

```
import numpy as np
import matplotlib.pyplot as plt
intervalos_array = np.asarray(predicciones[1])
intervalo_inf = intervalos_array[:,0]
intervalo_sup = intervalos_array[:,1]
curva_referencia = df['acumu']
curva_prediccion = df_pred
plt.figure(figsize=(12,12))
plt.plot(curva_referencia, linewidth=2)
plt.plot(curva_prediccion, linewidth=2, color='r')
plt.fill_between(curva_prediccion.index, intervalo_inf, intervalo_sup, color='r', alpha=0.1)
```

En la Figura 5.8, se presenta el resultado de la predicción realizada por el modelo ARIMA, representando con la línea color azul el comportamiento observado de la variable *acumu*, con la línea color rojo la predicción del modelo y sombreado en color rosa, los intervalos de confianza de la predicción, evidenciando que el modelo funciona pese a que la precisión necesita ajustarse.

Figura 5.8 Primera predicción ARIMA



Paso 11: Calcular el error absoluto de la predicción en cada año y el error absoluto medio, según se presenta en la Tabla 13, como elemento de comparación para los posteriores ajustes al modelo,

Tabla 13 Calculo de error absoluto del modelo ARIMA

Ronda 1 ARIMA			
Año	Valor observado	Predicción ARIMA	Error absoluto
2015	130.663	129.799,41	863,59
2016	139.794	136.278,84	3.515,16
2017	151.270	142.908,48	8.361,52
2018	162.497	149.630,56	12.866,44
2019	173.874	156.409,57	17.464,43
Suma errores			43.071,14
Error medio absoluto			8.614,23

5.3 Modelo 3 – Bosques Aleatorios

Los bosques aleatorios se basan en el modelo CART (*Classification and Regression Trees*) presentado por Leo Breiman, Jerome Friedman, Richard Olshen y Charles Stone (Breiman et al., 1984), que propone la fragmentación de predictores en agrupaciones, basadas en una condición deseable y la búsqueda de ella mediante arreglos que logren la mejor configuración de variables para cumplir con la condición establecida.

El punto inicial del modelo son los árboles de decisión, los que usados para problemas de regresión, se componen de raíz, nodos y ramas, donde la raíz es la variable a predecir, cada nodo intermedio es un pregunta sobre los datos (una configuración de variables que predice los valores de la raíz), cada nodo terminal (hoja) es la respuesta a esa pregunta (predicción) y las ramas son las divisiones generadas por las opciones de configuración de las variables seleccionadas, que

presentan diferentes respuestas a la pregunta, ajustándose por minimización de la suma residual de los cuadrados - RSS, entre la predicción y el valor real.

El proceso de los árboles de decisión para predecir el valor de una variable en función de otras variables que son independientes se puede resumir en:

- Selección de la variable que se utilizará como la raíz del árbol.
- División del conjunto de datos en subconjuntos, utilizando la variable seleccionada en el primer paso y asignando aleatoriamente las siguientes variables que apoyan la predicción.
- Predicción de cada observación de la variable raíz, iterando entre variables, valores y configuración de predictores, usando como RSS como optimizador.
- Presentación de la predicción final de cada observación, según el mejor ajuste encontrado.

Los modelos de bosques aleatorios se basan en la construcción de varios árboles de decisión, cada árbol dentro del bosque selecciona un subconjunto aleatorio de datos para entrenar el modelo. Los resultados de cada árbol individual son combinados con los otros, para obtener la estructura que mayor precisión logre en la predicción, realizando el siguiente proceso:

- Selección aleatoria de los datos de entrenamiento, basado en el *set* de datos presentado, construyendo árboles de decisión sobre este subconjunto de datos.
- Predicción del bosque aleatorio, toma la predicción individual de todos los árboles calculados, para combinar los resultados, usando promedios y RSS.
- Entrega como predicción el promedio de todos los resultados individuales y muestra cuáles son las combinaciones que mejor precisión logran.

De esta manera los bosques aleatorios actúan como un proceso de múltiples iteraciones de combinaciones de variables, para realizar predicción y establecer las combinaciones más exitosas (Aaldering et al., 2019).

La aplicación del modelo de bosques aleatorios se presenta siguiendo la forma de pasos usada anteriormente.

Paso 1: Importación de librerías necesarias y carga de datos en Google Colab leídos desde Google Drive.

```
import pandas as pd
from google.colab import drive
drive.mount('/content/drive')
```

Paso 2: Eliminación de variables no necesarias, para este caso *Fecha*, *WIPO1*.

```
data.drop(columns=['Fecha', 'WIPO1'], inplace=True)
```

En el modelo de bosques aleatorios, la estructura de acumulación (*acumu*) usada en los modelos anteriores no es procedente, dado que los valores no permiten una adecuada predicción al presentar saltos de valor entre cada año muy altos, especialmente para los últimos años del *set* de datos, por ello se usa la diferencia entre un año y el siguiente aplicando la función *diff*, que calcula las diferencias consecutivas entre filas del *set* de datos para la variable a predecir (Raschka & Mirjalili, 2019).

```
data['diffs'] = data['acumu'].diff()
```

Del resultado calculado se elimina la primera fila del *set* de datos obtenido, puesto que no contiene valor en la columna creada *diff* al no tener dato inicial para calcular la diferencia, quedando el *set* de datos con 49 años.

```
data = data.drop(index=[0])
```

Paso 3: Determinación de las variables de entrada y de la variable de salida del modelo (valores de X y Y) sobre el *set* de datos final, en el cual la última variable creada (*diff*) se ubica al final de la matriz, identificando esta como, la variable de salida (Y) y las restantes identificándolas como entradas (X).

```
X = data.iloc[:, :-1].values
Y = data.iloc[:, -1].values
print(X.shape)
print(Y.shape)

(49, 124)
(49,)
```

Paso 4: Creación de los *sets* de datos de entrenamiento y de validación, para lograr mayor robustez en el modelo se aplica validación cruzada (dado que el *set* de datos es reducido y un número de parámetros es amplio), asignando 45 registros para entrenamiento y cuatro registros para prueba.

```
X_train = X[:45, :]
Y_train = Y[:45]
X_val = X[45:, :]
Y_val = Y[45:]

print(f'Tamaños set de entrenamiento: {X_train.shape}, {Y_train.shape}')
```

```
print(f'Tamaños set de validación: {X_val.shape}, {Y_val.shape}')

```

Tamaños *set* de entrenamiento: (45, 125), (45,)

Tamaños *set* de validación: (4, 125), (4,)

Paso 5: Creación del bosque aleatorio, usando la librería *Scikit-Learn* y el módulo *Random Forest Regressor*, este módulo establece como hiper-parámetros del modelo:

max_features: cantidad de variables seleccionadas aleatoriamente dentro de las 124 variables del sistema sociotécnico, usadas para la predicción.

n_estimator: cantidad de árboles de regresión que contendrá el bosque.

```
from sklearn.ensemble import RandomForestRegressor
bosque = RandomForestRegressor(random_state=124)

```

Como se desconocen cuáles son las mejores configuraciones de estos hiper-parámetros se recurre al entrenamiento de múltiples árboles, cada uno con hiper-parámetros diferentes, de los cuales se obtiene el desempeño de cada caso, seleccionando el que mejor resultado genera, para luego integrarlo como valor del hiper-parámetro al modelo optimizado.

Paso 6: Validación cruzada, usando como métrica de desempeño el error cuadrático medio ECM, se entrena el modelo y se aplica validación cruzada, en la cual se toma el *set* de entrenamiento y divide en 5 particiones (*n_splits=5*) aleatorias (*shuffle=True*), así el algoritmo selecciona cuatro particiones para entrenar y una para validar. El resultado del entrenamiento se almacena en la variable *n_scores*.

```

from sklearn.model_selection import KFold, cross_val_score

cv = KFold(n_splits=5, shuffle=True, random_state=4)

n_scores = cross_val_score(bosque, X_train, Y_train,
scoring='neg_mean_squared_error', cv=cv)

```

Paso 7: Ajuste del hiper-parámetro $n_features$, el resultado del entrenamiento aleatorio es usado para ajustar el primer hiper-parámetro, construyendo contenedores para almacenar la métrica de desempeño (promedios y desviaciones standard), se crea el modelo con asignación de variables ($max_features$) igual al número de iteración, para evaluar todas las posibles configuraciones (125), se verifica nuevamente por validación cruzada (cv), almacenando los resultados en los contenedores.

```

n_features = np.arange(2, X_train.shape[1])
score_mean, score_std = [], []

for numero in n_features:
    print(f'Entrenando y validando con {numero}
características')
    bosque = RandomForestRegressor(random_state=125,
max_features=numero)

cv = KFold(n_splits=5, shuffle=True, random_state=4)
    n_scores = cross_val_score(bosque, X_train, Y_train,
scoring='neg_mean_squared_error', cv=cv)

    score_mean.append(-np.mean(n_scores))
    score_std.append(np.std(n_scores))

```

El proceso de entrenamiento se evidencia en el paso uno a uno, de las 125 características (variables) iniciando con validación y entrenamiento de dos variables y continuando hasta entrenar y validar con 125 variables, los resultados obtenidos permiten la selección de la mejor configuración.

```

Entrenando y validando con 2 características
Entrenando y validando con 3 características
Entrenando y validando con 4 características
Entrenando y validando con 5 características
Entrenando y validando con 6 características
...
Entrenando y validando con 120 características
Entrenando y validando con 121 características
Entrenando y validando con 122 características
Entrenando y validando con 123 características
Entrenando y validando con 124 características
Entrenando y validando con 125 características

```

El resultado obtenido numéricamente, se presenta en la Figura 5.9, permitiendo hallar el menor valor del ECM, del cual su posición determina la configuración óptima del hiper-parámetro $n_features$, según se indicó por el número de la iteración.

```

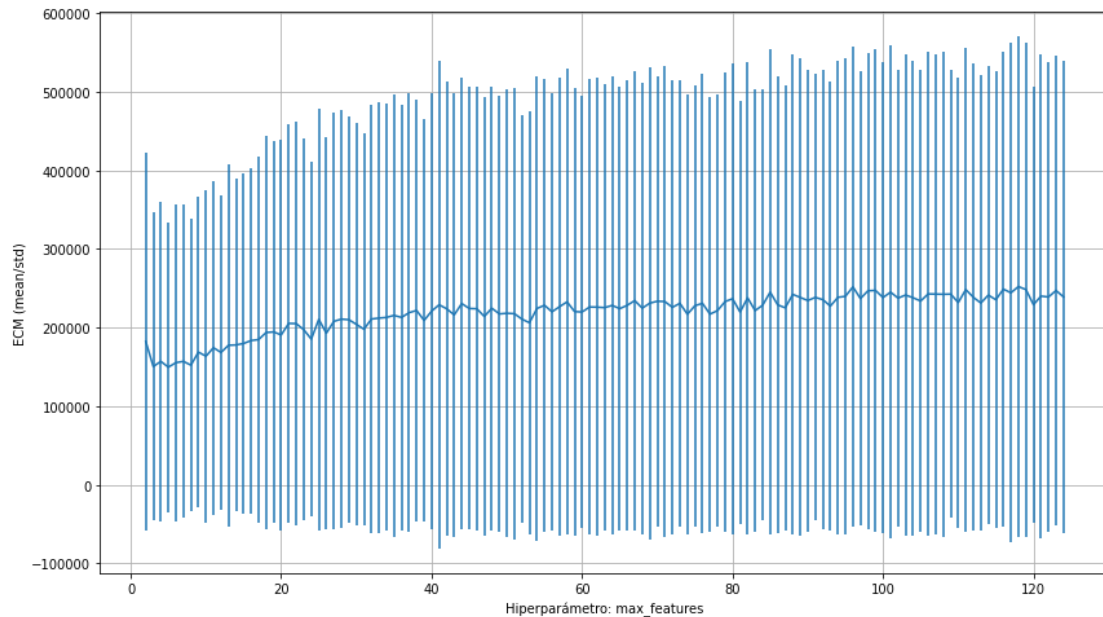
import matplotlib.pyplot as plt

plt.figure(figsize=(14,8))
plt.errorbar(x=n_features, y = score_mean, yerr = score_std
)
plt.xlabel('Hiperparámetro: max_features')
plt.ylabel('ECM (mean/std)')
plt.grid(axis='both')

print (score_std)

```

Como resultado se encuentra que el mejor desempeño se obtiene con cinco características (variables) por lo cual se ajusta el hiper-parámetro $n_features = 5$.

Figura 5.9 Desempeño para $n_features$ 

Paso 8: Ajuste del hiper-parámetro $n_estimators$, usando el valor óptimo de $n_features=5$ antes encontrado, se entrenan y validan diferentes tamaños de bosques, efectuando pasos de dos en dos, iniciando en dos bosques y terminando en 200, para lo cual se crean contenedores que almacenan la métrica de desempeño (promedios y desviaciones standard) de los bosques entrenados.

```
MAX_FEATURES = 5
n_estimators = np.arange(2, 200, 2)
score_mean, score_std = [], []
```

El modelo es creado manteniendo como optimizador ECM y se verifica por validación cruzada (cv), almacenando los resultados en los contenedores.

```
for numero in n_estimators:
    print(f'Entrenando y validando con {numero} árboles')
    bosque = RandomForestRegressor(random_state=125, max_features=MAX_FEATURES, n_estimators=numero)

    cv = KFold(n_splits=5, shuffle=True, random_state=4)
```

```

n_scores = cross_val_score(bosque, X_train, Y_train, scoring='neg_mean_squared_error', cv=cv)

score_mean.append(-np.mean(n_scores))
score_std.append(np.std(n_scores))

```

```

Entrenando y validando con 2 árboles
Entrenando y validando con 4 árboles
Entrenando y validando con 6 árboles
Entrenando y validando con 8 árboles
Entrenando y validando con 10 árboles
...
Entrenando y validando con 192 árboles
Entrenando y validando con 194 árboles
Entrenando y validando con 196 árboles
Entrenando y validando con 198 árboles

```

El proceso de entrenamiento y validación de las 100 iteraciones, con paso de dos en dos, es evaluado numérica y gráficamente, según se presenta en la Figura 5.10, encontrando la optimización del segundo hiper-parámetro.

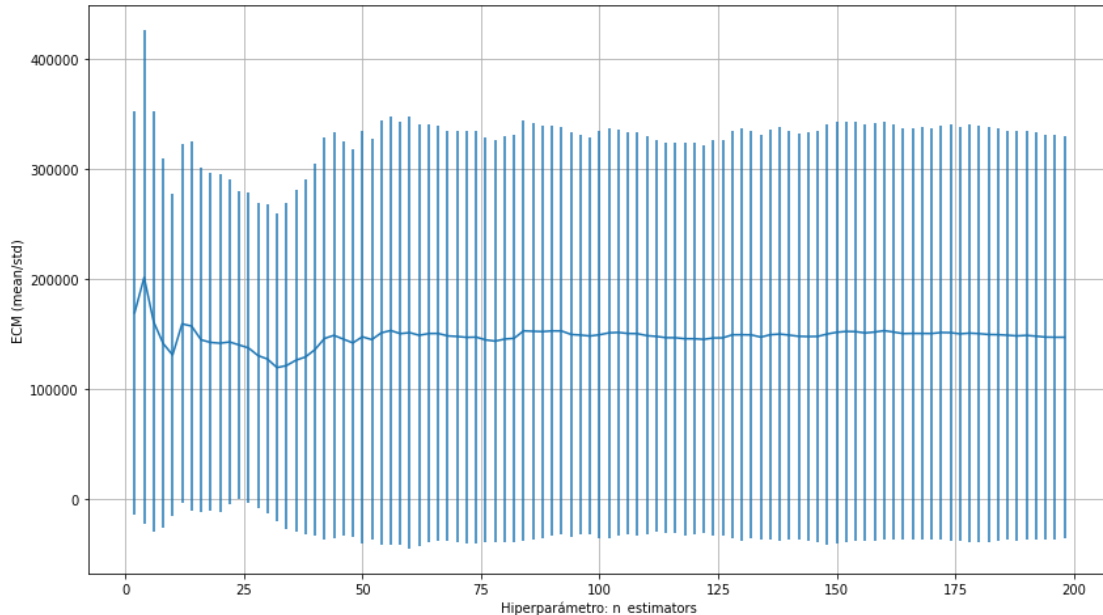
```

plt.figure(figsize=(14,8))
plt.errorbar(x=n_estimators, y = score_mean, yerr = score_std)
plt.xlabel('Hiperparámetro: n_estimators')
plt.ylabel('ECM (mean/std)')
plt.grid(axis='both')

print (score_std)

```

El mejor desempeño se obtiene con un bosque de 32 árboles, donde el ECM es menor, asignado de esta manera $n_estimators=32$.

Figura 5.10 Desempeño para `n_estimators`

Paso 9: Creación y entrenamiento del bosque optimizado, con los dos hiper-parámetros ajustados a su mejor configuración.

```
MAX_FEATURES = 5
N_ESTIMATORS = 32
bosque = RandomForestRegressor(random_state=125, max_features=MAX_FEATURES, n_estimators=N_ESTIMATORS)

bosque.fit(X_train, Y_train)
```

Paso 10: Predicción de la variable `X` y comparación con el *set* de datos de validación, según se asignó inicialmente en la creación del *set* de datos, el horizonte temporal de predicción es de cuatro años (validación) como se presentan en la Tabla 14, donde se exhiben los datos predichos, observando que el desempeño predictivo del modelo es bajo, al denotar diferencias significativas entre la variable esperada (validación) y la predicha (predicción).

```
bosque.predict X_val
Y_val
```

Tabla 14 Resultado predicción Bosque Aleatorio

Valores/Año	2016	2017	2018	2019
Predicción	5.882,93	6.056,65	5.967,93	5.790,15
Validación	9.131,00	11.476,00	11.227,00	11.377,00

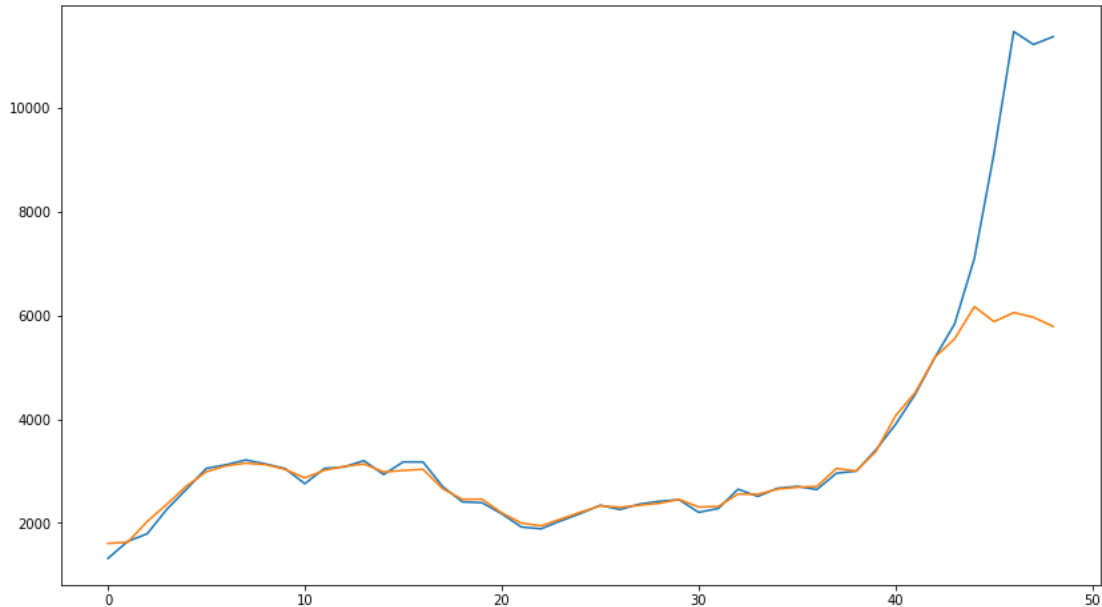
Paso 11: Graficación del desempeño del método, para comprobar los resultados del modelo en todo horizonte temporal, generando grafica comparativa del comportamiento de las variables.

```
Y_pred = bosque.predict(X)

plt.figure(figsize=(14, 8))
plt.plot(Y, label='Y')
plt.plot(Y_pred, label='Y_pred')
plt.show();
```

En la Figura 5.11 la línea azul representa frecuencia observada, calculada inicialmente por la función *diff* (presentando comportamiento similar a la frecuencia absoluta), la línea naranja representa la predicción, encontrando que la predicción se encuentra ajustada en los primeros 43 años, perdiendo eficacia en el último tramo de la curva, justo cuando el crecimiento de la frecuencia observada inicia un rápido ascenso.

