



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



**TECNOLÓGICO
NACIONAL DE MÉXICO**

INSTITUTO TECNOLÓGICO DE CIUDAD MADERO

División De Estudios De Posgrado E Investigación



INFORME FINAL DE AÑO SABÁTICO

DESARROLLO DE UN METODO DE APRENDIZAJE QUE RESUELVA EL
PROBLEMA DE PREDICCIÓN PARA PROBLEMAS DE SERIES DE TIEMPO
DE CASOS DE COMPETENCIA INTERNACIONAL TIPO M4

Que presenta

Dr. Juan Frausto Solis
1509 CVU Conahcyt 31308

Realizado en
Instituto Tecnológico De Ciudad Madero

Durante el Periodo
01/SEP/2023 AL 31/AGO/2024

Dictamen No.: AS-2-125/2023

Ciudad de México, **28 de agosto de 2023**

AUTORIZACIÓN No AS-2-125/2023

LUCIANO AGUILERA VÁZQUEZ
DIRECTOR DEL INSTITUTO TECNOLÓGICO
DE CIUDAD MADERO
PRESENTE

Con fundamento en el Reglamento Interior de Trabajo del Personal Docente de los Institutos Tecnológicos, en las Políticas Académicas Generales del Periodo Sabático y las bases de la Convocatoria correspondiente, el Tecnológico Nacional de México a través de la Comisión Central, ha revisado la solicitud del **Año Sabático** del **C. JUAN FRAUSTO SOLÍS** en el Programa de Investigación Científica, Tecnológica o Educativa; bajo la modalidad: **A.1. Proyecto de Investigación Científica: Desarrollo de un método de aprendizaje que resuelva el problema de predicción para problemas de series de tiempo de casos de competencia internacional tipo M4**, se tiene a bien emitir el siguiente dictamen de:

AUTORIZACIÓN

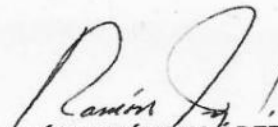
Del programa presentado a partir del **01 de septiembre de 2023** hasta el **31 de agosto de 2024**, quedando su seguimiento bajo la responsabilidad de la Dirección a su cargo, debiendo recibir y evaluar por parte de la Academia correspondiente y de la Comisión Dictaminadora, el reporte intermedio y final de las actividades desarrolladas durante el ejercicio del periodo sabático autorizado al docente. Así mismo, se deberá remitir a la Secretaría Académica, de Investigación e Innovación a través de la Dirección de Docencia e Innovación Educativa del Tecnológico Nacional de México, la evaluación de dichos reportes que comprenderán la totalidad de lo establecido en la calendarización, que se encuentra en el expediente de la solicitud del docente; además de un ejemplar en físico y en electrónico (en formato PDF) del producto final comprometido.



M00-PR-04-R08



Av. Universidad 1200, Col. Xoco, Alcaldía Benito Juárez, C.P. 06330 Ciudad de México
Tel. (55) 3601-7500, Ext. 65050, e-mail: direccion@tecnm.mx


RAMÓN JIMÉNEZ LÓPEZ
DIRECTOR GENERAL
TECNOLÓGICO NACIONAL DE MÉXICO


SECRETARÍA DE EDUCACIÓN PÚBLICA
TECNOLÓGICO NACIONAL
DE MÉXICO
DIRECCIÓN GENERAL



Versión 4



ÍNDICE DE CONTENIDO

1. IDENTIFICACIÓN DEL PROYECTO	3
1.1. DATOS GENERALES DEL PROYECTO	3
1.2. PROGRAMA DE ACTIVIDADES Y CALENDARIZACIÓN	3
1.3. PRODUCTOS ENTREGALES	3
2. RESUMEN DEL PROYECTO	4
3. INTRODUCCIÓN: EL PROBLEMA	5
3.2 OBJETIVOS	6
3.2.1 OBJETIVO GENERAL	6
3.2.2 OBJETIVOS ESPECÍFICOS	6
4. MARCO TEÓRICO	7
5. ACTIVIDADES	8
5.1 IDENTIFICACIÓN DE MÉTODOS INDIVIDUALES	8
5.1.1 NAIVE	9
5.1.2 ARIMA, AUTOARIMA Y NNETAR	9
5.1.3 ETS.....	9
5.1.4 SMYL.....	10
5.1.5 TBATS	10
5.1.6 STL-ar y RW-Drift.....	11
5.1.7 THETA.....	11
5.1.8 HOLT	12
5.1.9 FFORMA Y JAGANATHAN.....	13
5.1.10 BAGGING	13
5.2 DETERMINACIÓN DE MÉTODOS QUE CONFORMEN EL ENSAMBLE	13
5.3 DESARROLLO DEL MODELO DE OPTIMIZACIÓN.....	14
6. LOGROS OBTENIDOS DURANTE EL PERIODO SABÁTICO	15
6.1 ARTÍCULOS DE DIVULGACIÓN	15
6.2 PUBLICACIONES EN JOURNALS.....	15
6.3 CAPÍTULOS DE LIBRO.....	15
6.4 TESIS DE MAESTRÍA	15
REFERENCIAS	16

INFORME FINAL DE PERIODO SABÁTICO

1. IDENTIFICACIÓN DEL PROYECTO

1.1. DATOS GENERALES DEL PROYECTO

NOMBRE DE LA INSTITUCIÓN: Instituto Tecnológico de Ciudad Madero	
Responsable del Proyecto: Juan Frausto Solís Correo: juan.fs@cdmadero.tecnm.mx	Título del Proyecto: Desarrollo de un Método de Aprendizaje que resuelva el problema de predicción para problemas de series de tiempo de casos de competencia internacional tipo M4
Tipo de investigación: Científica (x) Tecnológica () Educativa ()	
Programa educativo en donde se desarrollará el proyecto: MCING-2011-45 Maestría en Ciencias de la Ingeniería	
Línea de investigación: Simulación y Cómputo Aplicado CMD-LGAC-01-17	