Figura 5.11 Resultado de la predicción Bosques Aleatorios V1



Paso 12: Determinación de la importancia de las variables (características) del sistema sociotécnico, una vez comprobado que el modelo está ajustado en la mayor parte de su recorrido, se procede con la identificación de la importancia de las variables que permitieron esa predicción, para ello se elimina la variable calculada *diff*, por no ser una variable predictora.

```
features = data.drop(columns=['diffs'])
```

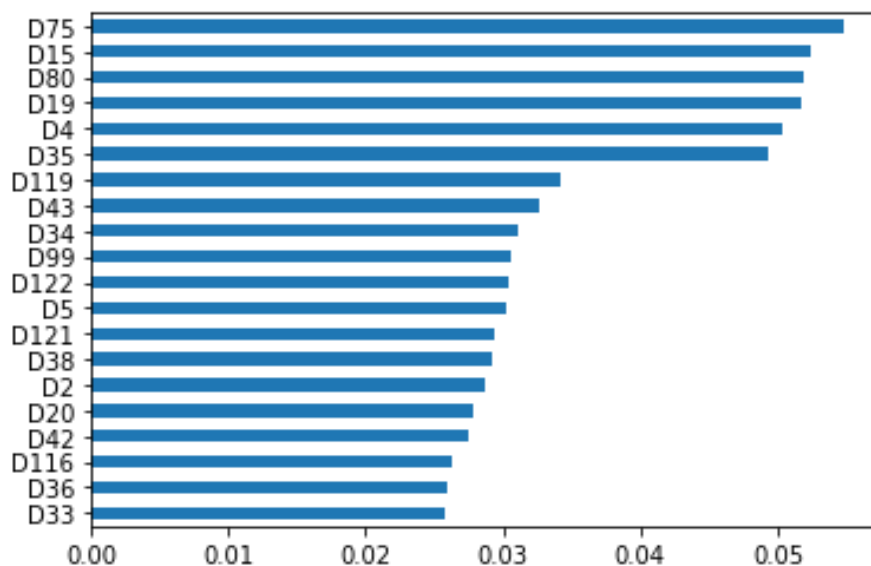
Accediendo al atributo *feature_importances_* del bosque entrenado, se genera un arreglo que contiene los valores numéricos de los niveles de importancia de las variables, justo con su nombre, ordenándolos descendientemente y presentando las 20 variables más importantes dentro del modelo de predicción entrenado.

```
importancias = bosque.feature_importances_
```

```
arreglo = pd.Series(importancias, index=features.columns.values)
arreglo.sort_values()[-20:].plot(kind='barh'); # Las top-20
```

El resultado se muestra en la Figura 5.12, en la que se presentan las 20 variables más importantes del modelo entrenado, ordenadas según su nivel de aporte al proceso de predicción, este resultado es un insumo básico para la futura optimización del modelo general de predicción del cambio incremental, puesto que identifica las variables del sistema sociotécnico más relevantes y permite depurar la matriz de datos.

Figura 5.12 Importancia de variables



Los métodos explorados permiten la optimización del modelo de predicción, encontrando que los tres métodos coexisten positivamente. El primero de ellos, la red LSTM presenta fortalezas en su capacidad predictiva para series de tiempo y múltiples hiper-parámetros de ajuste, el segundo modelo, ARIMA ostenta ventaja en la generación de intervalos de confianza y el tercero de Bosques Aleatorios, muestra relevancia en la identificación de las variables que más aportan en el proceso de predicción.

El resultado de los bosques de predicción permite la optimización de variables dentro del modelo, al usar solo las variables más importantes identificadas para realizar nuevas rondas de predicción en el modelo LSTM y en el modelo ARIMA simplificándolos y obteniendo una configuración más liviana sin perder la capacidad predictiva (Sokmen & Petrov, 2021).

Es necesario aclarar que para cada tecnología y para cada etapa del ciclo de vida las variables de mayor relevancia identificadas por el método de Bosques Aleatorios cambian y por ende es necesario mantener actualizado y alimentado con la última información disponible el *set* de datos, para que los resultados del método bosques aleatorios, sea tributario fidedigno del método LSTM.

Utilizando los datos obtenidos en la exploración de los tres métodos, se determina el proceso de ajuste y optimización del modelo así:

Ajuste 1: Selección de la mejor estructura de la variable a predecir, eligiendo entre frecuencia acumulada y frecuencia absoluta, según la etapa del ciclo de vida de la tecnología y su propio resultado predictivo.

Ajuste 2: Selección de la mejor configuración de hiper-parámetros en los métodos, por medio de comparación entre la predicción y los datos conocidos, teniendo en cuenta las condiciones específicas de la etapa del ciclo de vida donde se encuentre la tecnología.

Ajuste 3: Asignar el tamaño y forma de las variables del sistema sociotécnico, de acuerdo con la identificación de las variables con mayor aporte predictivo identificadas por el método de bosques aleatorios, la asignación está condicionada, a la etapa del ciclo de vida donde se encuentre la tecnología y las variables del pilar B (interés económico del área).

6 Prueba operativa del modelo

Desde la Organización Mundial de la Propiedad Intelectual – OMPI, se obtiene la base de datos de patentes de la subclase A01B (*Soil working in agriculture or forestry; Parts, details, or accessories of agricultural Machines or implements, in general*) la cual contiene un registro de 234.842 patentes para el periodo comprendido entre el año 1800 y el año 2021 (a corte 10 de febrero de 2022), la que se usara para la ejecución del modelo final de predicción del cambio incremental.

Según se indicó anteriormente, el rango temporal de las variables incluidas en el sistema sociotécnico para la predicción son las comprendidas entre 1970 a 2019, por lo cual se mantiene este periodo en la ejecución del modelo predictivo, sin que esto afecte la identificación de la etapa del ciclo de vida, la cual se desarrolla sin alteraciones temporales sobre los datos obtenidos, toda vez que es de vital importancia conocer el punto exacto de la emergencia tecnológica, por lo cual pese a que el horizonte de predicción inicia 1970, para la determinación de las etapas del ciclo de vida se tienen en cuenta todos los datos existentes entre 1800 a 2021 para la subclase seleccionada.

6.1 Descripción de la variable a predecir

El laboratorio de datos para la ejecución del modelo de predicción es el sector agricultura, seleccionado por el interés del investigador y el compromiso con la entidad financiadora de los estudios doctorales, aclarando que el modelo es válido para cualquier otro sector generando ajustes sobre las variables del sistema sociotécnico, en el caso de mantenerse el sector agricultura y solo modificando la subclase de interés, se debe cambiar la variable del pilar A, en cambios de sector económico, es necesario modificar la variable a predecir y las variables del sistema sociotécnico concernientes al pilar B.

Según la clasificación internacional de patentes (WIPO, 2019), la matriz de clasificación se divide en ocho secciones, las que a su vez contienen 131 clases y 645 subclases, Dentro de las cuales se encuentran 7.483 grupos de patentes y 67.020 subgrupos. El detalle de la configuración de la clasificación se presenta en la Tabla 15, donde se observa la estructura de secciones, clases, subclases, grupos y sus grupos.

Tabla 15 Estructura de la clasificación de patentes (2019)

Sección	Clases	Subclases	Grupos	Subgrupos	Subtotal grupos
A	16	84	1.136	8.048	9.184
B	38	169	1.983	15.259	17.242
C	21	87	1.322	13.387	14.709
D	9	39	350	2.726	3.076
E	8	31	323	3.122	3.445
F	18	99	1.099	8.121	9.220
G	15	85	723	7.831	8.554
H	6	51	547	8.526	9.073
Total	131	645	7.483	67.020	74.503

En específico la subclase A01B objeto de estudio tiene la siguiente ubicación dentro de la clasificación, incluyendo el índice de la subclase que presenta los grupos que la componen:

Sección A: *Human necessities* (Necesidades corrientes de la vida)

Clase: A01: *Agriculture; Forestry; Animal husbandry; Hunting; Trapping; Fishing*
(Agricultura; Silvicultura; Cría; Caza; Captura; Pesca)

Subclase A01B: *Soil working in agriculture or forestry; Parts, details, or accessories of agricultural machines or implements, in general* (Trabajo del suelo en agricultura o silvicultura; Partes, detalles o accesorios de máquinas o instrumentos agrícolas en general)

Índices de la subclase A01B: Herramientas de mano, Arados, rastrillos, elementos utilizables como arados o rastrillos o similares, Otras máquinas, Elementos o partes de máquinas o implementos, Transporte en la agricultura, Otras piezas, detalles o accesorios de agricultura maquinas o implementos, Métodos particulares para trabajar el suelo

Según se indicó la subclase A01B contiene 234.842 patentes registradas, por lo que es necesario dividir la búsqueda para generar grupos inferiores a 10.000 registros, según la restricción de descarga de la plataforma Patentscope, la división de grupos y sus ecuaciones correspondientes se presenta en la Tabla 16, específicamente para la regiones América del norte y Asia-Europa, fue necesario crear subdivisiones adicionales (por fechas) que permitieran el cumplimiento de la restricción de descarga, para las demás se usó el código asterisco (*) para indicar como fecha inicial el primer registro conocido.

Tabla 16 Grupos de descarga patentes A01B

Región	Patentes	Fecha inicial	Fecha final	Ecuación
África	186	*	31/12/2021	FP:(a01b) AND AD:([* TO 31.12.2021])
Arabplat	56	*	31/12/2021	FP:(a01b) AND AD:([* TO 31.12.2021])
América	2.605	*	31/12/2021	FP:(a01b) AND AD:([* TO 31.12.2021])
Latiplat	4.651	*	31/12/2021	FP:(a01b) AND AD:([* TO 31.12.2021])
América N	36.735	Dividida por fechas		FP:(a01b) AND AD:([* TO 31.12.2021])
América N	638	*	31/12/1886	FP:(a01b) AND AD:([* TO 31.12.1886])
América N	7.327	1/01/1887	31/12/1910	FP:(a01b) AND AD:([01.01.1887 TO 31.12.1910])
América N	7.390	1/01/1911	31/12/1930	FP:(a01b) AND AD:([01.01.1911 TO 31.12.1930])
América N	8.933	1/01/1931	31/12/1980	FP:(a01b) AND AD:([01.01.1931 TO 31.12.1980])
América N	6.566	1/01/1981	31/12/2010	FP:(a01b) AND AD:([01.01.1981 TO 31.12.2010])
América N	5.881	1/01/2011	31/12/2021	FP:(a01b) AND AD:([01.01.2011 TO 31.12.2021])
ASEAN	662	*	31/12/2021	FP:(a01b) AND AD:([* TO 31.12.2021])
ASIA-EUROPA	185.745	Dividida por fechas		FP:(a01b) AND AD:([* TO 31.12.2021])
ASIA-EUROPA	6.988	*	31/12/1940	FP:(a01b) AND AD:([* TO 31.12.1940])

Región	Patentes	Fecha inicial	Fecha final	Ecuación
ASIA-EUROPA	7.189	1/01/1941	31/12/1960	FP:(a01b) AND AD:([01.01.1941 TO 31.12.1960])
ASIA-EUROPA	7.384	1/01/1961	31/12/1970	FP:(a01b) AND AD:([01.01.1961 TO 31.12.1970])
ASIA-EUROPA	8.772	1/01/1971	31/12/1975	FP:(a01b) AND AD:([01.01.1971 TO 31.12.1975])
ASIA-EUROPA	8.445	1/01/1975	31/12/1978	FP:(a01b) AND AD:([01.01.1976 TO 31.12.1978])
ASIA-EUROPA	7.995	1/01/1979	31/12/1981	FP:(a01b) AND AD:([01.01.1979 TO 31.12.1981])
ASIA-EUROPA	8.486	1/01/1982	31/12/1984	FP:(a01b) AND AD:([01.01.1982 TO 31.12.1984])
ASIA-EUROPA	8.500	1/01/1985	31/12/1987	FP:(a01b) AND AD:([01.01.1985 TO 31.12.1987])
ASIA-EUROPA	8.737	1/01/1988	31/12/1991	FP:(a01b) AND AD:([01.01.1988 TO 31.12.1991])
ASIA-EUROPA	8.699	1/01/1992	31/12/1996	FP:(a01b) AND AD:([01.01.1992 TO 31.12.1996])
ASIA-EUROPA	9.364	1/01/1997	31/12/2001	FP:(a01b) AND AD:([01.01.1997 TO 31.12.2001])
ASIA-EUROPA	7.866	1/01/2002	31/12/2005	FP:(a01b) AND AD:([01.01.2002 TO 31.12.2005])
ASIA-EUROPA	9.221	1/01/2006	31/12/2009	FP:(a01b) AND AD:([01.01.2006 TO 31.12.2009])
ASIA-EUROPA	9.864	1/01/2010	31/12/2012	FP:(a01b) AND AD:([01.01.2010 TO 31.12.2012])
ASIA-EUROPA	9.145	1/01/2013	31/12/2014	FP:(a01b) AND AD:([01.01.2013 TO 31.12.2014])
ASIA-EUROPA	6.085	1/01/2015	31/12/2015	FP:(a01b) AND AD:([01.01.2015 TO 31.12.2015])
ASIA-EUROPA	7.971	1/01/2016	31/12/2016	FP:(a01b) AND AD:([01.01.2016 TO 31.12.2016])
ASIA-EUROPA	9.970	1/01/2017	31/12/2017	FP:(a01b) AND AD:([01.01.2017 TO 31.12.2017])
ASIA-EUROPA	9.790	1/01/2018	31/12/2018	FP:(a01b) AND AD:([01.01.2018 TO 31.12.2018])
ASIA-EUROPA	9.786	1/01/2019	31/12/2019	FP:(a01b) AND AD:([01.01.2019 TO 31.12.2019])
ASIA-EUROPA	9.475	1/01/2020	31/11/2020	FP:(a01b) AND AD:([01.01.2020 TO 31.11.2020])
ASIA-EUROPA	5.996	1/12/2020	31/12/2021	FP:(a01b) AND AD:([01.12.2020 TO 31.12.2021])
TCP-WO	4.219	*	31/12/2021	FP:(a01B) AND AD:([* TO 31.12.2021])

La última clasificación de región, llamada TCP, representa El Tratado de Cooperación en Materia de Patentes, el cual es un sistema de presentación de solicitudes que busca protección por patente en muchos países al mismo tiempo, mediante la radicación de una solicitud internacional.

Se llega así al proceso de compilación de las patentes en el que, sobre todos los 34 archivos de patentes descargados, se modifica la columna de fecha de aplicación, incorporando dos columnas adicionales para crear la estructura: día, mes y año, usando para ello la herramienta de texto en columnas en el software Microsoft Excel, guardando el resultado en formato de archivo plano separado por comas (.csv).

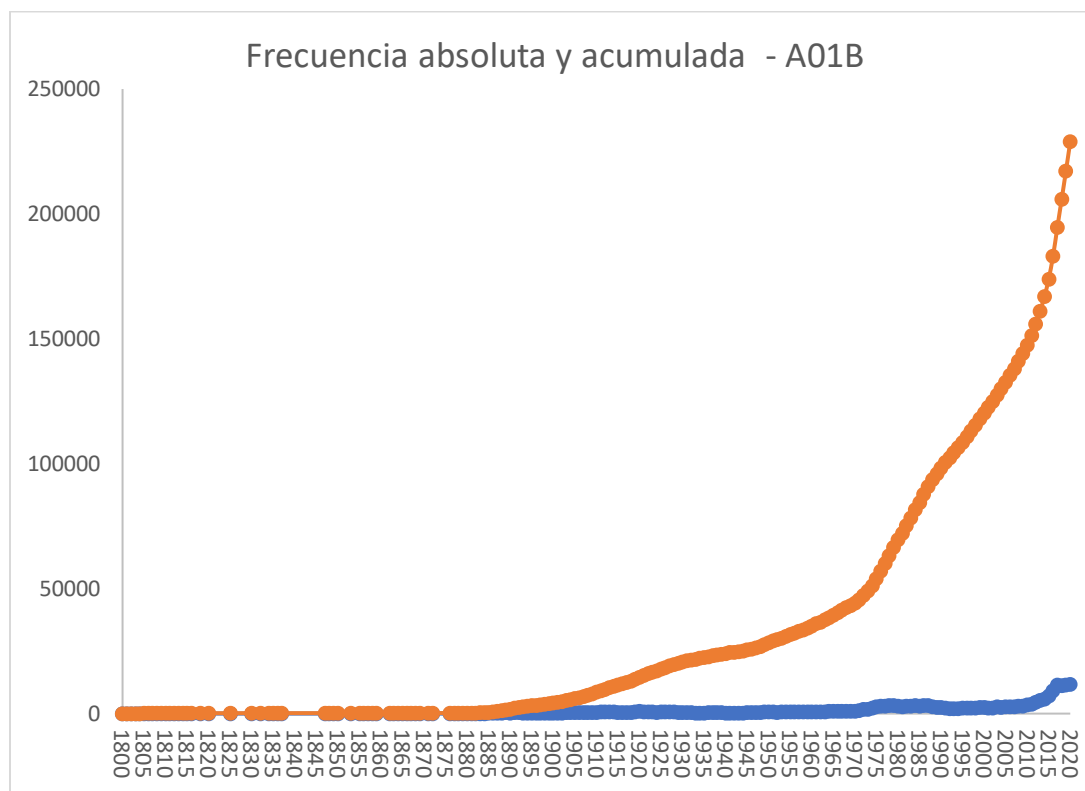
La compilación de los 34 archivos planos descargados se realiza en el software símbolo del sistema de Windows, en el cual se indica la ruta de ubicación de los archivos y se crea un nuevo archivo que los compila a todos, llamado *todo.csv*.

```
C:\Users\juana>cd C:\Users\juana\Desktop\TESIS
2022\primera_compilacion
C:\Users\juana\Desktop\TESIS 2022\primera_compilacion>copy
.\copia\*.csv todo.csv
.\copia\1_con_fecha_A01B_compilado_12_02_2022.csv
.\copia\2_con_fecha_A01B_compilado_12_02_2022.csv
.\copia\3_con_fecha_A01B_compilado_12_02_2022.csv
.\copia\4_con_fecha_A01B_compilado_12_02_2022.csv
1 archivo(s) copiado(s).
```

El resultado es un archivo plano que contiene todas las 234.842 patentes, el que permite su lectura en software de hojas de cálculo. Creando una tabla dinámica, se obtiene la cuenta de patentes presentadas por año, generando así la frecuencia absoluta anual de la variable a predecir.

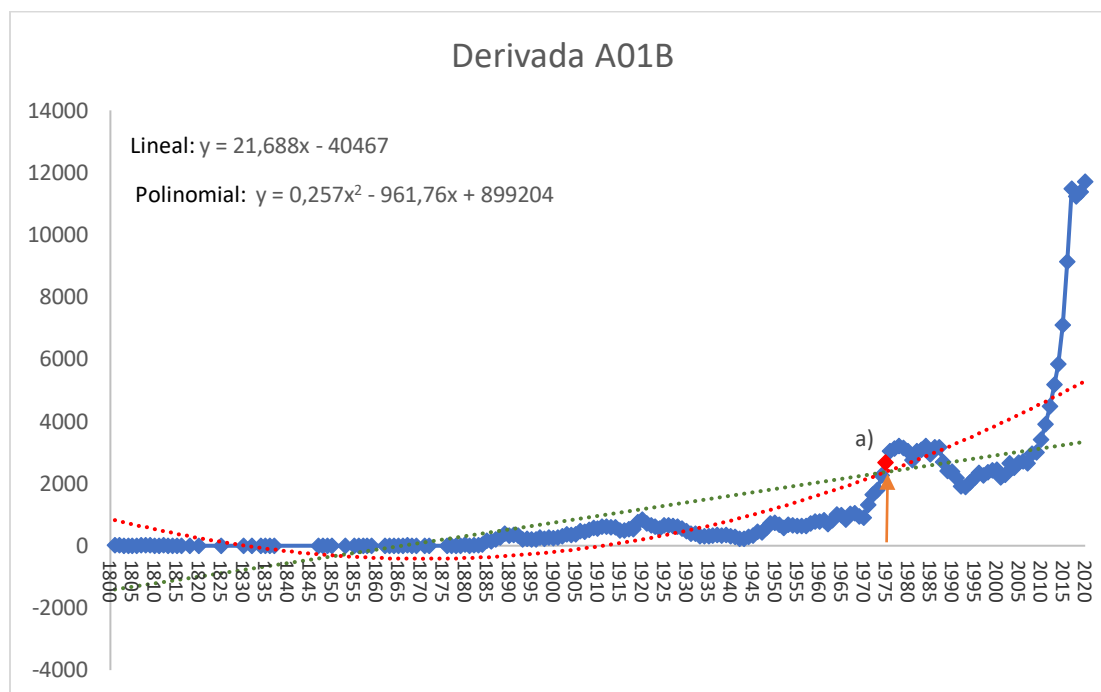
6.2 Identificación de la etapa en el ciclo de vida de A01B

Tomando la frecuencia absoluta de la variable a predecir A01B, se calcula la frecuencia acumulada usando la Ecuación 2, el resultado se presenta en la Figura 6.1, en la cual la línea color azul representa la frecuencia absoluta y la línea color naranja la frecuencia acumulada, en un rango temporal de 221 años.

Figura 6.1 Frecuencia absoluta y acumulada de la tecnología A01B

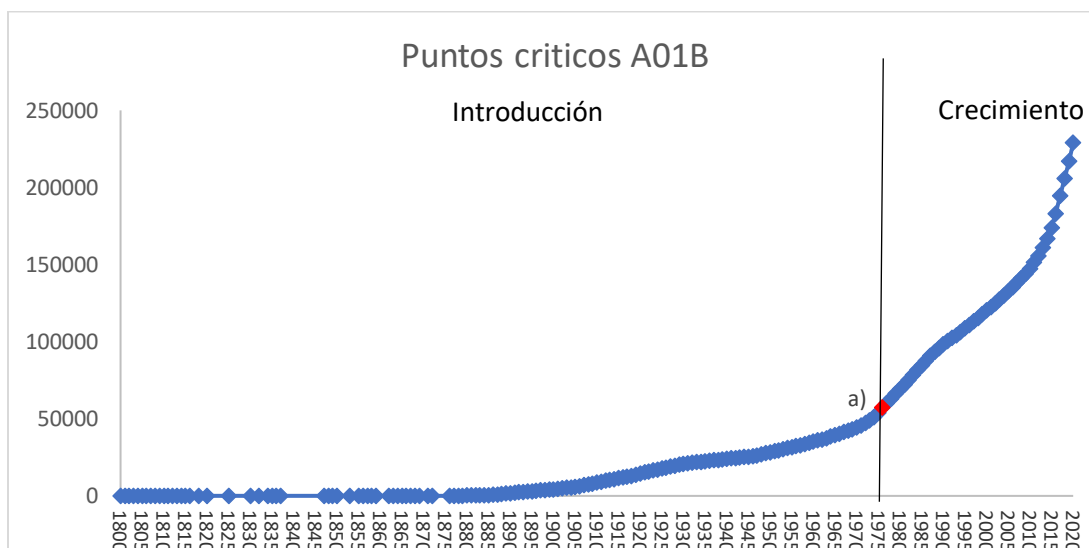
Como se estableció en el método de identificación de la etapa del ciclo de vida, aplicando la Ecuación 1, se obtiene la primera derivada de la frecuencia acumulada, sobre la cual se calculan y trazan la tendencia lineal y la tendencia polinomial, para hallar en su punto de corte sobre el cuadrante positivo, el primer punto crítico a), que determina el paso de la etapa de introducción a la etapa de crecimiento, el resultado se presenta en la Figura 6.2, observando que la tecnología a predecir, se encuentra en proceso de evolución sin llegar aun a la inflexión, por lo que su comportamiento se espera creciente para los siguientes periodos.

Figura 6.2 Uso de derivadas para identificación de las etapas en la tecnología A01B



El año identificado con el punto crítico a) es 1975, lo que presenta un periodo de 175 años en la primera etapa (introducción) del ciclo de vida para la tecnología representada en patentes de la subclase A01B y un periodo de permanencia de 46 años (hasta 31 de diciembre de 2021) en la etapa de crecimiento, sin que este sea su límite evolutivo, solo el máximo dato incluido en la base de datos disponible.

El punto crítico es trasladado a la curva sigmoïdal, según se presenta en la Figura 6.3, identificado con color rojo el año de cambio de etapa (1975) e incluyendo en la figura la línea recta vertical de división entre etapas.

Figura 6.3 Puntos crítico y etapas identificadas en la tecnología A01B

La comprobación del funcionamiento del método de identificación de las etapas de ciclo de vida la tecnología, es alcanzada, por cuanto el ciclo de la tecnología de estudio se encuentra en desarrollo, logrando identificaciones acertadas sin que sea necesaria la evolución completa de sus etapas (introducción, crecimiento, madurez, declive).

6.3 Aplicación de red LSTM

En primer lugar, es necesario segmentar la información de patentes para adecuarla al horizonte temporal del sistema sociotécnico, utilizando la frecuencia absoluta y acumulada de los años 1970 a 2019 de la subclase A01B, sin que esto elimine la información original, solo seleccionando el rango de 50 años de datos establecidos para entrenar el modelo de predicción de cambio incremental.

Siguiendo los pasos necesarios para la implementación del método LSTM según se presentaron anteriormente, sobre los cuales se realiza ajuste y optimización por medio de los hiperparámetros: Neuronas, *epoch*, tamaño del *set* de entrenamiento, tamaño del *set* de prueba y las variables incluidas en el sistema sociotécnico, buscando obtener la mayor precisión en la

predicción. El resumen del proceso se presenta en la Tabla 17, donde la sigla MAE corresponde al error absoluto medio(*mean_absolute_error*).

El ajuste de los hiper-parámetros responde a la comprobación de los resultados obtenidos en cada ronda, sobre la cual se generan cambios, se comparan los resultados y se ajusta nuevamente hasta obtener una configuración mejorada, medido por MAE.

Tabla 17 Rondas de optimización LSTM

Ronda	Variable a predecir	Variables	Set entrenamiento	Set prueba	Neuronas	Epoch	Predicción 2020	Valor observado 2019	% Error	Entrenamiento MAE	Prueba MAE	Valor de acumu 2020
1	Acumulada	124	45	5	16	20	111.867	173.874	0,357	0,03423	1,04297	111.868
2	Absoluta	124	45	5	16	50	5.033	11.377	0,558	0,07801	7,03105	178.907
3	Absoluta	124	46	4	16	50	5.633	11.377	0,125	0,12338	5,36790	183.830
4	Absoluta	124	47	3	16	50	7.297	11.377	0,359	0,10118	2,82868	181.171
5	Absoluta	124	48	2	16	50	8.492	11.377	0,254	0,11199	1,55455	182.366
6	Absoluta	124	48	2	32	50	9.689	11.377	0,148	0,05867	0,90932	183.563
7	Absoluta	124	48	2	64	50	9.956	11.377	0,125	0,04180	0,76572	183.830

El proceso de optimización presenta en la primera ronda la asignación de la variable a predecir frecuencia acumulada, incluyendo todo el *set* de datos del sistema sociotécnico con una asignación de 45 años para el *set* de entrenamiento, cinco años para el *set* de prueba, 16 neuronas y 20 *epoch*, Encontrando que el valor de la predicción es inferior al último valor observado de la frecuencia acumulada, en el entendido que no existe desacumulación es necesario realizar el primer ajuste al modelo, cambiando la variable a predecir por la frecuencia absoluta, el valor de la predicción será agregado al *set* de datos para generar el valor de la frecuencia acumulada predicha y así establecer la trayectoria de cambio incremental.

En la segunda ronda, se modifica la variable a predecir según se indicó, también aumentan

los *epoch* a 50, cambio generado por lo observado en la gráfica de función de pérdida, notando que 20 *epoch* no son suficientes para estabilizar la función, lo que indica que se requieren más repeticiones. Se mantienen los demás hiper parámetros sin alteración, el resultado de esta configuración es una predicción para el año 2020 de 5033 patentes, que comparado con el último dato observado en 2019 es menor con un 55.8%, en el entendido que la tecnología se encuentra en la etapa de crecimiento, se espera que el valor predicho supere al último valor observado, por lo cual se posibilita un nuevo espacio de ajuste para la configuración del modelo.

Los siguientes ajustes de las rondas tres, cuatro y cinco se realizan sobre el tamaño del *set* de entrenamiento y el tamaño del *set* de prueba, en los cuales se aumentan un año en cada ronda para el *set* de tratamiento y se reduce correspondientemente un año en el *set* de prueba, los resultados evidencian una disminución continuada de la media absoluta del error en el *set* de prueba.

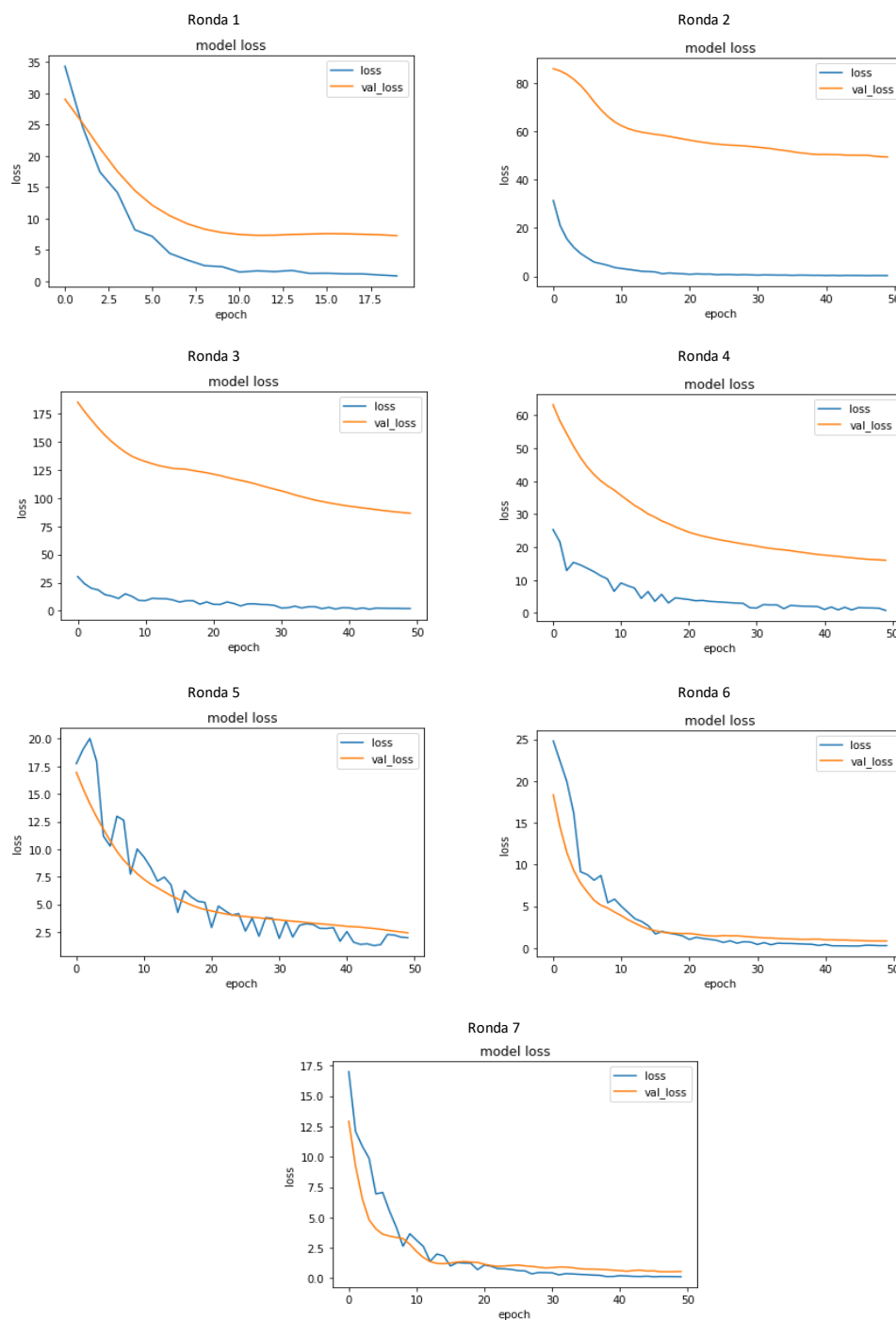
Buscando ampliar la precisión de la predicción se mantiene la configuración de 48 años en el *set* de tratamiento y dos años en el *set* de prueba, modificando la cantidad de neuronas en el sistema en las rondas seis y siete, obteniendo mejoras significativas en el valor de la predicción acercándose cada vez más al último valor observado y reduciendo ampliamente la media del error absoluto, explicado por la menor cantidad de datos en el *set* de prueba, sumado al propio cambio en las neuronas del modelo.

El máximo valor de la predicción es obtenido en la ronda siete con 9.956 patentes, con un error del 12.5% siendo esta la mejor configuración hasta el momento de la red LSTM.

Como evidencia del proceso se presenta en la Figura 6.4 en la cual se observa la evolución de la función de pérdida para las siete rondas realizadas, en la cual la línea color azul representa la

función de pérdida en el entrenamiento y la línea color naranja es la función de la pérdida en la validación (prueba), la función de pérdida del entrenamiento funciona mejor, por la existencia de una puerta de abandono de datos, mientras que en la fase de validación esta puerta es eliminada.

Figura 6.4 Funciones de perdida rondas 1 a 7 LSTM



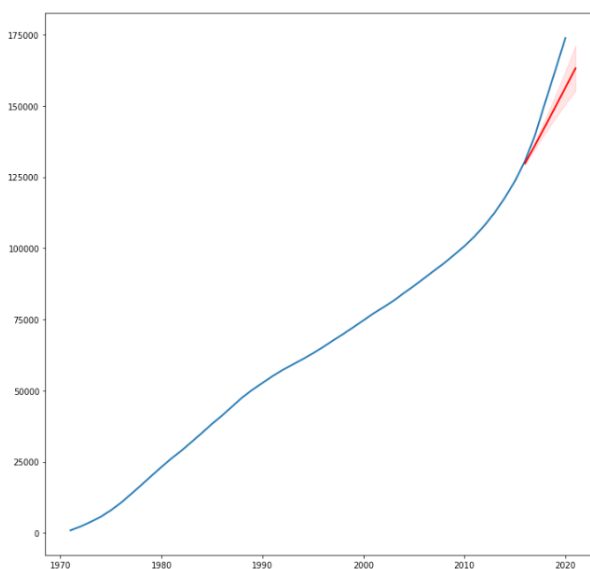
6.4 Aplicación de ARIMA

La primera ronda de ajuste del modelo ARIMA se realiza sobre la variable de predicción frecuencia acumulada y 123 variables del sistema sociotécnico puesto que se mantiene la eliminación de la variable D78, se asigna un *set* de entrenamiento de 45 años y un *set* de prueba de cinco años, encontrando un nivel alto de error medio absoluto, como se presenta en la Tabla 18 y corroborado en la Figura 6.5 donde las curvas del valor observado y el valor predicho son distantes, lo que muestra un bajo nivel de precisión predictiva.

Tabla 18 Ronda 1 ARIMA

Año	Valor observado	Predicción ARIMA	Error absoluto
2015	130.663	129.799,41	863,59
2016	139.794	136.278,84	3.515,16
2017	151.270	142.908,48	8.361,52
2018	162.497	149.630,56	12.866,44
2019	173.874	156.409,57	17.464,43
Suma errores			43.071,14
Error medio absoluto			8.614,23

Figura 6.5 ARIMA 1

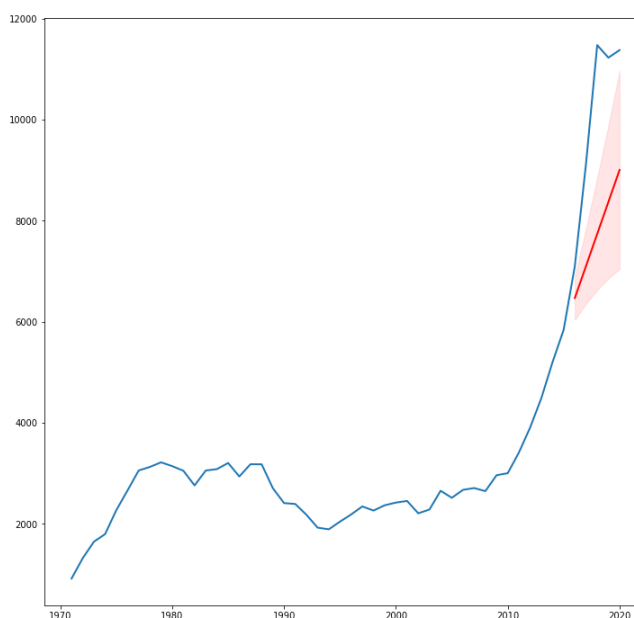


En la segunda ronda de ajuste del modelo ARIMA, se cambia la variable predecir por la frecuencia absoluta de las patentes (WIPO1), se mantienen el *set* de entrenamiento en 45 años, el *set* de prueba en cinco años y las 123 variables del sistema sociotécnico, los resultados se evidencian en la Tabla 19, encontrando reducción en el error medio, entendido por la disminución de la magnitud del valor a predecir, aunque la diferencia entre el valor observado y el predicho continúa siendo amplia como se observa en Figura 6.6, estando tan distanciadas las curvas que no existen cruces con el área del intervalo de predicción.