Area

Ciencias Químicas	()	Ingeniería Industrial, Administrativa y Desarrollo Regional	()
Ciencias Biológicas	()	Ciencias de la Educación	()
Ciencias de la Tierra y del Medio Ambiente	(x)	Ciencias del Mar	()

1.2. PROGRAMA DE ACTIVIDADES Y CALENDARIZACIÓN

ACTIVIDAD	MESES											
	1	2	3	4	5	6	7	8	9	10	11	12
A. Identificar métodos individuales.	XX XX	XX XX	XX XX									
B. Determinar métodos de pronóstico individuales				XX XX	XX XX							
C. Modelo de optimización para maximizar precisión						XX XX	XX XX					
D. Conformar método ensamble								XX	XX	XX	XX	XX
E. Difundir resultados.		XX	XX								XX	XX
ENTREGA REPORTE FINAL 100%	() Semestre Sabático: 1 septiembre/2023 -29 febrero/2024 (X) Año Sabático: 1 Septiembre/2023- 30 agosto/ 2024											

XX Actividad programada XX Actividad terminada PX Actividad en proceso (fecha de terminación planeada no se ha cumplido)

1.3. PRODUCTOS ENTREGABLES

PRODUCTOS ENTREGABLES DEL INICIO AL FINAL DEL PROYECTO			
Contribución en la formación de recursos humanos	Cant .	Productividad académica	Ca nt.
Incorporación de estudiantes de licenciatura al proyecto		Artículos científicos en revistas indizadas enviados:	
Incorporación estudiantes de servicio social:		Artículos científicos en revistas arbitradas enviados:	
Estudiantes con residencia concluida:		Artículos de divulgación enviados:	
Tesis en desarrollo de Licenciatura:		Artículos científicos en revistas indizadas publicados:	
Tesis concluida de Licenciatura:		Artículos científicos en revistas arbitradas publicados:	
Tesis en desarrollo de Maestría:		Artículos de divulgación publicados:	2
Tesis concluida de Maestría:	1	Memorias en extenso en congresos:	
Tesis en desarrollo de Doctorado:		Capítulos de libros enviados publicados	1

2. RESUMEN DEL PROYECTO

Predicción es un problema de relevancia en la ciencia e ingeniería que la humanidad ha tenido desde hace siglos. Una de las primeras estrategias planteadas fue consultar a expertos sobre su opinión de cual sería el valor de las variables en cuestión. Otra posibilidad es plantear modelos analíticos considerando el tiempo como variable dependiente.

Una estrategia de expertos, presenta dificultades debido a varios rubros: a) Vinculada con la identificación de expertos y obtención de su conomiento. En este rubro la primera dificultad es que no existe un método científico para demostrar que una persona realmente sea un experto; la segunda es que aún cuando, se acepte sin conceder, que el considerado experto realmente lo sea, no existe un mecanismo de garantizar que la opinión emitida por el experto realmente es una consecuencia lógicas con el razonamiento del mencionado experto. solución

Un método en base a modelos analíticos, implica considerar variaciones con respecto al tiempo de las variables involucradas; esos modelos analíticos llegan comunmente a plantear ecuaciones en derivadas las cuales consideran diferenciales con respecto al tiempo t y otras variables X involucradas, donde que X es un vector que comprende las variables dependientes conocida a veces como variable de estado. Esas ecuaciones pueden ser Ecuaciones Diferenciales Ordinarias o Ecuaciones Diferenciales Parciales (EDPs) variables en el tiempo. Como es sabido para esas ecuaciones existen métodos analíticos y numéricos para resolverlos dado que las condiciones iniciales son conocidas. Sin embargo, para que esos métodos pueden pronósticar una variable dependiente en el futuro, requerirían extimaciones con alto grado de precisión.

Cualquiera que sea el caso, es esencial la predicción de los valores en el futuro. De este modeo, los métodos de pronóstico son esenciales para la solución de problemas que requieren soluciones en el futuro. Por otra parte los métodos de pronóstico en series de tiempo, sobre los cuales se centra este reporte, son los más adecuados para predecir valores futuros para una gran gama de problemas. Además, para desarrolloar métodos de pronóstico de series tiempo se requiere contar con conjuntos de datos para los cuales existan suficientes métodos de gran precisión que hayan sido probados para una gran cantidad de series de tiempo. Como veremos M4 cumple con esas característica.

Es bien conocido, que en el área de Pronóstico, para un campo de aplicación particular, no existe un solo método que pueda garantizar que tiene el mejor desempeño. Por esa razón la estrategia que se aplica comunmente es la utilización no solo del mejor método identificado o publicado en un campo de aplicación para un problema particular; por el contrario se realiza el mejor ensamble para el pronóstico de valores de variables para un futuro de análisis. Esa estrategia es la que es aplicada en este proyecto.

3. INTRODUCCIÓN: EL PROBLEMA

Predicción con series de tiempo es esencial en muchas áreas de investigación importantes para el quehacer humano, tales como energía, industria, series financieras y cambio climático. Una serie de tiempo es un conjunto de observaciones las cuales se identifican con una variable, del tipo real o categórica. Las variables de tipo real comprenden, como su nombre lo indica, números del conjunto de los reales y son las más comúnmente usadas y, son utilizadas en este trabajo; por su parte las variables llamadas categóricas, se representan con cadenas de letras o de símbolos tales como 'SI', 'NO', 'ALTO', 'BARATO', y 'CARO'. Las series de tiempo se identifican comúnmente con letras o cadenas de letras como X_t , Y_t , T_t , las cuales para identificar observaciones en diversos instantes de tiempo se utiliza frecuentemente la letra t . Todo periodo de tiempo ocurre en un rango $[T_{min}, T_{max}]$. Para una serie de tiempo de interés X_t , la o las variables independientes con por lo regular, también series de tiempo $Y_1, Y_2, \dots, Y_k$, para los cuales se hace explícito el periodo de sus observaciones mediante el subíndice t , es decir; $Y_{1t}, Y_{2t}, \dots, Y_{kt}$. En muchos problemas de pronóstico, hay más de una variable de interés, es decir más de una variable dependiente X y también suele ocurrir que haya más de una variable independiente Y ; en este último caso tanto X como Y serán vectores de una o más variables simples cuyos valores pueden pertenecer a algún conjunto del área de interés, por ejemplo al conjunto $\Omega \in \mathbb{R}^n$, es decir de los números reales, al conjunto $\mathbb{N}$ de los números naturales, o al sistema lógico $L: \{V, F\}$ donde V y F representan valores de verdad de predicados de la lógica proposicional.

En [Brockwell & Davis] se presentan ejemplos de series de tiempo donde los elementos pueden ser de diversos tipos de conjuntos; en nuestro caso los elementos de las series de tiempo pertenecen a los números reales. Como se ha planteado antes, es conveniente no utilizar un método de pronóstico individual sino un ensamble de los pronósticos individuales. Para tal efecto, es importante señalar que para todos los métodos individuales se debe conocer su desempeño mediante métricas previamente establecidas.

Para el desarrollo de métodos de ensamble los métodos individuales son por lo regular probados en diferentes conjuntos de preferencia en aquellos originados con gran cantidad de series de tiempo para diversos problemas de pronóstico en ingeniería para los cuales se han probado los métodos de pronóstico más eficientes; de esa forma la selección de métodos individuales se realiza a partir de los mejores métodos del conjunto. M4 cumple con esas características, donde los mejores métodos de pronóstico han sido probados y cuya precisión se ha medido con métricas de error donde dos de las más utilizadas son las siguientes [Makridakis, 2020]:

- MAPE (Mean absolute percentage error),
- SMAPE (symmetric Mean Absolute Percentage Error) y
- MASE (Mean Absolute Scaled Error).

La primera de estas métricas es preferida para proporcionar información al usuario pues da una idea de su interpretación aun cuando no se tengan conocimientos profundos en el área de pronóstico. La otra métrica, es decir SMAPE, se han establecido con el objeto de definir lo que se conoce en el área de optimización como función objetivo; dicha función, permite establecer las condiciones para

diseñar un buen método de pronóstico. La primera métrica no es conveniente como función objetivo, debido a que tiene condiciones de indeterminación para ciertos valores de la serie de tiempo. No obstante, en general en lugar de eliminar los valores donde la métrica se indetermina es preferible utilizar una métrica sin esa peculiaridad y que ésta sea utilizada en el proceso de optimización. Lo anterior no significa que MAPE sea una métrica no útil debido a que no conviene que sea utilizada durante el proceso de optimización; no obstante, MAPE proporciona una gran ventaja como medida de comparación pues proporciona información fácilmente interpretable para el usuario. En otras palabras, muchos desarrolladores de métodos de pronóstico prefieren como función objetivo métricas como SMAPE u otra similar, pero para información al usuario conviene utilizar MAPE como información para el usuario final.

M4 consta de 100,000 series de tiempo organizadas en períodos que representan los periodos de registro de los datos de cada serie, pudiendo ser diario, semanal, mensual y anual. El método combinado consta de un conjunto de métodos individuales y como hemos dicho antes, se le denomina método de ensamble. El propósito de esta propuesta es desarrollar un método de ensamble para las series de M4. En M4 han participado diversos métodos para los cuales ya se han identificado métodos individuales con muy buen desempeño [Makidrakakis, 2020], [Smyl, 2020]. Un problema de la mayoría de métodos de pronóstico individuales, es que suponen que la series de tiempo son homocedásticas, es decir que la media y la varianza no cambian en diversas porciones de la serie de tiempo. No obstante, comúnmente, las series de tiempo que se presentan en la práctica, así como en M4, es que son heterocedasticas, es decir con media y varianza diferentes a lo largo de la serie de tiempo.

3.2 OBJETIVOS

3.2.1 OBJETIVO GENERAL

Desarrollar un método de ensamble para pronóstico de series de tiempo del conjunto M4 que tenga un desempeño estadísticamente equivalente a los mejores métodos de esta área

3.2.2 OBJETIVOS ESPECÍFICOS

OP1: Objetivo particular 1: Identificar métodos individuales que serán candidatos para conformar el método de ensamble a desarrollar y sus principios de funcionamiento.

OP2. Objetivo particular 2: Determinar los métodos de pronóstico individuales que conformarán el método de ensamble.

OP3. Objetivo particular 3: Desarrollar el modelo de optimización que considere la maximización de la precisión y las características dinámicas involucradas de las series de tiempo en particular.

OP4. Objetivo particular 4. Conformar el algoritmo de ensamble estableciendo diagramas de construcción e interacción de acuerdo con la práctica comúnmente aceptada para el diseño de dichos diagramas.

OP5. Objetivo particular 5 OP5. Difundir resultados en foros ciencia y tecnología acordes al tipo de investigación realizada

4. MARCO TEÓRICO

El problema de predicción es esencial en una gran cantidad de áreas asociadas al ser humano que involucran actividades para su bienestar o en fenómenos que pueden afectar su hábitat. En dichas áreas existen problemas de pronóstico relevantes asociados al quehacer humano tales como los que se ocurren en energía, industria, series financieras y cambio climático. De hecho, La predicción es una actividad tan antigua que incluso las competencias para evaluar los mejores métodos se remontan a varios miles de años citándose esa actividad desde los antiguos chinos y griegos. Las competencias formales de métodos de pronóstico han servido como un escaparate de aquéllos que tienen un mejor desempeño para conjuntos de series de tiempo. Las competencias entre métodos empiezan a organizarse con participación internacional y gran difusión hasta el siglo XX [Hyndman, 2019] [Makidrakis, et al, 2019].

En 1979 se busca medir la precisión de los métodos más populares para la época, que en ese entonces resultaban ser los siguientes: a) Estadísticos o clásicos y b) Métodos cualitativos basados en juicios de expertos [Makridakis, 1979]. Los métodos clásicos o estadísticos se pueden agrupar a su vez en varios tipos:

- a1) Tipo Suavizamiento exponencial (SE)
- a2) Tipo Box and Jenkins (B&J)
- a3) De variables de estado y descomposición.

Los métodos tipo SE ponderan el pronóstico más reciente y el o los datos más recientes. Los métodos del grupo a2 o B&J utilizan comúnmente dos partes: i) Una parte autorregresiva o parte AR y ii) Un promedio móvil conocido como MA [Young, 1977].