Tabla 19 Ronda 2 ARIMA

Año	Valor observado	Predicción ARIMA	Error absoluto
2015	7.099	6.468,68	630,32
2016	9.131	7.102,60	2.028,40
2017	11.476	7.736,52	3.739,48
2018	11.227	8.370,43	2.856,57
2019	11.377	9.004,35	2.372,65
Suma errores			11.627,42
Error medio absoluto			2.325,48

Figura 6.6 ARIMA 2

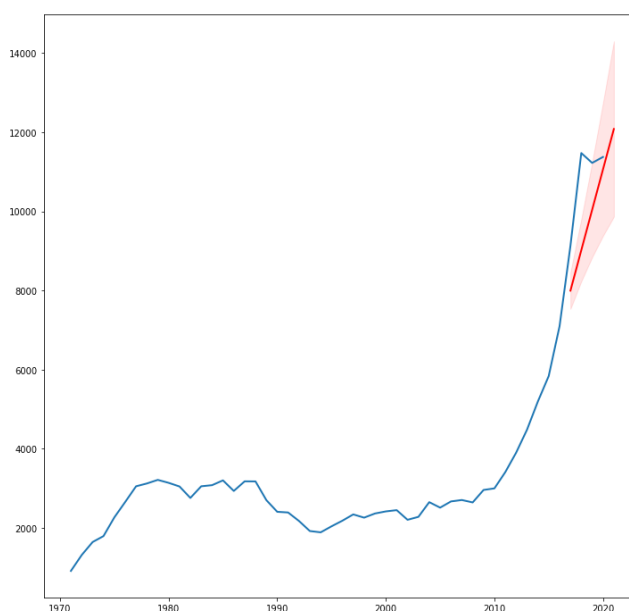


En la tercera ronda el modelo ARIMA es ajustado ampliando el *set* de entrenamiento a 46 años y reduciendo el *set* de validación a cuatro años, lo que genera una reducción en el error medio absoluto presentado en la Tabla 20, como se observa también en la Figura 6.7, en la cual se mantiene un período de predicción de cinco años, pese a tener un *set* de prueba de cuatro años, generando un primer espacio de predicción para el año 2020, se observa un acercamiento de la predicción hasta el punto que el último periodo de la curva observada se encuentra dentro del intervalo de predicción del método ARIMA.

Tabla 20 Ronda 3 ARIMA

Año	Valor observado	Predicción ARIMA	Error absoluto
2015	-	-	-
2016	9131	7.997,88	1.133,12
2017	11476	9.019,56	2.456,44
2018	11227	10.041,23	1.185,77
2019	11377	11.062,91	314,09
Suma errores			5.089,43
Error medio absoluto			1.272,36

Figura 6.7 ARIMA 3

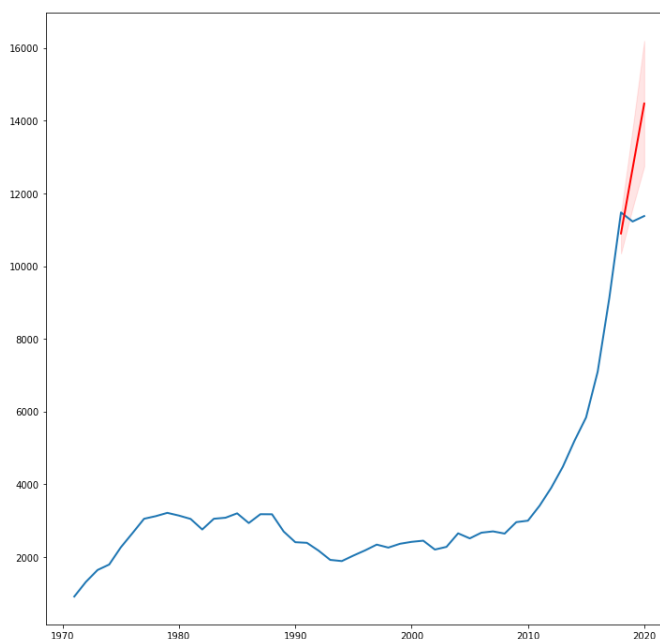


En la cuarta ronda de ajuste del modelo ARIMA, nuevamente se reduce el *set* de validación asignando tres años y se amplía el entrenamiento a 47 años, presentando una variación en el error hacia valores negativos, según se observa en la Tabla 21, representando predicciones de mayor valor que los datos observados, con este ajuste la curva de predicción se acerca más al valor esperado como se presenta en la Figura 6.8 manteniéndose dos años dentro del intervalo de predicción.

Tabla 21 Ronda 4 ARIMA

Año	Valor observado	Predicción ARIMA	Error absoluto
2015	-	-	-
2016	-	-	-
2017	11.476	10.898	578
2018	11.227	12.683,40	-1.456
2019	11.377	14.469,09	-3.092
Suma errores			-3.970
Error medio absoluto			-1.323,40

Figura 6.8 ARIMA 4

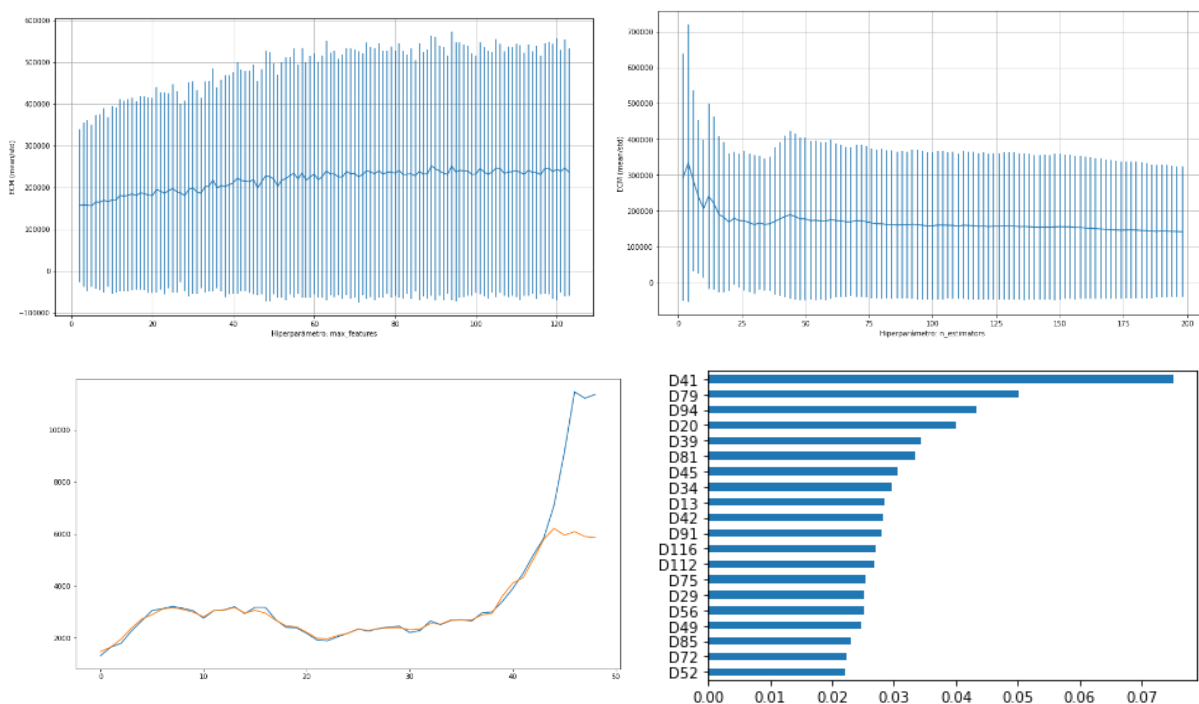


6.5 Aplicación de Bosques Aleatorios

La línea de base el método bosques aleatorios incluye 123 variables del sistema sociotécnico (excluyendo la variable D78 eliminada por comportamiento estacionario), un *set* de entrenamiento de 45 años y un *set* de previsión de cuatro años, de esta configuración inicial se obtienen como hiper-parámetros óptimos cuatro variables y 30 árboles de decisión, el resultado de la predicción presenta 48.5% de error para el año 2019, denotando una baja precisión, evaluado por la comparación entre el valor predicho y el valor observado.

En la Figura 6.9, se presentan los resultados gráficos que permiten ajustar de los hiper parámetros *max_feature* y *n_estimator*, así como la comparación entre la frecuencia observada *diff* (línea color azul) y el valor de la predicción (línea color naranja), finalmente se incluye el gráfico de las 20 variables más relevantes en el modelo de predicción.

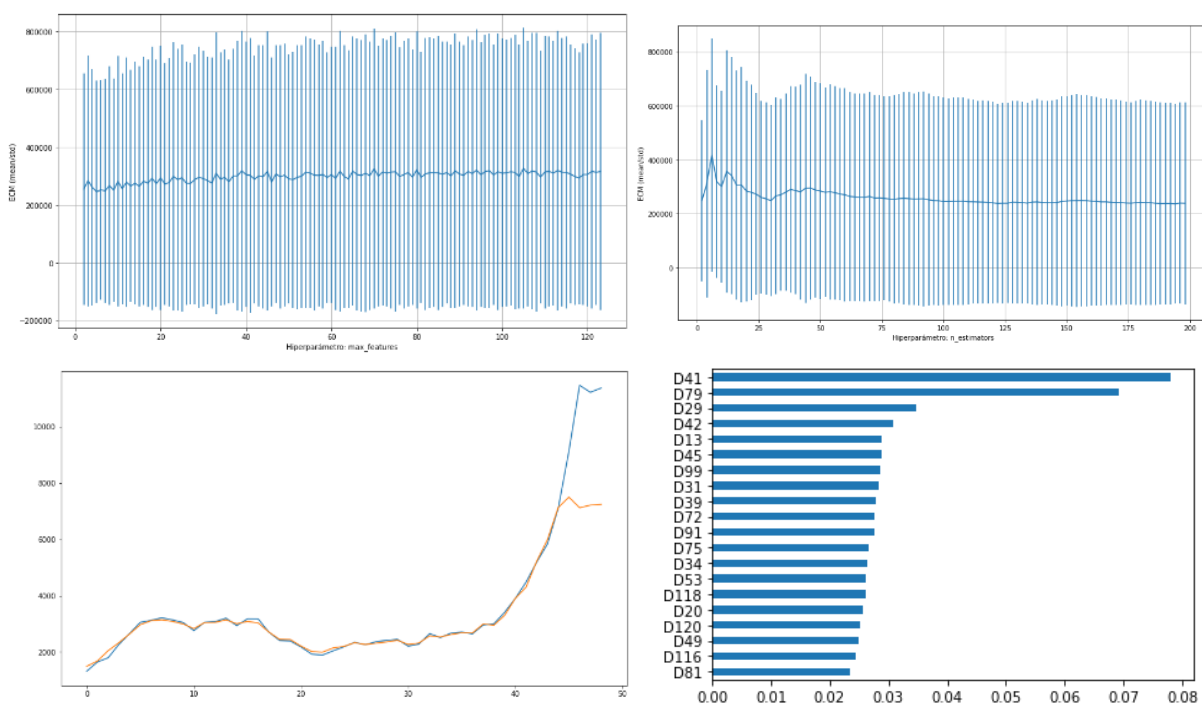
Figura 6.9 Resultados Ronda 1 Bosques Aleatorios



El primer ajuste de bosque aleatorio se realiza en la ronda dos, cambiando el *set* de entrenamiento a 46 años, mientras el *set* de prueba se reduce a tres años, con esta configuración se obtienen como hiper-parámetros cinco características o variables y 30 árboles de regresión, lo cual no presenta grandes variaciones frente a la configuración original, salvo el resultado de la predicción la cual mejora obteniendo un 36.3% de error para el valor predicho del año 2019.

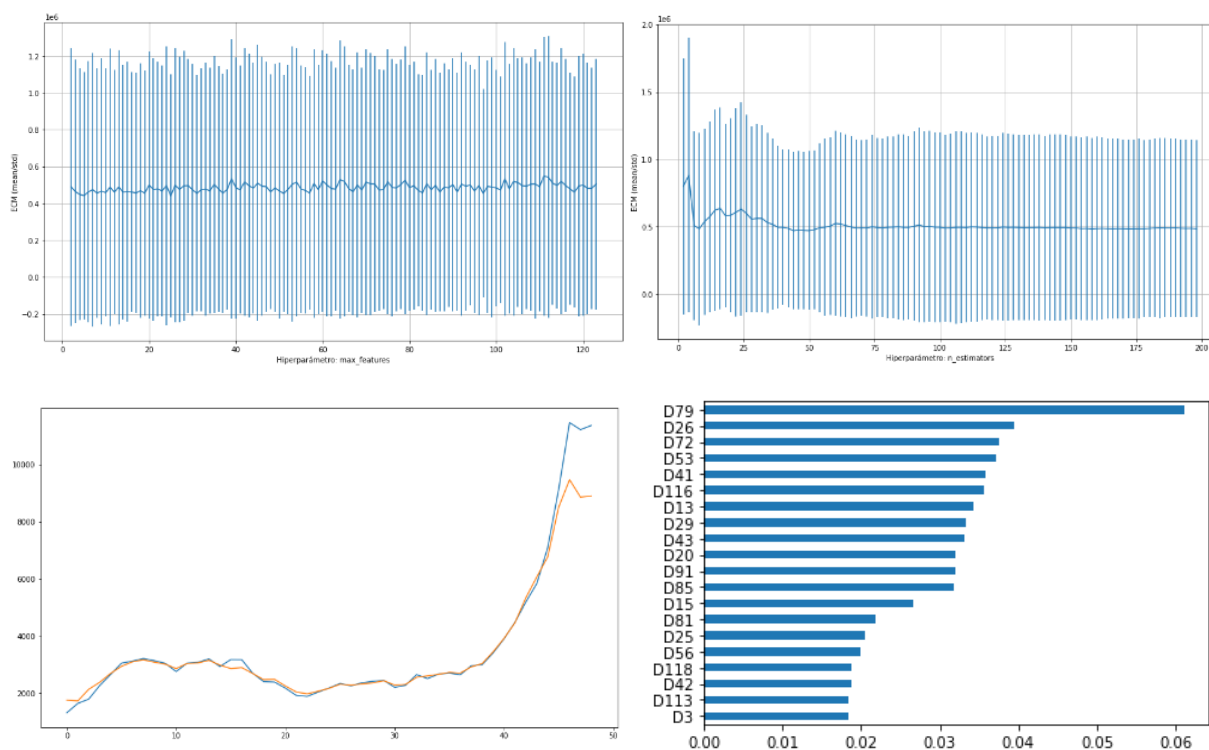
La determinación de los hiper-parámetros se presenta gráficamente en la Figura 6.10, denotándose una estabilidad en el comportamiento de las variables que permiten su cálculo, en este mismo espacio se presentan las curvas de frecuencia observada denominada *diff*, con una línea color azul y el valor comparado de su predicción como una línea en color naranja, mostrando reducción en la distancia entre ellas, comparativamente con la ronda anterior.

Figura 6.10 Resultados Ronda 2 Bosques Aleatorios



En la tercera ronda se continúa con la ampliación del *set* de entrenamiento, llegando a los 47 años, mientras el *set* de prueba se reduce a dos años, con esta configuración se halla que el hiper-parámetro *max_features* requiere 96 características, mientras que el hiper-parámetro *n_estimators* requiere 48 árboles, obteniendo un valor de predicción de 8.908 para 2019 comparado con el resultado observado de 11.377, equivalente a un error de 21.7%, notándose una mejora sustancial en el ajuste predictivo del método, como se muestra en la Figura 6.11, donde la curva de predicción (línea naranja) se acerca más a la curva observada (línea azul), explicado por la ampliación del periodo de entrenamiento y por la asignación de los más altos hiper-parámetros obtenidos hasta el momento.

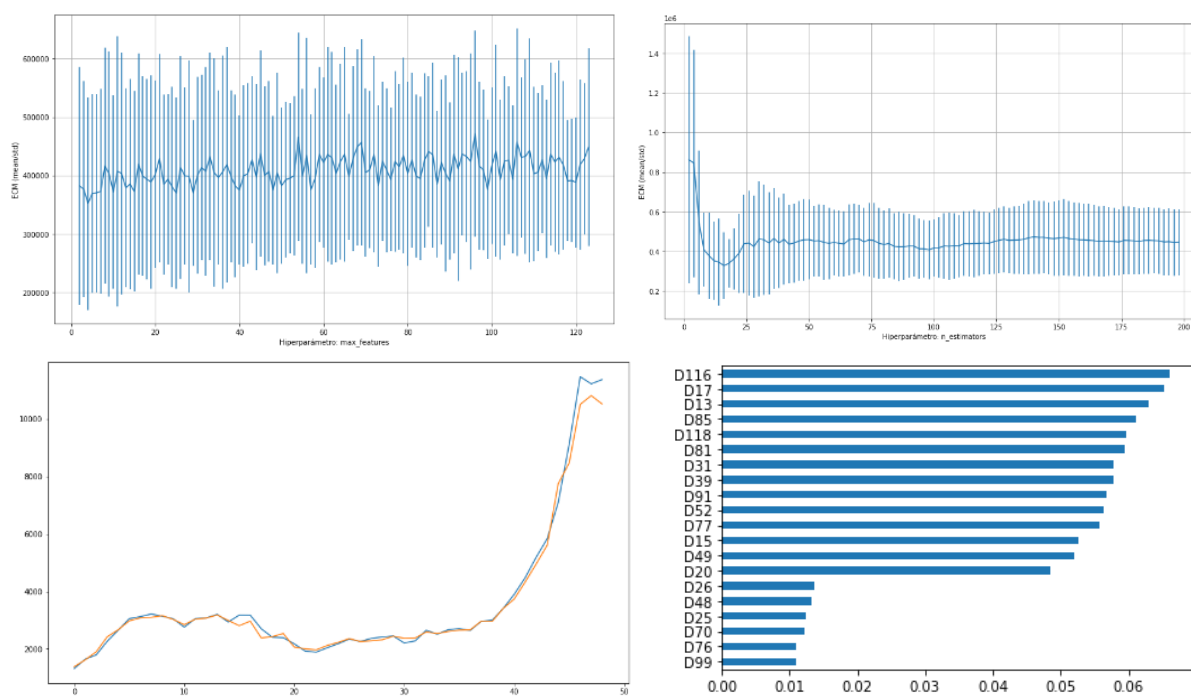
Figura 6.11 Resultados Ronda 3 Bosques Aleatorios



La ronda cuarta de ajuste de los bosques aleatorios, asigna el máximo periodo de 48 años como entrenamiento y un año para la prueba, con la cual se calcula el requerimiento de 117

características (*max_features*) y 14 árboles (*n_estimators*) para generar un valor predicho de 10.871 que sólo dista por error en un 4.4% del valor observado de 11.377, no queriendo indicar con esto que, el modelo se ajusta perfectamente, toda vez que el periodo de prueba es el menor posible, sin que se pueda comprobar más allá su evolución y error, pero logrando calcular las 20 variables más importantes para lograr ese resultado de predicción, así como una visualización de la cercanía de la curva de la frecuencia observada y la frecuencia predicha como se muestra en la Figura 6.12

Figura 6.12 Resultados Ronda 4 Bosques Aleatorios



El resumen de la configuración y los resultados de las cuatro rondas de ajuste del método de Bosques Aleatorios se presenta en la Tabla 22, el principal insumo obtenido de este método son las 20 variables más relevantes para la predicción obtenidas en cada una de las rondas, las que se compilan y se ordenan de acuerdo con la frecuencia de aparición, identificando 36 variables, de las cuales 25 presentan frecuencia mayor o igual a dos dentro de las cuatro rondas. El listado de

variables principales se presenta en la Tabla 23, siendo estas las de mayor representatividad estadística en el aporte al modelo predictivo, por lo cual esta simplificación del sistema sociotécnico es usada en las siguientes rondas de optimización y ajustes finales a la red LSTM y al método ARIMA.