A los métodos que conjuntan esas dos partes se conocen como métodos tipo ARMA, y existe una gran cantidad de métodos de ese tipo, entre los cuales los llamados ARIMA (Autoregressive Integrated Moving Average) tienen en general gran calidad de precisión [Brockwell & Davis, 1991]. Sin embargo, para utilizar un método ARIMA hay que determinar los parámetros adecuados para cada serie; para lo cual se han propuesto una gran cantidad de técnicas siendo las más comunes: redes neuronales y de variables de estado [Hussan, et al, 2018]; como consecuencia, una cantidad extremadamente grande de métodos ARIMA que es literalmente imposible analizarlos todos.

Los métodos de tipo a3 se basan en la teoría de variables de estados del área de control moderno [Ogata, 2011] siendo muy exitosos para ciertas aplicaciones, particularmente en pronóstico en tiempo real y a muy corto plazo (muy pocos periodos hacia adelante, del orden de 1 a 7); sin embargo, son poco utilizados debido a dificultades de su implementación para horizontes de pronóstico de un mayor largo plazo. Los métodos de Suavizamiento exponencial, así como los ARMA y los ARIMA son de aplicación más general que los del tipo a3 y su implementación es mucho menos complicada que los del tercer tipo. Hay que hacer notar, que la cantidad de métodos existentes en los grupos a1 y a2 es demasiado grande, de forma que la principal dificultad en estos tipos de métodos es justamente la selección adecuada para las series de tiempo a utilizar. Los métodos de tipo cualitativo son aún útiles hoy en día para ciertas aplicaciones, pero con muy limitada utilidad para el caso de pronóstico con series de tiempo. Existen diversas metodologías, entre las más conocidas se encuentran los métodos Delphi y muchos otros. Actualmente, para problemas de ese tipo, no se utiliza solo un método de

pronóstico sino una combinación de ellos. Es decir, el método a desarrollar se integra con un conjunto de métodos individuales de forma que se aprovechen las ventajas de unos y de otros. Para tal efecto, es importante señalar que para todos los métodos individuales se debe conocer su desempeño mediante métricas previamente establecidas; para M4 es común utilizar las famosas métricas de error MAPE y SMAPE presentadas en la introducción [Makdrakis, 2020]. Cuando se desarrollan métodos de pronóstico estos se prueban en conjuntos de series de tiempo que cuente con gran diversidad de casos, se hayan probado una gran cantidad de métodos; ese es el caso de M4 que cuenta con esas características teniendo, como se mencionó en la introducción, un número grande de series de tiempo donde las observaciones ocurren en periodos diario, semanal y mensual. Adicionalmente, diversos métodos de ensamble han participado en M4 y se tienen identificado los métodos individuales con el mejor desempeño [Makdrakis, 2020], [Smyl, 2020].

5. ACTIVIDADES

En esta sección se presentan en la sección 4.1 los métodos con buen desempeño de M4, a partir los cuales se identifican los métodos individuales; para ello se revisó la bibliografía para identificar cuáles de esos métodos cumplen con esa característica. En la sección 4.2 se seleccionan los métodos individuales que forman parte del ensamble de este proyecto; dicha selección tiene como objetivo seleccionar de entre los mejores métodos de M4 aquéllos que consumen menos recursos computacionales afín de poder probar el método desarrollado en este proyecto; la sección 4.3 plantea las estrategias para el desarrollo de métodos de pronóstico en base a métodos individuales desarrollados y publicados en la literatura especializada del área.

5.1 IDENTIFICACIÓN DE MÉTODOS INDIVIDUALES

En muchos proyectos se requiere conjuntos de series temporales univariadas, de tipo mensual, diario o semanal trabajo. Por otra parte, para obtener métodos de ensamble se necesitan seleccionar métodos de pronóstico individuales con buen desempeño y probados para una gran variedad de series de tiempo. En la Tabla A, 1 se muestran los métodos de pronóstico con mejor desempeño en M4, según una experimentación exhaustiva [Rodríguez, 2021]; estos son métodos individuales para conformar un ensamble, no obstante, que varios de ellos se conforman con varios métodos.

Tabla A1. Métodos de M4

No.	Método	Acró nimo	Individual(I) Combinación (C)	Descripción
1	Naïve	NA	I	Método Naive Clásico
2	ARIMA	AR	I	Método ARIMA usado para comparación
3	AutoArima	AA	I	AA determina el mejor método ARIMA
4	NNetar	NN	C	Método de pronóstico con redes neuronales
5	ETS	ET	C	Usa suavizamiento exponencial y Holt
6	TBATS	TB	C	Suavizamiento exponencial-Box-Cox-ARIMA
7	Holt	HO	I	Método Holt clásico Suavizado exponencial
8	RW-Drift	RW	C	Random Walk con drift activado
9	STLM-ar	ST	C	STLM con método AR.

10	Theta method	TH	C	Suavizamiento exponencial y Regresión lineal
11	Smyl	SM	C	Combina diversos modelos de redes neuronales
12	Bagging	BA	C	Método que busca reducir la variabilidad
13	FFORMA	FF	C	Uso de promedio basado en características
14	Jaganathan	JA	C	Método de Jaganathan

5.1.1 NAIVE

Este método considera que el pronóstico del siguiente periodo es igual al pronóstico del último valor de la serie de tiempo. Como podrá observarse en la ecuación (2) existirá por lo regular un error; el método Naive busca solucionar ese problema realizando un modelo en el periodo de validación de los residuales y luego pronosticando estos con algún método estadístico o de aprendizaje o bien un ensamble de métodos, lo que resulte mejor [Makdrakis, 2020].