Tabla 22 Resumen rondas de optimización Bosques Aleatorios

Ronda	Set entrenamiento	Set prueba	Max_features	N_estimators	Error año 2019
1	45	4	4	30	48.5%
2	46	3	5	30	36.3%
3	47	2	96	48	21.7%
4	48	1	117	14	4.4%

Tabla 23 Variables de mayor aporte a la predicción

Variable	Pilar	Fuente	Descripción	Código
D13	C	Banco Mundial	Tasa de natalidad, nacidos vivos en un año (por cada 1.000 personas)	SP.DYN.CBRT.IN
D15	C	Banco Mundial	Tasa de mortalidad, adultos, varones (por cada 1.000 varones adultos)	SP.DYN.AMRT.MA
D20	C	Banco Mundial	Tasa de alfabetización, total de adultos (% de personas de 15 años o más)	SE.ADT.LITR.ZS
D25	C	Banco Mundial	Rentas del petróleo (% del PIB)	NY.GDP.PETR.RT.ZS
D26	C	Banco Mundial	Rentas del carbón (% del PIB)	NY.GDP.COAL.RT.ZS
D29	C	Banco Mundial	Remesas de trabajadores y compensación de empleados, recibidas (% del PIB)	BX.TRF.PWKR.DT.GD.ZS
D31	C	Banco Mundial	Proporción de inscripciones de mujeres con respecto a varones en la educación terciaria (%)	SE.ENR.TERT.FM.ZS
D34	C	Banco Mundial	Producción pesquera total (toneladas métricas)	ER.FSH.PROD.MT
D39	C	Banco Mundial	Población, mujeres	SP.POP.TOTL.FE.IN
D41	B	Banco Mundial	Población rural	SP.RUR.TOTL
D42	C	Banco Mundial	Población entre 15 y 64 años de edad, total	SP.POP.1564.TO
D45	C	Banco Mundial	PIB per cápita (US\$ a precios actuales)	NY.GDP.PCAP.CD
D49	C	Banco Mundial	Masa monetaria (% del PIB)	FM.LBL.BMNY.GD.ZS
D52	C	Banco Mundial	Inscripción escolar, nivel terciario (% bruto)	SE.TER.ENRR

Variable	Pilar	Fuente	Descripción	Código
D53	C	Banco Mundial	Inscripción escolar, nivel secundario (% bruto)	SE.SEC.ENRR
D56	C	Banco Mundial	Ingreso nacional bruto (ING) (US\$)	NY.GNP.MKTP.CD
D72	C	Banco Mundial	Esperanza de vida al nacer, total (años)	SP.DYN.LE00.IN
D75	C	Banco Mundial	Emisiones de CO2 (kt)	EN.ATM.CO2E.KT
D79	C	Banco Mundial	Educación de nivel secundario, alumnos	SE.SEC.ENRL
D81	C	Banco Mundial	Densidad de población (personas por kilómetro)	EN.POP.DNST
D85	C	Banco Mundial	Crecimiento de la población urbana (% anual)	SP.URB.GROW
D91	C	Banco Mundial	Cargos por el uso de propiedad intelectual, pagos (balanza de pagos, US\$ a precios actuales)	BM.GSR.ROYL.CD
D99	B	Banco Mundial	Agricultura, valor agregado (US\$ a precios actuales)	NV.AGR.TOTL.CD
D116	B	FAO	Gross Production Value (constant 2014-2016 thousand US\$)(217 productos _1000 US\$)	PROD.VAL.217
D118	C	FAO	Población total (1000 personas)	TT.POB.

6.6 Optimización de red LSTM

Una vez establecidas por el método de Bosques Aleatorios las 25 variables que más aportan al modelo de predicción, se efectúa reemplazo de estas como sistema sociotécnico, en la red LSTM, probado si el ajuste mejora los resultados de predicción.

El resumen del proceso de optimización se presenta en la Tabla 24, donde la octava ronda mantiene su foco sobre la variable de frecuencia absoluta, más las 25 variables identificadas como sistema sociotécnico simplificado, con periodo de entrenamiento de 48 años, periodo de prueba de dos años, 64 neuronas en el modelo y 50 *epoch*, logrando un valor de predicción de 10.828, superando el último resultado alcanzado en la ronda siete, lo que representa un error de 4.82%, lo cual comparado con el 12.23% de error de la ronda anterior, demuestra mejora por el uso del sistema sociotécnico simplificado.

Tabla 24 Optimización de red LSTM

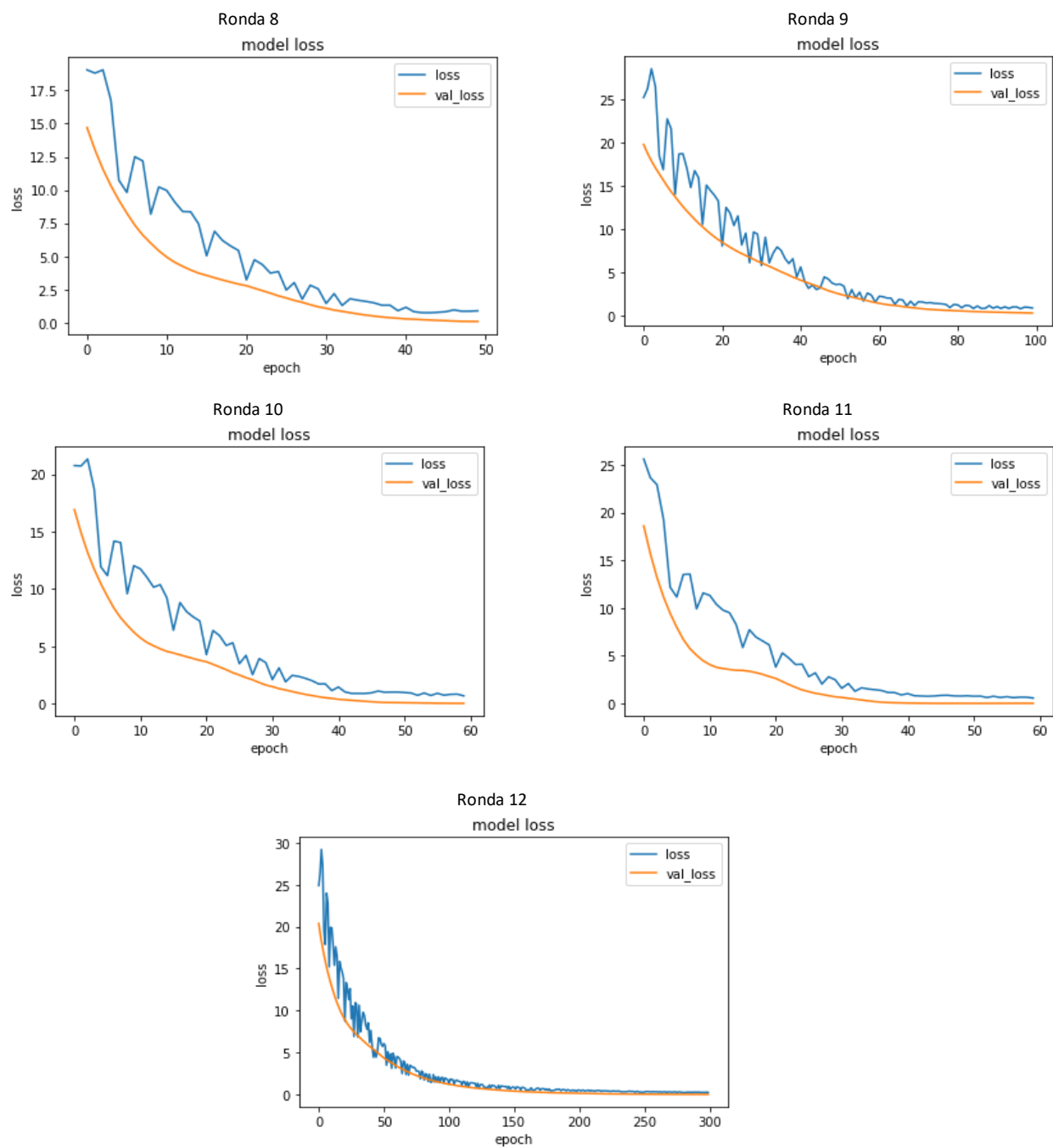
Ronda	Variable a predecir	Variables	Set entrenamiento	Set prueba	Neuronas	Epoch	Predicción 2020	Valor observado 2019	% Error	Entrenamiento MAE	Prueba MAE	Valor de acumu 2020
8	Absoluta	26	48	2	64	50	10.828	11.377	-0,048	0,122	0,296	184.702
9	Absoluta	26	48	2	32	100	10.436	11.377	-0,083	0,120	0,507	184.310
10	Absoluta	26	48	2	64	60	11.148	11.377	-0,020	0,117	0,123	185.022
11	Absoluta	26	48	2	128	60	11.517	11.377	0,012	0,113	0,075	185.391
12	Absoluta	26	48	2	16	300	11.333	11.377	-0,004	0,078	0,024	185.207

En la novena ronda se asigna una reducción a 32 neuronas y aumento a 100 *epoch*, manteniendo los demás parámetros constantes, lo que obtiene una predicción levemente menor a la obtenida en la configuración anterior, aumentando a 8.27% el error, por lo cual se retira la

reducción de neuronas y se procede con la décima ronda de optimización en la cual se retorna a 64 neuronas y se asignan 60 *epoch*, siguiendo el comportamiento de la función de pérdida tanto del entrenamiento como de la validación obtenida en la ronda nueve, en la que se observa una minimización cercana a los 60 *epoch*, por lo cual se asigna esta cantidad en esta ronda, obteniendo una predicción de 11.148 siendo la de mejor ajuste hasta el momento con un error cercano al 2% sobre la verificación del último dato conocido.

Según se presenta en la Figura 6.13, las curvas de la función de pérdida de entrenamiento y predicción de cada ronda tienden a estabilizarse en periodos distintos según la cantidad de neuronas, por lo cual se procede con la ronda once de optimización en la cual se aumenta a 128 neuronas, se mantienen los *epoch* a 60 con el fin de verificar la relación de entre estos dos hiper-parámetros, obteniendo una predicción para el año 2020 de 11.517 correspondiente a un error del 1.23% sobre el último dato conocido, siendo ahora este el nuevo valor de predicción más alto en las once rondas de optimización.

Finalmente se genera una configuración inversa a la anterior reduciendo a 16 neuronas y aumentando a 300 *epoch*, buscando comprobar la estabilidad de las variables predictoras independiente de las configuraciones extremas de los hiper-parámetros, con lo cual se obtiene una predicción de 11.333 3 un error de 4% demostrando que el sistema es estable y se encuentra en condiciones de funcionamiento para realizar la predicción del cambio incremental.

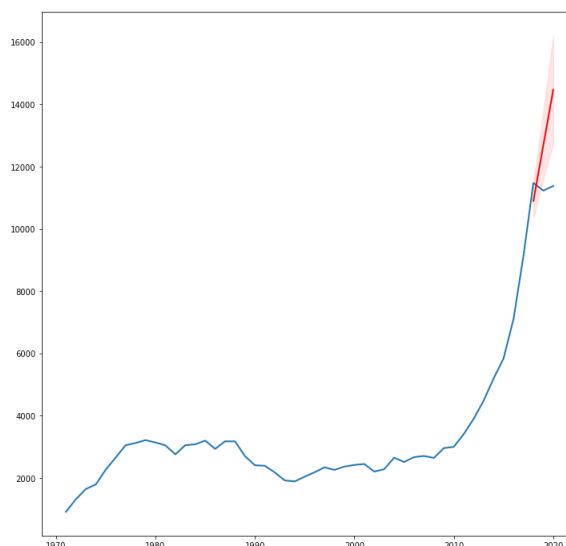
Figura 6.13 Optimización de LSTM

6.7 Optimización de ARIMA

En la quinta ronda del método ARIMA se modifica las variables del sistema sociotécnico según la identificación generada por el método Bosques Aleatorios de las 25 variables de mayor aporte a la predicción, sobre esta se mantienen 47 años de entrenamiento y tres de predicción, encontrando que el error absoluto medio se mantiene constante con lo obtenido en la ronda cuatro, pese a la reducción 99 variables del sistema sociotécnico, según se presenta en la Tabla 25 y la Figura 6.14, lo que corrobora la importancia y la relevancia predictiva de las 25 variables identificadas, de esta forma la curva de predicción, así como el intervalo de confianza de ésta se mantiene, solo reduciendo el tiempo de procesamiento de los datos, dado que con 25 variables el proceso tarda 1.7 segundos, mientras que con 124 variables se toma 6.7 segundos

Tabla 25 Ronda 5 ARIMA

Año	Valor observado	Predicción ARIMA	Error absoluto
2015	-	-	-
2016	-	-	-
2017	11.476	10.897,71	578
2018	11.227	12.683,40	-1.456
2019	11.377	14.469,09	-3.092
Suma errores			-3.970
Error medio absoluto			-1.323,40

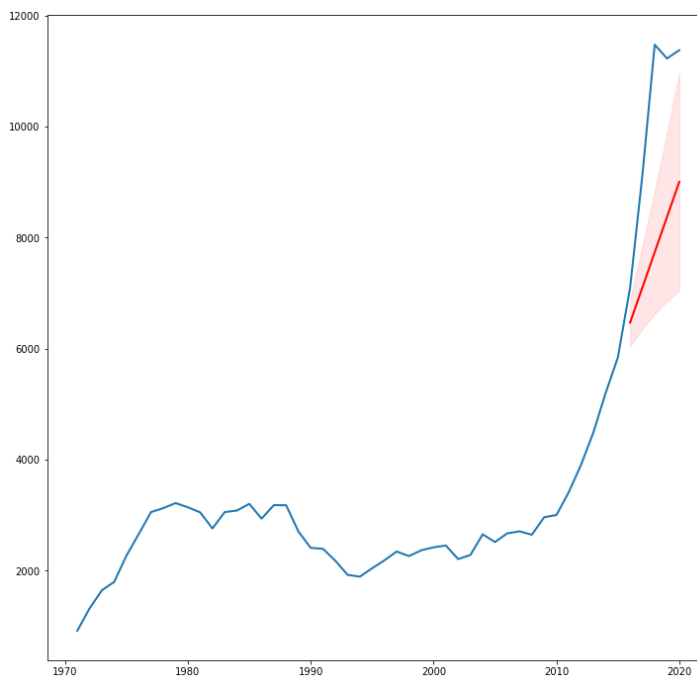
Figura 6.14 ARIMA 5

El último ajuste del método ARIMA se aplica para comprobar sobre la estructura de las 25 variables más relevantes el comportamiento al regresar el *set* de entrenamiento 45 años y el *set* de prueba a cinco años, encontrando que el error retorna a los mismos valores obtenidos en la ronda dos, los resultados son presentados en la Tabla 26 y la

Figura 6.15, confirmando la estabilidad del modelo y ratificando que el mejor ajuste posible se ha obtenido en la ronda cuatro y cinco independientemente de la cantidad de variables del sistema sociotécnico.

Tabla 26 Ronda 6 ARIMA

Año	Valor observado	Predicción ARIMA	Error absoluto
2015	7.099	6.468,68	630
2016	9.131	7.102,60	2.028
2017	11.476	7.736,52	3.739
2018	11.227	8.370,43	2.857
2019	11.377	9.004,35	2.373
Suma errores			11.627
Error medio absoluto			2.325

Figura 6.15 ARIMA 6

Independientemente de la configuración del sistema sociotécnico, el método ARIMA es replicable, pudiéndose optar por un modelo complejo con muchas variables predictoras o por un modelo sencillo con pocas variables en el sistema, en ambos casos se obtiene un nivel de predicción con error cercano al 20%, por lo cual, después de evaluarlo comparativamente con las redes LSTM, este presenta un rendimiento predictivo menor y se descarta su uso en el modelo.

6.8 Evaluación del modelo

El resultado de la optimización de los métodos permite identificar a las redes LSTM con el mejor desempeño medido por su precisión predictiva (Aradi & Hewahi, 2020), obtenido con la configuración de 26 variables, 25 predictoras y una a predecir (frecuencia absoluta), asignación

de 48 años para el entrenamiento y dos años para la prueba, 128 neuronas y 60 *epoch* como hiper-parámetros.

Debido a la ausencia de información completa de las variables del sistema sociotécnico fue necesario eliminar el año 2020 de la matriz de datos, pese a contar con los datos de patentes registradas para el mismo año, este caso excepcional permite comparar el comportamiento del modelo predictivo sobre datos reales no entrenados, de esta manera el número de patentes registradas es 11.715 mientras la predicción del modelo para 2020 es de 11.517, encontrando un error absoluto de 198 patentes, correspondiente al 1.69%, equivalente a 98.31% de precisión, al agregar este resultado a *set* de datos de frecuencia absoluta y calcular su acumulada, se encuentra que esta responde a lo esperado en la etapa del ciclo de vida donde se encuentra la tecnología, de esta manera indicando que la trayectoria tecnológica de la sub clase A01B se encuentra en pleno crecimiento.

En referencia al método ARIMA, este no permite mayores optimizaciones dado que se estabiliza entre sus configuraciones según se demostró, aunque se consiguen mejoras con la simplificación del sistema sociotécnico y es relevante la inclusión que los intervalos de confianza de la predicción observadas como el cono de posibilidades del cambio incremental, este comparativamente tiene una baja precisión predictiva obteniendo error de 27.17% en las rondas cuatro y cinco, por lo cual se descarta del modelo.

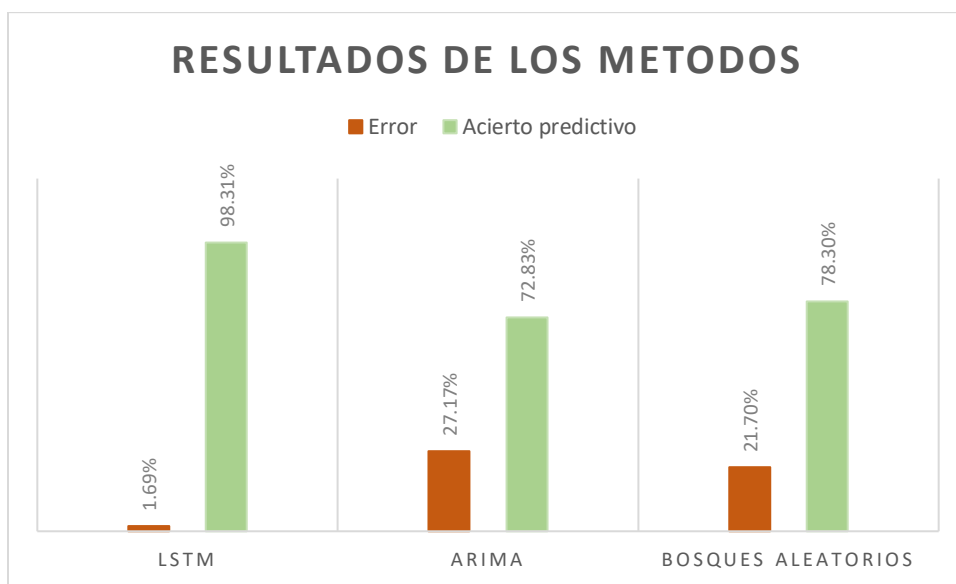
Finalmente, el método de Bosques Aleatorios logra error de predicción inferior al 5% en su ronda cuatro de optimización, pero limitado a un solo año en el *set* de prueba, para homogenizarlo con la configuración de los *set* de entrenamiento y prueba de la red LSTM se usan los resultados obtenidos con el *set* de entrenamiento en 47 años y la prueba en dos años, con lo

cual se logra un error del 21.7%, razón por la cual la precisión del método no es suficiente para usarlo como predictor, pero si contribuye a la identificación de las variables de mayor aporte en el proceso predictivo, las que pueden ser trasladadas a la red LSTM, simplificando el *set* de datos del sistema sociotécnico.