5.1.2 ARIMA, AUTOARIMA Y NNETAR

ARIMA es un modelo ARIMA muy utilizado para una gran gama de aplicaciones y que con frecuencia se utiliza para fines prácticos cuando se desea implementaciones rápidas, además se ha empleado para propósitos de comparación de diversos métodos en competencias como las de M4 [Makdrakis, 2020]. Autorima es una forma de ejecutar ARIMA; para ello se buscan los parámetros de ARIMA, mediante diversas implementaciones que presuntamente encuentran los mejores parámetros de dicho modelo, pero comúnmente se basan en el criterio de información de Akaike o BIC. Sin embargo, debido que las distribuciones son por lo general heterocedástica (diferentes en cada porción) no hay garantía de un alto desempeño [Box & Jenkins, et. al, 2016]. Por su parte, NNeTAR (Neural Network Auto-regression) es un modelo de redes neuronales que usa el enfoque auto-regresivo de ARIMA. Este método se identifica comúnmente como NNTAR (p,k), donde p es el número de lags y k especifica cuantos nodos hay en la capa oculta. Este método consume demasiado tiempo de cómputo.

5.1.3 ETS

ETS (Exponential smoothing with Trend and Stationality) es una extensión del método suavizamiento exponencial (SE) propuesto por Brown a mitad del siglo XX y que básicamente consiste en obtener el pronóstico de un periodo hacia adelante mediante la ponderación de los datos previos y un pronóstico anterior; puesto que el último dato y el último pronóstico son Y_t y $\widehat{Y}_t$, entonces el pronóstico de un periodo hacia el futuro es dado por la ecuación (1) [Ponce-Flores, 2021]:

$$\widehat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \widehat{Y}_t \quad (1)$$

Se podría pensar que pronosticar con la ecuación (1) obtendría resultados muy modestos; sin embargo, hay una gran cantidad de métodos de SE, pues hay muchas maneras de realizar la ponderación y de obtener $\widehat{Y}_t$; una ellas es justamente ETS [Hyndman, Koehler, Snyder, & Grose, 2002]. Puesto que el valor pronosticado se puede asociar a cualquier método, ETS combina diversos métodos de Holt&Winters y utiliza una formulación en variables de estado al combinar métodos de pronóstico.

5.1.4 SMYL

Smyl es un framework que consta de un suavizamiento exponencial y un conjunto de redes neuronales previamente entrenadas. El pronóstico se obtiene por un ensamble de todos los pronósticos que es afinado por un modelo de suavizamiento exponencial [Smyl, 2020]; los resultados son de una calidad excelente, y se calificó como el método más preciso de M4. Sin embargo, el costo computacional es extremadamente elevado debido a cuestiones asociadas a los requisitos de entrenamiento. Los investigadores del área buscan la manera de mejorar ese problema.

5.1.5 TBATS

TBATS (Trigonometric seasonality, Box-Cox transformation, ARMA, Trend, Seasonal components) fue propuesto por en el grupo de Hyndman [Alysha, Hyndman, Snyder, 2011], buscando tratar con diversos tipos de periodos estacionales en una misma serie. TBATS puede tratar con series sigan patrones estacionales anuales, mensuales, semanales; en consecuencia, los métodos de manejo de residuos deben tratarse tomando en cuenta esa situación.

Nótese que modelos con diferentes estacionalidades tal (por ejemplo, doble o triple estacionalidad) han sido comunes en el área de pronóstico, entre los que se encuentran los de Holt&Winters, sin embargo, en el caso de TBATS, puede haber un número mayor de estacionalidades, mientras que los residuos son tratados en TBATS por ARIMA. Además, TBATS puede tratar series de tiempo heterocedasticas ya que utiliza la transformación de Box&Cox. [Box & Jenkins, et. al, 2016], [Alysham, Hyndman, Snyder, 2012].

5.1.6 STL-ar y RW-Drift

STL (Stacionality Trend Loess) es un método que divide la serie en:

- Estacionalidad,
- Tendencia y
- Residuales

STL modela cada componente con el modelo de regresión de loess el cual busca suavizar la curva para luego realizar una regresión [Cleveland, 1990]. En STL-ar, los residuales son pronosticados con ARIMA. El método es relativamente simple con una precisión aceptable, sin embargo, se ha notado que este método es menos preciso que Autoarima, ETS y TBATS [Xiaoqian, et. al., 2021].

Por otra parte, RW-Drift (Random Walk with Drift es un método de pronóstico que en primer lugar sigue una caminata aleatoria, la cual equivale a un proceso donde es imposible conocer el valor del estado de las variables de un instante dado al siguiente. RW_drift es un método que combina el método Naive, en el cual pronóstico del futuro es completamente determinada por el último valor del presente; la diferencia es que RW_Drift combina el método Naive con una caminata aleatoria como se muestra enseguida. Supongamos que la variable de estado en el momento presente t es X_t y que conocemos su valor; entonces un método de pronóstico Naive a partir del valor X_t del presente indicaría que el valor el instante posterior $t+1$ se define por la ecuación:

$$X_{t+1} = X_t + \epsilon_t \quad (2)$$

Donde ϵ_t es una variable aleatoria que establece el error del pronóstico. El pronóstico de X_{t+1} que utiliza un proceso que sigue una caminata aleatoria con Drift a partir del valor presente X_t se define a partir de la ecuación siguiente:

$$X_{t+1} = b_0 + b_1 X_t + \epsilon_t \quad (3)$$

Las caminatas aleatorias no tienen Drift si la media es cero o diferente de cero respectivamente. Desafortunadamente RW-Drift no tiene en general muy buena precisión.

5.1.7 THETA

Theta es un método que combina Suavizamiento Exponencial, Regresión lineal y descomposición matricial. Theta busca obtener componentes de la serie de tiempo sin curvaturas pronunciadas. Para ello, se parte de que para los siguientes dos casos:

- Cualquier función en el caso continuo tendrá primeras derivadas cuyas inclinaciones serán más suaves que el gradiente determinado por las primeras diferencias.
- Cualquier función, en el caso discreto, tendrá inclinaciones más suaves conforme las segundas derivadas sean cercanas a cero.