Los resultados evidencian que la red LSTM genera la mayor precisión en la predicción, logrando esto por la propia estructura de las celdas que la componen, especialmente la puerta de olvido que elimina información no relevante para el modelo predictivo, así en cada neurona y en cada *epoch*, se depura y optimiza la predicción, hasta obtener un bajo error y por ende una predicción con alto nivel de precisión.

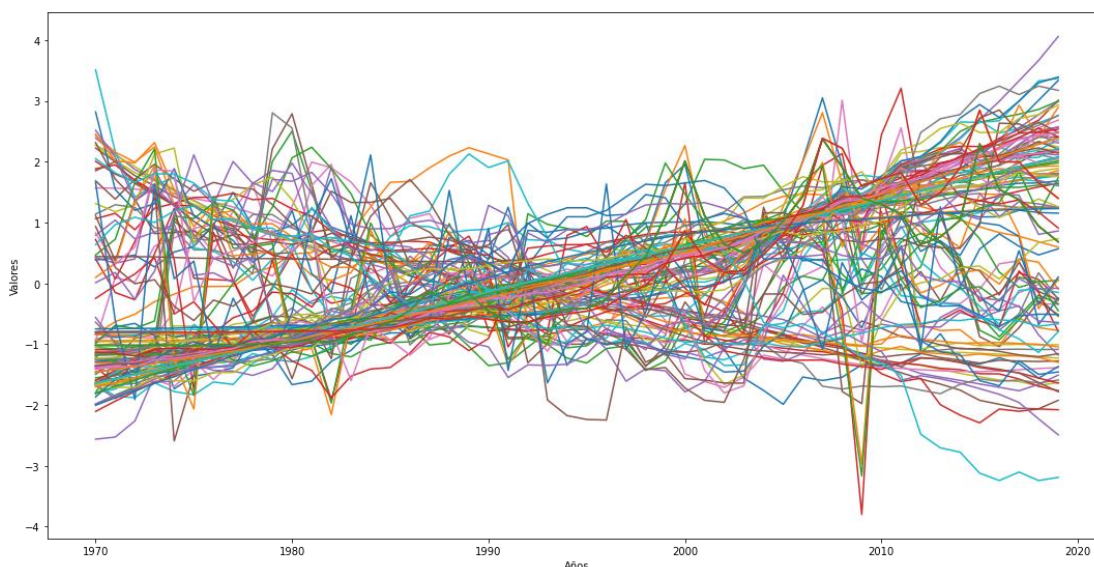
Los resultados de la mejor configuración de cada uno de los métodos explorados se presentan en la Figura 6.16 Mejores resultados de los métodos Figura 6.16, comprobando gráficamente que es la red LSTM la que mayor precisión obtiene en la predicción y por ende permitiendo una respuesta positiva a la hipótesis de investigación.

Figura 6.16 Mejores resultados de los métodos



La obtención de la simplificación del sistema sociotécnico, permite pasar del estado inicial que contiene 125 variables, es presentado en la Figura 6.17, donde se evidencian tendencias, crecientes y decrecientes claras por el volumen de variables, pero difícilmente identificables individualmente, a un sistema reducido, aunque el primer sistema posibilita la ejecución del modelo de predicción completo dando cuenta de múltiples relaciones con la variable a predecir, puede ser mejorado sin sacrificar la precisión.

Figura 6.17 Comportamiento de las 125 variables



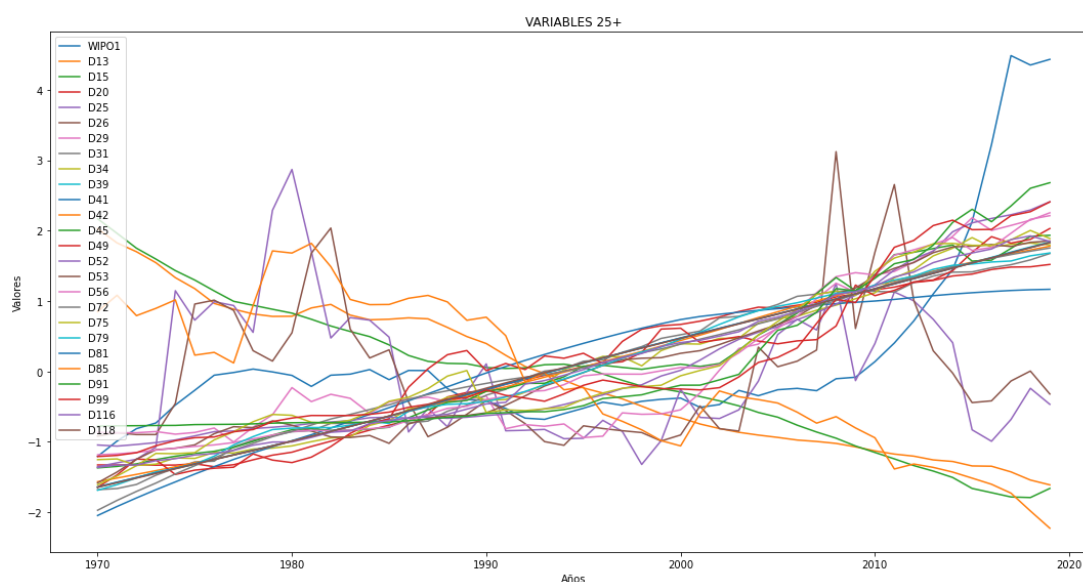
El comportamiento de las 25 variables identificadas como las de mayor aporte predictivo se presentan en la Figura 6.18, correspondiendo a aquellas que, en cuatro rondas de ajuste del método de bosques aleatorios, son identificadas por el atributo *feature_importances_* como las más importantes para la predicción en cada ronda, los resultados de las cuatro rondas son compilados y seleccionadas las variables con frecuencia mayor o igual a dos.

El resultado agrupado por áreas de interés de las 25 variables se presenta en la Tabla 27, de las cuales hacen parte tres de ellas (12%) al pilar B y 22 (88%) al pilar C, contrastado con la estructura inicial del sistema sociotécnico donde la distribución era 27.42% para el pilar B y 72.58% para el pilar C, muestra un cambio en su configuración interna, encontrando que siete áreas permanecen, de las 12 planteadas inicialmente, desapareciendo las variables de las áreas de Agricultura, Comercio exterior, Energía, Transporte, y Gasto militar, puesto que no se identifica alta significancia de estas en el modelo predictivo.

Tabla 27 Distribución del sistema sociotécnico simplificado

Área	Variables
Demografía	9
Educación	5
Producción	4
Financieras	3
Rentabilidad	2
Laboral	1
Medio ambiente	1
Total	25

Figura 6.18 Comportamiento de las 26 variables



En el entendido que el modelo busca la predicción del cambio incremental, el estudio de las relaciones internas entre las variables del sistema socio técnico simplificado y la variable a predecir supera el alcance propuesto para la investigación y abre una ventana de oportunidad para futuros desarrollos investigativos sobre la relación de causalidad entre las variables, que permita identificar la lógica subyacente entre ellas y en el caso de existir también correlaciones meramente matemáticas sin que se presenten relaciones claras, requiriendo de análisis más profundos en busca de identificación de variables intermedias que las conecten y permitan su entendimiento relacional de segundo nivel.

Cabe resaltar que la etapa de ciclo de vida en la que se encuentra la tecnología es un delimitante para las condiciones de su cambio incremental, por ello tecnologías que se encuentran en la fase de introducción hacen esperar comportamientos de leve crecimiento, en la cual cada nueva predicción estará muy cercana a los últimos valores conocidos, caso contrario ocurre en la fase de crecimiento en la que, cada nueva predicción se espera sea superior a los últimos valores conocidos, toda vez que esta fase es la de mayor velocidad de cambio, este comportamiento se revierte a partir del punto de inflexión, donde inicia la etapa de madurez esperando que los cambios sean cada vez menores, con tendencia de reducción de diferencias entre años, finalmente en la etapa de declive los cambios se esperan mínimos, hasta el punto de estabilizar la cantidad de patentes registradas.

Este tipo de comportamientos permiten la evaluación de los resultados del modelo de predicción usando redes LSTM en la combinación con bosques aleatorios (Aithal et al., 2022) para identificar y reducir las variables de mayor aporte al sistema, en ese sentido se logra depurar el *set* de datos, reducir el ruido y obtener un modelo más liviano sin perder precisión.

El resumen de los pasos del modelo se presenta a continuación:

1. Definición de la tecnología a predecir, mínimo a nivel de subclase de la OMPI, descarga de datos y organización en frecuencias absolutas anuales.
2. Generar la frecuencia acumulada de la variable a predecir y sobre esta calcular la primera derivada, la tendencia lineal y la tendencia polinómica, con esta información hallar los puntos críticos de cambio de etapa de la tecnología.
3. Identificar la etapa del ciclo de vida donde se encuentra la tecnología y reconocer su comportamiento esperado.
4. Recolectar y organizar las variables del sistema sociotécnico, especialmente las concernientes al pilar B (interés económico del área), relacionadas al sector de la tecnología seleccionada a estudiar.
5. Depurar la base de datos del sistema sociotécnico, para obtener el mayor lapso con datos completos que puedan ser incluidos en el *set* de predicción.
6. Construir y ejecutar una red LSTM Multivariada, donde se incluyan ventanas de cinco años, 128 neuronas, 60 *epoch*, con 96% de los datos asignados al entrenamiento y 4% asignados a la prueba, logrando la mayor cantidad de información disponible para predecir un solo momento en el futuro, configuraciones diferentes tienen rendimientos degradados según se demostró.
7. Realizar cuatro rondas de Bosques Aleatorios, en cada una de ellas ampliando el periodo de entrenamiento y reduciendo el *set* de prueba, obteniendo de cada ronda las variables que mayor aporte tienen en la predicción.
8. Compilar las variables identificadas en las cuatro rondas de Bosques Aleatorios, organizarlas para filtrar aquellas que se encuentren presentes en mínimo dos de las

cuatro rondas, las cuales se usaran para simplificar el *set* de datos del sistema sociotécnico usado en la red LSTM para la predicción.

9. Reemplazar en la red LSTM construida las variables del sistema sociotécnico, por el grupo simplificado y evaluar capacidad predictiva comparada.
10. Incluir el dato predicho en la frecuencia absoluta de la variable a predecir y calcular la acumulada para evidenciar la trayectoria tecnológica futura.

De esta forma se configura el modelo de predicción del cambio tecnológico incremental, que conjuga la identificación de la etapa del ciclo de vida de la tecnología y métodos de *machine learning*, para obtener la predicción del siguiente paso en la trayectoria tecnológica, comprobando su efectividad y viabilidad de aplicación.

7 Conclusiones

Se logró cumplir con los objetivos específicos propuestos y se mantiene una línea conductora con la definición conceptual y teórica, que permite desarrollar el método de identificación de etapas en el ciclo de vida, caracterizar el sistema sociotécnico, explorar y seleccionar de métodos de predicción, ejecutar el modelo sobre una tecnología seleccionada a nivel de subclase y, finalmente, evaluar la efectividad del modelo de predicción.

El método de identificación de los puntos críticos que separan las etapas del ciclo de vida de la tecnología, alcanzado por el uso combinado de la frecuencia acumulada de patentes, el cálculo de la primera derivada y las tendencias lineales y polinómicas es viable y replicable a cualquier tecnología, logrando determinar con exactitud la etapa donde se encuentra una tecnología, independientemente de cuál de las cuatro etapas esté recorriendo en el momento. Este resultado permite la segmentación del comportamiento esperado de su trayectoria tecnológica y reduce la incertidumbre al estrechar el área de predicción. Las pruebas realizadas en cinco tecnologías de ciclo de vida completo (Máquina de escribir, *Floppy*, VHS, *Pager* y FAX), cada una con diferentes lapsos evolutivos, cantidad de patentes y comportamientos, confirman el funcionamiento del método y posibilitan la ejecución del modelo de predicción.

Es necesario diferenciar los momentos del modelo de predicción. Mientras el comportamiento del ciclo de vida de la tecnología se observa por medio de la frecuencia acumulada (sigmoideal), desde la cual se identifican sus etapas, la predicción requiere que se realice sobre la frecuencia absoluta, permitiendo un mejor manejo de datos en el modelo propuesto, donde el resultado de la predicción se agrega a la frecuencia acumulada, obteniendo la trayectoria tecnológica.

En cada etapa del ciclo de vida se identifican comportamientos esperados de trayectoria tecnológica, en la etapa de introducción la predicción del comportamiento futuro se facilita, dado que la tendencia de crecimiento de la curva no tiene grandes cambios. En ese punto también son pocas las variables del sistema sociotécnico que le afectan.

En la etapa de crecimiento la complejidad de la predicción aumenta, puesto que a medida que avanza la tecnología dentro de la etapa de crecimiento, más rápidos y grandes son los cambios y por ende los datos del pasado difícilmente trazan el comportamiento futuro, adicionalmente una mayor cantidad de variables del sistema sociotécnico influye en el comportamiento.

La tercera etapa de madurez presenta una desaceleración del cambio incremental, reduciendo paulatinamente la cantidad de patentes, en este momento se encuentran presentes en el entorno tecnologías sustitutas, siendo todas afectadas por el sistema sociotécnico.

Finalmente, en la etapa de declive, vuelve a facilitarse la predicción, puesto que su comportamiento tiende a estabilizarse, reduciéndose la cantidad de patentes y la cantidad de variables del sistema sociotécnico que le afectan, declinando el interés en la tecnología y desplazando el paradigma dominante, comportamientos que reafirman los planteamientos de Bass, Rogers, Foster y Levitt.

La caracterización del sistema sociotécnico, dividiéndolo en tres pilares, permite identificar por separado el origen de los datos incluidos en el modelo y usar una configuración de variables por cada pilar, según la tecnología de interés, logrando un modelo de predicción de cambio incremental que se ajusta a diversas áreas.

El conjunto de variables y sus fuentes permiten tener el panorama del comportamiento mundial histórico, tanto del interés económico en el área (pilar B) donde se ubica la tecnología en

estudio, como de las condiciones sociales circundantes (pilar C), sin que estas sean exhaustivas para el contexto, pero sí indicando el rumbo general del sistema, lo que implica una abstracción de la realidad en períodos anuales, puesto que la mayoría de los datos se encuentra en este rango.

El proceso de depuración del sistema sociotécnico permitió la reducción de dimensiones, pasando de 548 variables que contenían más de 31 millones de datos, a un grupo intermedio de 268 variables con 2.617.715 datos para, finalmente, obtener un conjunto de 125 variables con 6.426 datos. Sobre este conjunto sintetizado se probó el modelo y se confirmó su funcionamiento.

Las variables del sistema sociotécnico tienen diferentes niveles de importancia según sea la etapa del ciclo de vida donde se encuentre una tecnología, así que las mismas condiciones sociotécnicas afectan de manera diferenciada a las tecnologías dependiendo de su estado de evolución.

El uso de patentes en el nivel de subclase como variable a predecir, dista de la completitud. Su imperfección radica en el parcial reflejo de los cambios tecnológicos de un sector, excluyendo el saber hacer y la protección por secreto industrial, entre otros, pero es válido como predictor general a falta de un índice más preciso.

El uso de métodos de *machine learning* en específico el aprendizaje supervisado para ejecutar el modelo de predicción, beneficia la velocidad y potencia de cálculo, además de posibilitar el ajuste de los hiper-parámetros para obtener mejor precisión predictiva. En esta vía se evidencio que las redes LSTM son las más potentes y precisas para predecir series temporales multivariadas, mientras que el método ARIMA genera resultados de predicción con errores cercanos al 27%, difícilmente optimizables, su mayor aporte es la generación de conos de posibilidades de predicción, entendidos como intervalos de confianza. Por su parte, los Bosques

Aleatorios permiten la identificación de las variables de mayor aporte al modelo predictivo, información que es usada para construir un conjunto de datos de sistema sociotécnico simplificado, que retroalimenta la red LSTM, reduciendo la complejidad del modelo sin perder su capacidad predictiva.

Los métodos de predicción explorados requieren de un proceso de iteración, donde los bosques aleatorios tributan la identificación de variables más relevantes a la red LSTM, logrando ajustes de predicción robustos en períodos de tiempo y conjunto de datos relativamente pequeños. Esta simplificación del sistema sociotécnico, en asocio con los métodos de predicción, permiten la alta precisión del modelo de cambio incremental logrado.

Con la aplicación del modelo de predicción se ha identificado que la tecnología de la subclase A01B se encuentra en etapa de crecimiento, con punto crítico de cambio de etapa en 1975. Así, la primera etapa de introducción permaneció durante 175 años iniciando en 1800, y el desarrollo de la etapa de crecimiento cuenta hasta diciembre de 2021, con 46 años de evolución. Como predicción final se obtuvieron 11.517 patentes para el año 2020, lo que representan una precisión mayor al 98%, al compararlo con el reporte oficial de la OMPI.

El modelo es novedoso en el sentido en que aporta el proceso necesario para corroborar las teorías de evolución tecnológica, ciclo de vida tecnológico, curvas sigmoideas y sistemas sociotécnicos, entendidos como un conjunto interrelacionado que se aborda con métodos estadísticos soportados en la capacidad de procesamiento generada por la inteligencia artificial (aprendizaje de máquina con redes neuronales), a su vez permite su uso en diferentes campos donde la evolución de tipo sigmoideal es relevante, así como la aplicación para cualquier tecnología, independientemente de la tapa del ciclo de vida donde se encuentre.

7.1 Limitaciones

La disponibilidad de datos de fuentes oficiales abiertas, con datos mundiales, es reducida, encontrando dificultades en la homogeneidad de los datos, la estabilidad de la medición y el periodo disponible, por lo cual algunas variables de gran interés se excluyeron del sistema sociotécnico final, al incumplir los requisitos establecidos. Esta situación limita la configuración del sistema sociotécnico y abstrae más allá de lo deseable las entradas al modelo.

El desconocimiento de los cambios tecnológicos que no se encuentran registrados en la Organización Mundial de la Propiedad Intelectual, dificulta la exactitud de los datos, generando vacíos sobre la realidad completa de la evolución tecnológica, toda vez que los mejores datos disponibles son los de las patentes y sobre estas mismas se ejecuta la predicción, manteniendo el subregistro como un factor propio del cambio tecnológico.