Debido a lo anteriormente señalado, el método Theta, a partir de los elementos de la serie de tiempo X_t : $\{X_1, X_2, X_3, X_4, \dots, X_n\}$, se calculan las primeras y segundas diferencias de X_t :

$$X'_t = \theta_1 \cdot X_t \dots (4)$$

$$X''_t = \theta_2 \cdot X'_t \quad (5)$$

Nótese que θ_1 representa los coeficientes del gradiente de la X_t , mientras θ_2 representa los gradientes de X'_t . Theta considera los puntos donde los gradientes en las ecuaciones anteriores son más pequeños. La ecuación (4) es lineal, ya que los coeficientes existentes en θ_1 son simples números reales, en consecuencia se puede obtener esta ecuación por una regresión lineal con los valores de X'_t y X_t . Un razonamiento similar se aplica a la ecuación (2). Acto seguido se aplica un método de ensamble para obtener el pronóstico [Makdrakis, 2020].

5.1.8 HOLT

El método de Holt considera que la ecuación (1) de suavizamiento exponencial proporcionada anteriormente puede ser corregida. Para ello tomamos nuevamente esa ecuación, que repetimos aquí por simplicidad de la explicación:

$$\text{Dado un modelo exponencial simple: } \widehat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \widehat{Y}_t \quad (6)$$

Si en el lado derecho de (6) a Y_t lo concebimos como un nivel del pronóstico entonces por consistencia lógica, la variable del lado izquierdo debe ser un nivel haciéndolo igual a una variable l_t y si a esta variable le aplicamos un nuevo suavizamiento exponencial tendremos $l_t = \alpha Y_t + (1 - \alpha) \widehat{Y}_t$. En consecuencia, es posible formular el suavizamiento exponencial como sigue [Ponce 2021]:

$$\widehat{Y}_{t+1} = l_t \quad (7a)$$

$$\text{Donde; } l_t = \alpha Y_t + (1 - \alpha) \widehat{Y}_t \quad (7b)$$

El modelo (6) solo es capaz de predecir con buena precisión un periodo hacia adelante. No obstante, el método de Holt extiende el modelo de suavizamiento exponencial simple de las ecuaciones (7) como sigue:

$$\widehat{Y}_{t+h} = l_t \text{ con } l_t = \alpha Y_t + (1 - \alpha) l_{t-1} \dots (8)$$

La ecuación (8) puede pronosticar varios periodos hacia adelante. En el método de Holt se plantea replantear las ecuaciones (7) con una regresión lineal sobre $\widehat{Y}_{t+1}$ que considere el nivel l_t y e intersección b:

$$\widehat{Y}_{t+h} = l_t + b_t h \quad (9)$$

$$l_{t+h} = \alpha Y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (10)$$

$$b_t = \beta(l_t - b_{t-1}) \quad (11)$$

Una forma común de expresar el método de Hot& Winters es utilizar las variables S para los valores suavizados y F para el valor del pronóstico como sigue [Hansun. 2016], [Brown& Meyer, 1961]:

$$F_{t+h} = a_t + b_th \quad (12)$$

$$a_t = 2S_t - S'_t \quad (13)$$

$$b_t = \frac{a}{1-a} (S_t - S'_t) \quad (14)$$

$$S_t = \alpha Y_t + (1 - \alpha)Y'_{t-1} \quad (15)$$

$$S'_t = \alpha S'_t + (1 - \alpha)S'_{t-1} \quad (16)$$

5.1.9 FFORMA Y JAGANATHAN

FFORMA es un ensamble de métodos de pronóstico, en el cual los pesos son obtenidos relacionando las características de las series de tiempo con la eficacia de los métodos; para ello, el cada método individual requiere entrenar un gran número de series de tiempo para lograr determinar las ponderaciones adecuadas, lo cual consume demasiado tiempo computacional; sin embargo, fue uno de los mejores métodos de la competencia M4. Jaganathan por su parte, es un método de ensamble que determina los pesos considerando dos estrategias: función de las ponderaciones que ha obtenido en las series de tiempo a participar o bien usando una distribución uniforme [Jaganathan, 2020]. Aun cuando obtuvo una buena posición en M4, el tiempo de aprendizaje en la determinación de los pesos es excesivamente alto [Ponce-Flores, 2021].

5.1.10 BAGGING

Bagging es un método que consiste en utilizar muchos predictores a partir de los cuales se obtiene la predicción final [Breiman, 1996]. La idea principal es realizar una combinación lineal durante el periodo de validación buscando reducir la varianza. En el caso de M4, cada serie es descompuesta en sus componentes Tendencia, Estacionalidad y Complemento, y pronosticada por separada por métodos individuales a cada componente obteniendo el pronóstico del ensamble por una combinación lineal de los pronosticadores individuales, finalmente el error es corregido por suavizamiento exponencial [Petropolous, 2018]

5.2 DETERMINACIÓN DE MÉTODOS QUE CONFORMEN EL ENSAMBLE

Después de analizar los métodos presentados en la sección anterior y luego de analizar diversas pruebas realizadas sobre M4 probar con diversas métricas podemos determinar que los métodos individuales de buena calidad y con tiempos de ejecución no excesivos que podrán conforma el ensamble son los siguientes [Ponce-Flores, 2021]:

Tabla 2. Métodos individuales para conformar el ensamble

Nombre del método	Nemónico	Tipo de método
AutoArima	AA	Estadístico
Naïve Method	NA	Clásico
ARIMA	AR	Clásico
Holt	HO	Clásico

5.3 DESARROLLO DEL MODELO DE OPTIMIZACIÓN

Una forma eficaz de realizar un ensamble es combinar los pronósticos de métodos individuales utilizando ponderaciones (o pesos) dependiendo de las métricas de los errores de forma que el problema se puede formular como sigue [Bates&Granger,1969]; esa idea se ha extendido hasta la fecha y se ha utilizado recientemente en M4 en los métodos FFORMA Jaganathan; este último autor propuso también utilizar pesos de esa forma de manera considerando una distribución uniforme . Si aplicamos ese principio, entonces la función objetivo que determinaría el ensamble debe ser:

$$\text{Min } \Phi = \sum_{i=1}^n \omega_i \chi_i \text{ s. a } \sum_{i=1}^n \omega_i = 1 \quad (17)$$

No obstante, es fácil comprobar que el ensamble obtenido con el modelo (17) no es mejor que mejor de los métodos individuales [Frausto, 2022A, 2022B]. Otra alternativa que se propone es la siguiente:

- A. Evaluar los métodos participantes con las métricas de desempeño de Akaike y Criterio bayesiano de la información [Frausto 2022 A].
- B. Caracterizar los métodos participantes y sintonizarlos en función de las series de tiempo que sean elegidas.
- C. Establecer un modelo de portafolio de métodos de pronóstico para el subconjunto específico de las series de tiempo seleccionadas.
- D. Refinar el modelo de optimización planteado.
- E. Probar el modelo planteado

6. LOGROS OBTENIDOS DURANTE EL PERIODO SABÁTICO

6.1 ARTÍCULOS DE DIVULGACIÓN

Erick Estrada, Guadalupe Castilla, Juan Frausto, Multivariable Weather Forecasting by Learning and Hurrican occurrences, in Mexican Caribbean, SIMDIA 2023, página 87, 9-10 Nov. 2023, Spain ISBN: 978-84-09-58680-6, <https://www.iberamia.org/iberamia/simdia/wp-content/uploads/sites/16/2024/02/Actas-SIMDIA2023.pdf>

6.2 PUBLICACIONES EN JOURNALS

Pérez-Ortega J.,Moreno-Calderón C, Roblero-Aguilar S., Almanza-Ortega N, Frausto-Solís J, Pazos-Rangel R., Rodríguez J.A, New Criterion for Improving Convergence of Fuzzy C-Means Clustering, Axioms, Jan 2024. <https://www.mdpi.com/2075-1680/13/1/35>

Estrada-Patiño E; Castilla-Valdez G; Frausto-Solís, J, González-Barbosa, J; Sánchez-Hernández J.; A novel approach for temperature forecasting in climate change using ensemble decomposition of time series, International Journal of Computational Intelligence Systems, Springer, 2024, doi: 10.1007/s44196-024-00667-6

6.3 CAPÍTULOS DE LIBRO

Frausto J., Galicia-González J, Gonzalez Barbosa J, Castilla-Valdez G, Sánchez J., SSA-Deep Learning Forecasting Methodology with KF and SMA Filters, Mathematical and Computational Applications, Feb2024, ISSN 2297-8747. <https://www.mdpi.com/2297-8747/29/2/19>

Juan Frausto Solís, Javier González Barbosa, Mirna Ponce Flores, Oscar Bozada Zuñiga, Juan Paulo Sánchez Hernández, FMarkNet Forecasting model based on Neural networks and the Markowitz Model. Part of the book series: Studies in Computational intelligence, Springer, Innovative Applications of Neural Networks to Data Analytics and Signal Processing, 2024, pp 44-49, SCI, Volume 1171, https://link.springer.com/chapter/10.1007/978-3-031-69769-2_3

6.4 TESIS DE MAESTRÍA

1) Rebeca Daniela Yáñez Díaz

Nombre de la tesis: “Métodos de Predicción de la Temperatura como Variable Relevante en el Cambio Climático en Regiones de México.”

Terminación de la tesis: 14 diciembre de 2023

Acta de Examen de grado: 24 de mayo de 2024.

2) Oscar Román Bozada Zuñiga

Nombre de la tesis: “El uso de técnicas de aprendizaje automática en aplicaciones de Cambio Climático”

Terminación de la tesis: 13 mayo de 2023

Acta de Examen de grado: 19 de septiembre de 2024.

REFERENCIAS

- [Alysha, Hyndman, Snyder, 2011] Alysha M. De Livera, Rob J. Hyndman & Ralph D. Snyder (2011) Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing, *Journal of the American Statistical Association*, 106:496, 1513-1527, DOI: 10.1198/jasa.2011.tm09771
- [Alshammai, 2022] Alshammari, F.S.; Akyildiz, F.T.; Khan, M.A.; Din, A.; Sunthrayuth, P. A. Stochastic Mathematical Model for Understanding the COVID-19 Infection Using Real Data. *Symmetry* 2022, 14, 2521. <https://doi.org/10.3390/sym14122521>
- [Assimakopoulo, 2000] V. Assimakopoulos and K. Nikolopoulos, “The theta model: a decomposition approach to forecasting,” *International Journal of Forecasting*, vol. 16, no. 4, pp. 521–530, 2000.
- [Bates & Granger, 1969] Bates, J.M., and Granger C. W., The Combination of Forecasts. *OR, JSTOR*, Vol. 20, No. 4, Dec., 1969, pp. 451-468.
- [Brockwell & Davis, 1991] Brockwell P., Davis R., *Time Series: Theory and Methods*, Springer, 1991
- [Box & Jenkins, et. al, 2016] Box E.P. George, Jenkins Gwilym., Reinsel Gregory, “Time Series. Forecasting and Control, John Wiley & Sons, 2016
- [Box&Cox, 1964] Box E.P.George, Cox D.R., Analysis of Transformations, *Journal of the Royal Statistical Society, Serie B*, 1964, pp.211-252.
- [Breiman, 1996] Breiman, L, Bagging Predictors. *Machine Learning Jr*, Kluwer, 24, pp 123-140, 1996
- [Brown & Meyer, 1961] Brown R, Meyer R, The Fundamental Theorem of Exponential Smoothing, *Operations Research*, vol 9, No. 5, 1961.
- [Cleveland, 1990] Cleveland R, Cleveland S, MacRae J, Terpenning I., STL: A Seasonal Trend Decomposition Based on Loess. *Journal of Official Statistics*, Vol 6, No. 1, 1990, pp 3-73
- [Duarte Santos, 2022] Duarte-Santos F; Lopes Ferreira P; Tristan-Pedersen J.S. The Climate Change Challenge: A Review of the Barriers and Solutions to Deliver a Paris Solution. *Climate 2022, Climate 2022*, Vol 10, Number 75, pp. 2-32; doi: <https://doi.org/10.3390/cli10050075>
- [Estrada Patiño, 2024], Estrada-Patiño E; Castilla-Valdez G; Frausto-Solís, J, González-Barbosa, J; Sánchez-Hernández J.; A novel approach for temperature forecasting in climate change using ensemble decomposition of time series, Submitted paper to *International Journal of Computational Intelligence Systems*, Springer, Accepted for its publication, 2024.