Lograr predicciones en tiempo real, o con periodicidad inferior a un año, permanece en el terreno de lo deseable, pero no de lo posible, puesto que la heterogeneidad de los reportes de datos del sistema sociotécnico lo impiden y, además requiere del desarrollo de código *frontend* que sea amigable con los usuarios para interactuar y aprovechar sus resultados.

El aprendizaje del lenguaje Python, incluidas las librerías específicas de *machine learning* para series de tiempo, es una barrera tanto para el autor, en la búsqueda de un código más refinado, como para los usuarios e investigadores, toda vez que obliga el conocimiento de nuevas herramientas para intervenir y usar el modelo de predicción.

7.2 Aportes

La teoría de Giovanni Dosi, expresada en 1982 sobre trayectorias tecnológicas (Dosi, 1982), en la cual la discontinuidad tecnológica surge como resultado del cambio de paradigma dominante,

y presenta un primer acercamiento a los factores que rigen el proceso de selección de la nueva tecnología dentro de un grupo de posibilidades, es controvertida en el supuesto de discontinuidad, ya que se comprobó que las tecnologías se ralentizan, son absorbidas por otras, o reemplazadas en su etapa de declive, pero no por ello desapareciendo, solo cesando su evolución.

El profesor Geels ha desarrollado una teoría del comportamiento sigmooidal de la curva de ciclo de vida tecnológico (Geels, 2005d; Geels & Kemp, 2007), que considera basado en una red de multi actores en regímenes sociotécnicos, como una explicación al cambio tecnológico, lo cual es comprobado con el uso de frecuencias acumuladas y el modelo multivariado de los pilares B y C, demostrando que existen relaciones fuertes entre las variables índices del sistema y no solo entre los actores, sino también evidenciando cómo estas relaciones permiten identificar la dinámica del cambio incremental.

El símil entre ciclo de vida del producto y el ciclo de vida de la tecnología y las etapas que lo componen ha sido ampliamente abordado, pero carecía de un método de identificación. En este trabajo se tomaron las etapas propuestas por Theodore Levitt: introducción, crecimiento, madurez y declive (Levitt, 1965b), y se logró un método de identificación de los puntos críticos de cambio entre etapas y, desde allí, la propia separación de estas. El método matemático de identificación aporta a la teoría, creando una verificación aplicable a cualquier tecnología, independientemente de su estado evolutivo.

Keith Pavitt plantea el cambio tecnológico desde los sectores que lo generan y como éstos trazan una trayectoria de cambio (Pavitt, 1984b). Ese trabajo demuestra que dicho patrón de cambio puede ser estudiado. Específicamente la trayectoria tecnológica es evidenciada por patentes a nivel de subclase, pudiendo incluso especificarse más, a nivel de grupo o subgrupo, evidenciando que, en efecto, las trayectorias son claras, presentando comportamientos predecibles,

sujetos a restricciones de ciclo de vida y condiciones circundantes, lo que se llamó originalmente taxonomía del cambio, se presenta ahora como el pilar A (cambio tecnológico generador), creando una métrica para la predicción de la trayectoria y posibilitando su gestión.

Al modelo de Bass de adopción tecnológica (Bass, 1969b), que se encuentra fuertemente relacionado con la curva sigmoïdal y la evolución tecnológica en la que los innovadores, los adoptantes tempranos, tardíos y los rezagados, son parte constitutiva del cambio incremental, se adicionan las expresiones de los pilares B y C sobre los cuales se edifica el proceso de acumulación de frecuencias de interesados en el desarrollo tecnológico específico, aportando en su cálculo específico.

Entender la tecnología como un elemento vivo que evoluciona y se puede estudiar matemática y estadísticamente, contribuye al corpus general de la gestión de la tecnología, más aún la incorporación de métodos de *machine learning* actualizan la técnica y proponen una vertiente de modelo neoshumpeteriano de medición de cambio tecnológico incremental.

7.3 Trabajo Futuro

Ante el avance de la inteligencia artificial, en especial el *machine learning* para series de tiempo, se abre un espacio de mejora de código que permita la mejora del proceso y de la depuración, inclusive la integración con *software* como SQL y la migración a otros lenguajes de programación como R o Julia, los que permitan el desarrollo de un proceso automatizado para evaluar y predecir diferentes tecnologías, generando salidas graficas tipo *dashboard* para la toma de decisiones sobre los datos obtenidos.

La potencia de las redes neuronales *transformers* posibilita el desarrollo de optimizaciones en el modelo de predicción, mejorando su desempeño, así como el procesamiento de secuencias en paralelo que determinen relaciones entre las diferentes secuencias simultáneamente.

Para tipificar los comportamientos evolutivos de las tecnologías, es posible iniciar investigación sobre las reglas o condiciones específicas de cada etapa del ciclo de vida, requiriendo el estudio de múltiples subclases de patentes para estandarizar su métrica, tales como tiempo de duración por etapa, condicionado a los tiempos en etapas anteriores, altura máxima esperada, pendiente de la curva, entre otros. Con estos datos es viable construir trayectorias tecnológicas de largo alcance para influir en su desarrollo.

En el entendido de que cada etapa del ciclo de vida tiene variables con mayor efecto que otras, es posible obtener del conjunto simplificado del sistema sociotécnico la fuerza de la relación entre variables, así como la relación de éstas con la tecnología. Estas relaciones despliegan toda una nueva área de profundización, en la cual se establezcan la causalidad entre ellas y se identifique la existencia de rezagos temporales entre unas y otras.

Las tecnologías están correlacionadas, aunque se estudian independientemente, ya que una tecnología incorpora otras, por lo cual estudiar y predecir el comportamiento de todas las tecnologías que sean constitutivas de otra o de un artefacto, permite identificar los frenos al cambio o restricción evolutiva, fundando las bases para la gestión específica de componentes y su relación, así como la medición de su impacto en las tecnologías con las que se relaciona.

Profundizando en las relaciones entre el sistema sociotécnico y una tecnología, es posible la inclusión de los ciclos de Kondratief y/o la inclusión de rezagos temporales en las variables predictoras para buscar evidencia de relaciones de causalidad con desfase temporal entre variables,

donde los resultados en una variable hoy producen resultados en otra variable en el futuro, en especial sobre la identificación de variables con alto nivel de impacto en los cambios tecnológicos futuros.

La incorporación de la dimensión geográfica ha sido expuesta anteriormente y es altamente deseable para la identificación de los lugares donde se generan los cambios incrementales, creando curvas sigmoideas separadas por países o regiones, con sus respectivas etapas de ciclo de vida, dando cuenta del interés de una región por una tecnología y el relevo de dominadores tecnológicos.

Una interesante exploración futura es la creación de las relaciones duales, como la existente entre el número de patentes y los datos demográficos. En especial para el sector agricultura, la relación entre el cambio tecnológico del sector y la producción de alimentos en el sentido de David Ricardo sobre los rendimientos decrecientes y productividad marginal, en la cual se definen las tierras fértiles como finitas y por ende, luego de ocupadas todas, solo queda el camino de la mejora tecnológica para aumentar la producción que atienda un número creciente de demandantes, reviste de pertinencia a los pilares B y C como detonantes del cambio tecnológico.

8 Referencias

- Aaldering, L. J., Leker, J., & Song, C. H. (2019). Analysis of technological knowledge stock and prediction of its future development potential: The case of lithium-ion batteries. *Journal of Cleaner Production*, 223, 301-311. Scopus.
<https://doi.org/10.1016/j.jclepro.2019.03.174>
- Abernathy, W. J., & Clark, K. B. (1985). Innovation: Mapping the winds of creative destruction. *Research Policy*, 14(1), 3-22. [https://doi.org/10.1016/0048-7333\(85\)90021-6](https://doi.org/10.1016/0048-7333(85)90021-6)
- Adner, R., & Kapoor, R. (2010). Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. *Strategic Management Journal*, 31(3), 306-333. Scopus. <https://doi.org/10.1002/smj.821>
- Ahuja, G., Lampert, C. M., & Tandon, V. (2008). Moving Beyond Schumpeter: Management Research on the Determinants of Technological Innovation. *Academy of Management Annals*, 2, 1-98. <https://doi.org/10.1080/19416520802211446>
- Aithal, P. K., Acharya, U. D., Geetha, M., Sagar, R., & Abraham, R. (2022). *A Comparative Study of Deep Neural Network and Statistical Models for Stock Price Prediction*. 2022 3rd International Conference for Emerging Technology, INCET 2022. Scopus.
<https://doi.org/10.1109/INCET54531.2022.9824487>
- Aizstrauta, D., Ginters, E., & Eroles, M.-A. P. (2015). *Applying theory of diffusion of innovations to evaluate technology acceptance and sustainability*. 43, 69-77. Scopus.
<https://doi.org/10.1016/j.procs.2014.12.010>

- Altuntas, S., Dereli, T., & Kusiak, A. (2015). Forecasting technology success based on patent data. *Technological Forecasting and Social Change*, *96*, 202-214. Scopus.
<https://doi.org/10.1016/j.techfore.2015.03.011>
- Anderson, C. R., & Zeithaml, C. P. (1984). Stage of the Product Life Cycle, Business Strategy, and Business Performance. *The Academy of Management Journal*, *27*(1), 5-24. JSTOR.
<https://doi.org/10.2307/255954>
- Anderson, P., & Tushman, M. L. (1990). Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change. *Administrative Science Quarterly*, *35*(4), 604-633. JSTOR. <https://doi.org/10.2307/2393511>
- Aradi, M. A., & Hewahi, N. (2020). *Prediction of Stock Price and Direction Using Neural Networks: Datasets Hybrid Modeling Approach*. 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy, ICDABI 2020. Scopus. <https://doi.org/10.1109/ICDABI51230.2020.9325697>
- Bass, F. M. (1969a). A New Product Growth for Model Consumer Durables. *Management Science*, *15*(5), Art. 5. JSTOR.
- Bass, F. M. (1969b). A New Product Growth for Model Consumer Durables. *Management Science*, *15*(5), 215-227. JSTOR.
- Bass, F. M. (2004). Comments on "A New Product Growth for Model Consumer Durables The Bass Model". *Management Science*, *50*(12_supplement), 1833-1840.
<https://doi.org/10.1287/mnsc.1040.0300>
- Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, *23*(1), 4-17. Scopus.
<https://doi.org/10.1016/j.intcom.2010.07.003>

- Becker, M. C., Knudsen, T., & Swedberg, R. (2012). Schumpeter's Theory of Economic Development: 100 years of development. *Journal of Evolutionary Economics*, 22(5), 917-933. <https://doi.org/10.1007/s00191-012-0297-x>
- Bergek, A., Jacobsson, S., Carlsson, B., Lindmark, S., & Rickne, A. (2008). Analyzing the functional dynamics of technological innovation systems: A scheme of analysis. *Research Policy*, 37(3), 407-429. Scopus. <https://doi.org/10.1016/j.respol.2007.12.003>
- Block, J. H., Fisch, C. O., & van Praag, M. (2017). The Schumpeterian entrepreneur: A review of the empirical evidence on the antecedents, behaviour and consequences of innovative entrepreneurship. *Industry and Innovation*, 24(1), 61-95. <https://doi.org/10.1080/13662716.2016.1216397>
- Box, G., & Jenkins, G. (1970). *Time series analysis Forecasting and control*. Wiley. <http://entsphere.com/pub/pdf/Papers/Signal%20Processing/Box,%20Time%20Series%20Analysis,%20Forecasting%20and%20control%20BOOK.pdf>
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees*. Routledge; <https://doi.org/10.1201/9781315139470>.
- Chang, S.-H., & Fan, C.-Y. (2015). Telematics technology development forecasting: The patent analysis and technology life cycle perspective. *Lecture Notes in Electrical Engineering*, 349, 149-158. Scopus. https://doi.org/10.1007/978-3-662-47200-2_17
- Chang, S.-H., & Fan, C.-Y. (2016). Identification of the technology life cycle of telematics a patent-based analytical perspective. *Technological Forecasting and Social Change*, 105, 1-10. Scopus. <https://doi.org/10.1016/j.techfore.2016.01.023>

- Chen, L., & Klasky, H. B. (2022). *Six Machine-Learning Methods for Predicting Hospital-Stay Duration for Patients with Sepsis: A Comparative Study*. 2022-March, 302-309. Scopus. <https://doi.org/10.1109/SoutheastCon48659.2022.9764052>
- Christensen, C. M. (1992a). EXPLORING THE LIMITS OF THE TECHNOLOGY S-CURVE. PART I: COMPONENT TECHNOLOGIES. *Production and Operations Management*, *1*(4), Art. 4. Scopus. <https://doi.org/10.1111/j.1937-5956.1992.tb00001.x>
- Christensen, C. M. (1992b). EXPLORING THE LIMITS OF THE TECHNOLOGY S-CURVE. PART I: COMPONENT TECHNOLOGIES. *Production and Operations Management*, *1*(4), 334-357. Scopus. <https://doi.org/10.1111/j.1937-5956.1992.tb00001.x>
- Christensen, C. M. (1992c). EXPLORING THE LIMITS OF THE TECHNOLOGY S-CURVE. PART II: ARCHITECTURAL TECHNOLOGIES. *Production and Operations Management*, *1*(4), 358-366. Scopus. <https://doi.org/10.1111/j.1937-5956.1992.tb00002.x>
- Christensen, C. M., & Bower, J. L. (1996). Customer power, strategic investment, and the failure of leading firms. *Strategic Management Journal*, *17*(3), 197-218. Scopus.
- Daim, T. U., Chiavetta, D., Porter, A., & Saritas, O. (2016). *Anticipating future innovation pathways through large data analysis*. Springer.
- Darwin, C. (1859). *EL ORIGEN DE LAS ESPECIES POR LA SELECCION NATURAL* (2.^a-1963.^a ed.). https://books.google.es/books?hl=es&lr=&id=pKNNeMliJRwC&oi=fnd&pg=PA5&dq=e+l+origen+de+las+especies&ots=yI0rsOTXo4&sig=pxXy1bdtkvSG8_oUj8cw603sORo#v=onepage&q=el%20origen%20de%20las%20especies&f=false
- Dean, J. (1950a). Pricing Policies for New Products. *Harvard Business Review*, November 1976, Art. November 1976. <https://hbr.org/1976/11/pricing-policies-for-new-products>

- Dean, J. (1950b). Pricing Policies for New Products. *Harvard Business Review*, November 1976.
<https://hbr.org/1976/11/pricing-policies-for-new-products>
- Dosi, G. (1982). Technological paradigms and technological trajectories. A suggested interpretation of the determinants and directions of technical change. *Research Policy*, 11(3), 147-162. Scopus. [https://doi.org/10.1016/0048-7333\(82\)90016-6](https://doi.org/10.1016/0048-7333(82)90016-6)
- Emery, F. E., & Trist, E. L. (1960). Socio-technical Systems. *Management Sciences Models and Techniques*, vol 2.
- Faber, A., & Frenken, K. (2009). Models in evolutionary economics and environmental policy: Towards an evolutionary environmental economics. *Technological Forecasting and Social Change*, 76(4), 462-470. <https://doi.org/10.1016/j.techfore.2008.04.009>
- Foster, R. N. (1985a). Timing technological transitions. *Technology in Society*, 7(2-3), Art. 2-3. Scopus.
- Foster, R. N. (1985b). Timing technological transitions. *Technology in Society*, 7(2-3), 127-141. Scopus.
- Fox, H. W. (1973). MARKETING: THE BASICS. *IEEE Engineering Management Review*, 1(1), 31-37. Scopus.
- Freeman, C. (1994). The economics of technical change. *Cambridge Journal of Economics*, 18(5), 463-514. Scopus. <https://doi.org/10.1093/oxfordjournals.cje.a035286>
- Freeman, C. (2007). A Schumpeterian renaissance? En *Elgar Companion to Neo-Schumpeterian Economics* (pp. 130-141). Scopus.
- Freeman, C., Clark, J., & Soete, L. (1982a). *Unemployment and Technical Innovation: A study of long waves and economic development*. <http://collections.unu.edu/view/UNU:672>

- Freeman, C., Clark, J., & Soete, L. (1982b). *Unemployment and Technical Innovation: A study of long waves and economic development*. <http://collections.unu.edu/view/UNU:672>
- Fuenfschilling, L., & Truffer, B. (2014). The structuration of socio-technical regimes—Conceptual foundations from institutional theory. *Research Policy*, 43(4), 772-791. Scopus. <https://doi.org/10.1016/j.respol.2013.10.010>
- Fujimoto, T. (2007). Architecture-Based Comparative Advantage—A Design Information View of Manufacturing. *Evolutionary and Institutional Economics Review*, 4(1), 55-112. <https://doi.org/10.14441/eier.4.55>
- Gao, L., Porter, A. L., Wang, J., Fang, S., Zhang, X., Ma, T., Wang, W., & Huang, L. (2013). Technology life cycle analysis method based on patent documents. *Technological Forecasting and Social Change*, 80(3), 398-407. Scopus. <https://doi.org/10.1016/j.techfore.2012.10.003>
- Garcia, J., Molina, J. M., Berlanga, A., Patricio, M., Bustamante, A., & Padilla, W. (2018). *Ciencia de datos, tecnicas analiticas y de aprendizaje estadistico*. Alfaomega.
- Geels, F. W. (2004). From sectoral systems of innovation to socio-technical systems: Insights about dynamics and change from sociology and institutional theory. *Research Policy*, 33(6-7), 897-920. Scopus. <https://doi.org/10.1016/j.respol.2004.01.015>
- Geels, F. W. (2005a). Co-evolution of technology and society: The transition in water supply and personal hygiene in the Netherlands (1850-1930) - A case study in multi-level perspective. *Technology in Society*, 27(3), Art. 3. Scopus. <https://doi.org/10.1016/j.techsoc.2005.04.008>
- Geels, F. W. (2005b). Co-evolution of technology and society: The transition in water supply and personal hygiene in the Netherlands (1850-1930) - A case study in multi-level

perspective. *Technology in Society*, 27(3), 363-397. Scopus.

<https://doi.org/10.1016/j.techsoc.2005.04.008>

Geels, F. W. (2005c). The dynamics of transitions in socio-technical systems: A multi-level analysis of the transition pathway from horse-drawn carriages to automobiles (1860-1930). *Technology Analysis and Strategic Management*, 17(4), Art. 4. Scopus.

<https://doi.org/10.1080/09537320500357319>

Geels, F. W. (2005d). The dynamics of transitions in socio-technical systems: A multi-level analysis of the transition pathway from horse-drawn carriages to automobiles (1860-1930). *Technology Analysis and Strategic Management*, 17(4), 445-476. Scopus.

<https://doi.org/10.1080/09537320500357319>

Geels, F. W., & Kemp, R. (2007). Dynamics in socio-technical systems: Typology of change processes and contrasting case studies. *Technology in Society*, 29(4), 441-455. Scopus.

<https://doi.org/10.1016/j.techsoc.2007.08.009>

Geels, F. W., & Raven, R. (2006a). Non-linearity and expectations in niche-development trajectories: Ups and downs in Dutch biogas development (1973-2003). *Technology Analysis and Strategic Management*, 18(3-4), Art. 3-4. Scopus.

<https://doi.org/10.1080/09537320600777143>

Geels, F. W., & Raven, R. (2006b). Non-linearity and expectations in niche-development trajectories: Ups and downs in Dutch biogas development (1973-2003). *Technology Analysis and Strategic Management*, 18(3-4), 375-392. Scopus.