[Hansun S., 2016} Hansen S., A New Approach of Brown's Double Exponential Smoothing Method in Time Series Analysis, Balkan Journal of Electrical and Computer Engineering · October 2016, pp 75-78.

[Hussan, et al,2018] Hussan Al-Chalabi, Yamur K. Al-Douri and Jan Lundberg, Time Series Forecasting using ARIMA Model, ADVCOMP 2018: The Twelfth International Conference on Advanced Engineering Computing and Applications in Sciences, November, 2018 - Athens, Greece.

[Hyndman, 2002] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose, "A state space framework for automatic forecasting using exponential smoothing methods," International Journal of Forecasting, vol. 18, no. 4, pp. 439–454, 2002.

[Hyndman & Khandakar, 2008] Hyndman, R. J., & Khandakar, Y. "Automatic Time Series Forecasting: The forecast Package for R. Journal of Statistical Software", 27(3), 1–22. <https://doi.org/10.18637/jss.v027.i03ds>, 2008

[Hyndman, 2019] Rob Hyndman, A brief history of forecasting competitions, Monash Business School, 2011, <https://www.monash.edu/business/ebs/research/publications/ebs/wp03-2019.pdf>

[Frausto-Solis,2021] Frausto, J.; Hernández, L.; González, J.; Sánchez, J.P.; Román, E. Convolutional Neural Network Component Transformation (CNN-CT) for Confirmed COVID-19 Cases. Math. Comput. Appl. 2021, vol 26, No. 29. <https://doi.org/10.3390/mca26020029>

[Frausto, 2022A] J. Frausto, L. Rodriguez-Moya, J. González-Barbosa, G. Castilla, M. Ponce, "FCTA: A Forecasting Combined Methodology with a Threshold Accepting Approach", Mathematical Problems in Engineering, Volume 2022, Article ID 6206037, pp 1-19, <https://doi.org/10.1155/2022/6206037>.

[Frausto, 2022B] Frausto J; Purata J.L.; González del Angel M.; González; Castilla G, SAIPO-TAIPPO and Genetic Algorithms for Investment Portfolios. Axioms 2022, 11, 42. pp 1-20. <https://doi.org/10.3390/axioms11020042>

[Guckenheimer J, Holmes P, 2002] Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields, 6th Printing 2002 Edition, Springer

[Jaganathan, 2020] Jaganathan S., Prakash P., A combination-based forecasting method for the M4-competition, International Journal of Forecasting, Vol 36, Issue 1, January–March 2020, Pages 98-104.

[Makidrakis., 1979] Makridakis Spyros, Hibon Michele, Claus Moser Claus, Accuracy of Forecasting: An Empirical Investigation, Journal of the Royal Statistical Society Series A (General), Vol. 142, No. 2 (1979), pp. 97-145, Oxford University Press, 1979.

[Makidrakis, 2020] Makridakis Spyros, Evangelos Spiliotis, Vassilios Assimakopoulos, The M4 Competition: 100,000 time series and 61 forecasting methods, International Journal of Forecasting, Volume 36, Issue 1, January–March 2020, pp 54-74, <https://doi.org/10.1016/j.ijforecast.2019.04.014>.

[Montero, 2019] Montero-Manso P, Athanasopoulos G., Hyndman R., Talagala T, FFORMA: Feature-based Forecast Model Averaging, International Journal of Forecasting, Vol 36, Issue 1, January–March 2020, Pages 86-92

[Ogata, 2011] Ogata K., Modern Control Engineering, Pearson edition, 2011.

[Petropoulos, 2018] Petropoulos, F., Hyndman, R., & Bergmeir, C. Exploring the sources of uncertainty: Why does bagging for time series forecasting work? European Journal of Operational Research, 268 (2), pp 545-554.

[Ponce-Flores, 2021] Ponce-Flores Mirna, Métodos Híbridos para Pronóstico de Series Financieras basadas en Suavizamiento Exponencial, Tesis Doctoral, Tecnológico de Ciudad Madero, 2021.

[Ponce-Flores & Frausto 2021] Ponce-Flores M, Frausto J, Santamaría-Bonfil G, Pérez J, González JJ. Time Series Complexities and Their Relationship to Forecasting Performance. Entropy Journal 2020; vol 22, No 1, <https://doi.org/10.3390/e22010089>

[Smyl, 2020] Smyl, S, A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting, International Journal of Forecasting, vol. 36, no. 1, pp. 75–85, 2020.

[Xiaoqian et al., 2021] Xiaoqian, Khan Y., Petropoulos W., Fen L, et al., 2021] The Uncertainty estimation of features based on forecast combinations, Journal of the Operations Research Society, March 2021, pp. 979-993.

[Yannik Hahn, 2023] Yannik Hahn; Tristan Langer; Richard Meyes; Tobias Meisen. Time Series Dataset Survey for Forecasting with Deep Learning, Journal of Forecasting, March 2023, 5, 315–335,

[Young, 1977] Young Warren, The Box-Jenkins approach to time series analysis and forecasting : principles and applications, Revue française d'automatique, d'informatique et de recherche opérationnelle. Recherche opérationnelle, tome 11, no 2 (1977), p. 129-143.

NOMBRE DEL TRABAJO

Proy-Frausto 29Sep2024V6.docx

AUTOR

Juan Frausto Solís

RECuento DE PALABRAS

6348 Words

RECuento DE CARACTERES

35556 Characters

RECuento DE PÁGINAS

18 Pages

TAMAÑO DEL ARCHIVO

94.1KB

FECHA DE ENTREGA

Sep 29, 2024 7:49 PM CST

FECHA DEL INFORME

Sep 29, 2024 7:50 PM CST**● 6% de similitud general**

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos.

- 5% Base de datos de Internet
- Base de datos de Crossref
- 5% Base de datos de trabajos entregados
- 1% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

● Excluir del Reporte de Similitud

- Material bibliográfico
- Material citado
- Material citado
- Coincidencia baja (menos de 10 palabras)