<https://doi.org/10.1080/09537320600777143>

Geels, F. W., & Schot, J. (2007). Typology of sociotechnical transition pathways. *Research Policy*, 36(3), 399-417. Scopus. <https://doi.org/10.1016/j.respol.2007.01.003>

- Gerón, A. (2019). *Hands on machine learning with Scikit-Learn, Keras & TensorFlow*. O’reilly.
- Godin, B. (2008). In the shadow of schumpeter: W. Rupert Maclaurin and the study of technological innovation. *Minerva*, 46(3), 343-360. <https://doi.org/10.1007/s11024-008-9100-4>
- Haupt, R., Kloyer, M., & Lange, M. (2007). Patent indicators for the technology life cycle development. *Research Policy*, 36(3), 387-398. Scopus. <https://doi.org/10.1016/j.respol.2006.12.004>
- Hekkert, M. P., & Negro, S. O. (2009). Functions of innovation systems as a framework to understand sustainable technological change: Empirical evidence for earlier claims. *Technological Forecasting and Social Change*, 76(4), 584-594. Scopus. <https://doi.org/10.1016/j.techfore.2008.04.013>
- Hekkert, M. P., Suurs, R. A. A., Negro, S. O., Kuhlmann, S., & Smits, R. E. H. M. (2007). Functions of innovation systems: A new approach for analysing technological change. *Technological Forecasting and Social Change*, 74(4), 413-432. Scopus. <https://doi.org/10.1016/j.techfore.2006.03.002>
- Hingley, P., & Dikta, G. (2022). Use of Information Criteria for finding Box-Jenkins Time Series Models for Patent Filings Counts Forecasts. *IAENG International Journal of Applied Mathematics*, 52(1). Scopus.
- Huenteler, J., Schmidt, T. S., Ossenbrink, J., & Hoffmann, V. H. (2016a). Technology life-cycles in the energy sector—Technological characteristics and the role of deployment for innovation. *Technological Forecasting and Social Change*, 104, 102-121. <https://doi.org/10.1016/j.techfore.2015.09.022>

- Huenteler, J., Schmidt, T. S., Ossenbrink, J., & Hoffmann, V. H. (2016b). Technology life-cycles in the energy sector—Technological characteristics and the role of deployment for innovation. *Technological Forecasting and Social Change*, *104*, 102-121.
<https://doi.org/10.1016/j.techfore.2015.09.022>
- Hummon, N. P., & Dereian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks*, *11*(1), 39-63. Scopus. [https://doi.org/10.1016/0378-8733\(89\)90017-8](https://doi.org/10.1016/0378-8733(89)90017-8)
- Jamison, A., & Hård, M. (2003). The story-lines of technological change: Innovation, construction and appropriation. *Technology Analysis and Strategic Management*, *15*(1), 81-92. Scopus. <https://doi.org/10.1080/0953732032000046060>
- Järvenpää, H. M., Mäkinen, S. J., & Seppänen, M. (2011). Patent and publishing activity sequence over a technology's life cycle. *Technological Forecasting and Social Change*, *78*(2), 283-293. Scopus. <https://doi.org/10.1016/j.techfore.2010.06.020>
- Jha, S., Yang, E., Almagrabi, A. O., Bashir, A. K., & Joshi, G. P. (2021). Comparative analysis of time series model and machine testing systems for crime forecasting. *Neural Computing and Applications*, *33*(17), 10621-10636. Scopus.
<https://doi.org/10.1007/s00521-020-04998-1>
- Jones, G., Schilling, M., & Hill, C. (1998). *Strategic Management: An Integrated Approach*, 12e. https://www.researchgate.net/publication/298791895_Strategic_Management_An_Integrated_Approach_12e
- Jun, S., Park, S. S., & Jang, D. S. (2012). Patent Management for Technology Forecasting: A Case study of the bio-industry. *Journal of Intellectual Property Rights*, *17*(6), 539-546. Scopus.

- Kaplan, S., & Tripsas, M. (2008). Thinking about technology: Applying a cognitive lens to technical change. *Research Policy*, 37(5), 790-805.
<https://doi.org/10.1016/j.respol.2008.02.002>
- Kemp, R., Loorbach, D., & Rotmans, J. (2007). Transition management as a model for managing processes of co-evolution towards sustainable development. *International Journal of Sustainable Development and World Ecology*, 14(1), 78-91. Scopus.
<https://doi.org/10.1080/13504500709469709>
- Kim, D.-J., & Kogut, B. (1996). Technological Platforms and Diversification. *Organization Science*, 7(3), 283-301. Scopus. <https://doi.org/10.1287/orsc.7.3.283>
- Kivimaa, P., & Kern, F. (2016). Creative destruction or mere niche support? Innovation policy mixes for sustainability transitions. *Research Policy*, 45(1), 205-217. Scopus.
<https://doi.org/10.1016/j.respol.2015.09.008>
- Klepper, S. (1997). Industry Life Cycles. *Industrial and Corporate Change*, 6(1), 145-182.
<https://doi.org/10.1093/icc/6.1.145>
- Korotayev, A., Zinkina, J., & Bogevolnov, J. (2011). Kondratieff waves in global invention activity (1900-2008). *Technological Forecasting and Social Change*, 78(7), 1280-1284.
<https://doi.org/10.1016/j.techfore.2011.02.011>
- Kuhn, T. S. (1962). *The structure of scientific revolutions* (Vol. 1).
- Lee, C., Kim, J., Noh, M., Woo, H.-G., & Gang, K. (2017). Patterns of technology life cycles: Stochastic analysis based on patent citations. *Technology Analysis and Strategic Management*, 29(1), 53-67. Scopus. <https://doi.org/10.1080/09537325.2016.1194974>

- Lee, J., & Berente, N. (2013). The era of incremental change in the technology innovation life cycle: An analysis of the automotive emission control industry. *Research Policy*, *42*(8), 1469-1481. <https://doi.org/10.1016/j.respol.2013.05.004>
- Levitt, T. (1965a). Exploit the Product Life Cycle. *Graduate School of Business Administration, Harvard University.*, Vol. 43.
http://marketch.cloud/pdf/Exploit_the_Product_Life_Cycle.pdf
- Levitt, T. (1965b). Exploit the Product Life Cycle. *Graduate School of Business Administration, Harvard University.*, Vol. 43.
http://marketch.cloud/pdf/Exploit_the_Product_Life_Cycle.pdf
- Leydesdorff, L. (2015). Can technology life-cycles be indicated by diversity in patent classifications? The crucial role of variety. *Scientometrics*, *105*(3), 1441-1451. Scopus.
<https://doi.org/10.1007/s11192-015-1639-x>
- Li, X., Yin, Y., Manrique, D. V., & Bäck, T. (2021). Lifecycle forecast for consumer technology products with limited sales data. *International Journal of Production Economics*, *239*. Scopus. <https://doi.org/10.1016/j.ijpe.2021.108206>
- Little, A. D. (1981). *The Strategic Management of Technology*. Arthur D. Little.
- Liu, Y., & Zhao, M. (2022). An obsolescence forecasting method based on improved radial basis function neural network. *Ain Shams Engineering Journal*, *13*(6). Scopus.
<https://doi.org/10.1016/j.asej.2022.101775>
- Maclaurin, W. R. (1950). The Process of Technological Innovation: The Launching of a New Scientific Industry. *The American Economic Review*, *40*(1), 90-112. JSTOR.

- Maclaurin, W. R. (1953). The sequence from invention to innovation and its relation to economic growth. *Quarterly Journal of Economics*, 67(1), 97-111. Scopus.
<https://doi.org/10.2307/1884150>
- Maclaurin, W. R. (1954). Technological Progress in Some American Industries. *The American Economic Review*, 44(2), 178-189. JSTOR.
- Martínez, C. (1998). *Estadística y muestreo* (9.^a ed.). Ecoe ediciones.
- McKinney, W. (2017). *Python for data analysis*. O'reilly.
- Medina, F., & Galván, M. (2007). *Imputación de datos: Teoría y práctica*. Naciones Unidas.
https://repositorio.cepal.org/bitstream/handle/11362/4755/1/S0700590_es.pdf
- Mina, A., Ramlogan, R., Tampubolon, G., & Metcalfe, J. S. (2007). Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge. *Research Policy*, 36(5), 789-806. Scopus. <https://doi.org/10.1016/j.respol.2006.12.007>
- Muller, A., & Guido, S. (2016). *Introduction to machine learning with Python*. O'reilly.
- Mumford, E. (2006). The story of socio-technical design: Reflections on its successes, failures and potential. *Information Systems Journal*, 16(4), 317-342. Scopus.
<https://doi.org/10.1111/j.1365-2575.2006.00221.x>
- Murphy, J. T. (2015). Human geography and socio-technical transition studies: Promising intersections. *Environmental Innovation and Societal Transitions*, 17, 73-91. Scopus.
<https://doi.org/10.1016/j.eist.2015.03.002>
- Nawaz Sharif, M., & Kabir, C. (1976). A generalized model for forecasting technological substitution. *Technological Forecasting and Social Change*, 8(4), 353-364.
[https://doi.org/10.1016/0040-1625\(76\)90027-5](https://doi.org/10.1016/0040-1625(76)90027-5)

- Nelson, R. R., & Winter, S. G. (1982). *An Evolutionary Theory of Economic Change—Richard R. Nelson, Sidney G. Winter / Harvard University Press*.
<http://www.hup.harvard.edu/catalog.php?isbn=9780674272286>
- Nieto, M., López, F., & Cruz, F. (1998). Performance analysis of technology using the S curve model: The case of digital signal processing (DSP) technologies. *Technovation*, 18(6), 439-457. [https://doi.org/10.1016/S0166-4972\(98\)00021-2](https://doi.org/10.1016/S0166-4972(98)00021-2)
- OCDE. (2018). *Manual de Oslo*.
- Oliveira, A. S., Santos, R. O. D., Silva, B. C. D. S., Guarieiro, L. L. N., Angerhausen, M., Reisgen, U., Sampaio, R. R., Machado, B. A. S., Droguett, E. L., Da Silva, P. H. F., & Coelho, R. S. (2021). A detailed forecast of the technologies based on lifecycle analysis of gma and cmtwelding processes. *Sustainability (Switzerland)*, 13(7). Scopus.
<https://doi.org/10.3390/su13073766>
- Parvin, A. J., Jr., Beruvides, M. G., & Tercero-Gómez, V. G. (2022). In Situ Technological Innovation Diffusion Rate Accuracy Assessment. *Systems*, 10(2). Scopus.
<https://doi.org/10.3390/systems10020025>
- Pavitt, K. (1982a). R&D, patenting and innovative activities. A statistical exploration. *Research Policy*, 11(1), Art. 1. Scopus. [https://doi.org/10.1016/0048-7333\(82\)90005-1](https://doi.org/10.1016/0048-7333(82)90005-1)
- Pavitt, K. (1982b). R&D, patenting and innovative activities. A statistical exploration. *Research Policy*, 11(1), 33-51. Scopus. [https://doi.org/10.1016/0048-7333\(82\)90005-1](https://doi.org/10.1016/0048-7333(82)90005-1)
- Pavitt, K. (1984a). Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy*, 13(6), Art. 6. Scopus. [https://doi.org/10.1016/0048-7333\(84\)90018-0](https://doi.org/10.1016/0048-7333(84)90018-0)
- Pavitt, K. (1984b). Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy*, 13(6), 343-373. Scopus. [https://doi.org/10.1016/0048-7333\(84\)90018-0](https://doi.org/10.1016/0048-7333(84)90018-0)

- Pérez. (2009). Technological revolutions and techno-economic paradigms. *Cambridge Journal of Economics*, 34(1), 185-202. Scopus. <https://doi.org/10.1093/cje/bep051>
- Pérez. (2012). Revoluciones tecnológicas y paradigmas tecnoeconómicos. *Tecnología y Construcción*, 21(1), Art. 1. http://saber.ucv.ve/ojs/index.php/rev_tc/article/view/2886
- Perez, C. (2010). Technological Revolutions and Techno-economic paradigms. *Cambridge Journal of Economics*, 34(1), 185-202.
- Ramadhan, M. H., Malik, V. I., & Sjafrizal, T. (2018). *Artificial neural network approach for technology life cycle construction on patent data*. 499-503. Scopus. <https://doi.org/10.1109/IEA.2018.8387152>
- Raschka, S., & Mirjalili, V. (2019). *Python machine learning*. Marcombo.
- Ravena, R., Schota, J., & Berkhoutb, F. (2012a). Space and scale in socio-Technical transitions. *Environmental Innovation and Societal Transitions*, 4, 63-78. Scopus. <https://doi.org/10.1016/j.eist.2012.08.001>
- Ravena, R., Schota, J., & Berkhoutb, F. (2012b). Space and scale in socio-Technical transitions. *Environmental Innovation and Societal Transitions*, 4, 63-78. Scopus. <https://doi.org/10.1016/j.eist.2012.08.001>
- Rink, D. R., & Swan, J. E. (1979). Product life cycle research: A literature review. *Journal of Business Research*, 7(3), 219-242. Scopus. [https://doi.org/10.1016/0148-2963\(79\)90030-4](https://doi.org/10.1016/0148-2963(79)90030-4)
- Robledo, J. (2013). *Introducción a la Gestión de la Tecnología y la Innovación*. Universidad Nacional de Colombia Sede–Medellín.
- Rogers, E. M. (1976a). New Product Adoption and Diffusion. *Journal of Consumer Research*, 2(4), Art. 4. JSTOR.

- Rogers, E. M. (1976b). New Product Adoption and Diffusion. *Journal of Consumer Research*, 2(4), 290-301. JSTOR.
- Rogers, E. M. (1983). *Diffusion of innovations* (3rd ed). Free Press ; Collier Macmillan.
- Roussel, P. A. (1984). Technological Maturity Proves a Valid and Important Concept. *Research Management*, 27(1), 29-34. <https://doi.org/10.1080/00345334.1984.11756815>
- Schumpeter, J. A. (1935a). A Theorist's Comment on the Current Business Cycle. *Journal of the American Statistical Association*, 30(189), 167-168. Scopus. <https://doi.org/10.1080/01621459.1935.10504150>
- Schumpeter, J. A. (1935b). *BUSINESS CYCLES. A Theoretical, Historical and Statistical Analysis of the Capitalist Process*. 385.
- Schumpeter, J. A. (1939). *BUSINESS CYCLES. A Theoretical, Historical and Statistical Analysis of the Capitalist Process*. (1.^a ed.).
- Smith, M., & Agrawal, R. (2015). *A comparison of time series model forecasting methods on patent groups*. 1353, 167-173. Scopus.
- Sokmen, N., & Petrov, A. E. (2021). *Forecasting of Radar Technologies Using Patent Information*. 15th IEEE International Conference on Application of Information and Communication Technologies, AICT 2021. Scopus. <https://doi.org/10.1109/AICT52784.2021.9620514>
- Sood, A., & Tellis, G. J. (2005a). Technological evolution and radical innovation. *Journal of Marketing*, 69(3), 152-168. Scopus. <https://doi.org/10.1509/jmkg.69.3.152.66361>
- Sood, A., & Tellis, G. J. (2005b). Technological evolution and radical innovation. *Journal of Marketing*, 69(3), Art. 3. Scopus. <https://doi.org/10.1509/jmkg.69.3.152.66361>

- Taylor, M., & Taylor, A. (2012). The technology life cycle: Conceptualization and managerial implications. *International Journal of Production Economics*, 140(1), 541-553. Scopus.
<https://doi.org/10.1016/j.ijpe.2012.07.006>
- Utterback, J. M., & Abernathy, W. J. (1975). A dynamic model of process and product innovation. *Omega*, 3(6), 639-656. [https://doi.org/10.1016/0305-0483\(75\)90068-7](https://doi.org/10.1016/0305-0483(75)90068-7)
- Velásquez, M. A., & Martínez, J. (2009). Estimacion de observaciones faltantes en series de tiempo usando metodos multivariados con restricciones. *Comunicaciones en Estadística*, 2(1), 17.
- Veraszto, E. V., Silva, D. da, Miranda, N. A., & Simon, F. O. (2009). Tecnologia: Buscando uma definição para o conceito. *PRISMA.COM*, 0(8), 19-46.
- Vermeulen, B., Pyka, A., La Poutre, J. A., & de Kok, A. G. (2018). Capability-based governance patterns over the product life-cycle: An agent-based model. *Journal of Economic Interaction and Coordination*, 13(2), 311-349. <https://doi.org/10.1007/s11403-016-0184-x>
- Vernon, R. (1966). International investment and international trade in the product cycle. *Quarterly Journal of Economics*, 80(2), 190-207. Scopus.
<https://doi.org/10.2307/1880689>
- Verspagen, B. (2007). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems*, 10(1), 93-115. Scopus.
<https://doi.org/10.1142/S0219525907000945>
- Vora, V., Shah, M., Chouhan, A., & Tawde, P. (2022). Stock Market Prices and Returns Forecasting Using Deep Learning Based on Technical and Fundamental Analysis.

Lecture Notes in Networks and Systems, 401, 717-728. Scopus.

https://doi.org/10.1007/978-981-19-0098-3_68

Wasson, C. R. (1976). The importance of the product life cycle to the industrial marketer.

Industrial Marketing Management, 5(6), 299-308. Scopus. [https://doi.org/10.1016/0019-8501\(76\)90012-2](https://doi.org/10.1016/0019-8501(76)90012-2)

Wells, L. T. (1968). A product life cycle for international trade? *Journal of Marketing*, 32 (3), 1-6.

Windrum, P., & Garcia-Goni, M. (2008). A neo-Schumpeterian model of health services

innovation. *Research Policy*, 37(4), 649-672. <https://doi.org/10.1016/j.respol.2007.12.011>

WIPO. (2019). *International Patent Classification (IPC)*. World Intellectual Property Organization.

You, H., Li, M., Hipel, K. W., Jiang, J., Ge, B., & Duan, H. (2017). Development trend

forecasting for coherent light generator technology based on patent citation network analysis. *Scientometrics*, 111(1), 297-315. Scopus. <https://doi.org/10.1007/s11192-017-2252-y>

Anexo - Divulgación

Ponencia 1

Título: *Technology lifecycle and innovation trajectories*

Evento: *Science and Lunch for You - Business School Berlin*

País: Alemania.

Fecha: 9 de julio de 2020.

Ponencia 2

Título: Modelo de predicción del cambio incremental

Evento: XIX Congreso Latino-Iberoamericano de Gestión Tecnológica y de la Innovación - ALTEC 2021

País: Perú

Fecha: 29 de octubre de 2021

Proyecto de grado Maestría

Título: Propuesta modelo de predicción del cambio incremental dentro del ciclo de vida tecnológico

Programa: Maestría en gestión tecnológica- Universidad Pontificia Bolivariana

País: Colombia

Fecha: 12 de octubre de 2021

Ponencia y publicación

Título: *Predicting incremental technological change based the technological life cycle*

Evento: IAMOT 2023

País: Brasil

Fecha: Aceptado para ponencia. Abril-Mayo 2023

Pasantía

Universidad de negocios y leyes de Berlín - BSP

Fecha: 4 de octubre de 2021 a 2 de junio de 2022

Redes académicas

International Entrepreneurship - Lab Smart Money – IELSM

Coordinador académico plataforma

Universidad de negocios y leyes de Berlín (Alemania) y Universidad Tecnológica de Querétaro (México)

Anexo – Proceso metodológico

