



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ingeniería de Sistemas e Informática

Escuela Profesional de Ingeniería de Sistemas

**Framework usando un componente de pre-
procesamiento y MARS para la predicción de ventas
en las pymes del Perú**

TESIS

Para optar el Título Profesional de Ingeniero de Sistemas

AUTOR

Felix Ricardo BERTO CASTRO

ASESOR

Cayo Víctor LEÓN FERNÁNDEZ

Lima, Perú

2018

FICHA CATALOGRÁFICA

APELLIDOS Y NOMBRES DEL AUTOR:

BERTO CASTRO, Felix Ricardo

TÍTULO DE LA TESIS DE TITULACIÓN:

“FRAMEWORK USANDO UN COMPONENTE DE PRE-PROCESAMIENTO Y MARS PARA LA PREDICCIÓN DE VENTAS EN LAS PYMES DEL PERÚ”

Programa/ línea de investigación: Inteligencia Artificial

(Lima-Perú 2018)

Tesis de titulación, Facultad de Ingeniería de Sistemas e Informática, pregrado, Universidad Nacional Mayor de San Marcos.

Formato 21x29.7 cm, paginas xv, 97

DEDICATORIA

Le dedico esta tesis de titulación a:
a mis padres Felix y Alejandrina.
A esa persona especial.
A Dios.

AGRADECIMIENTOS

Agradezco a mis padres por enseñarme que la constancia y el trabajo duro son las claves del éxito en la vida, a todos mis maestros que me guiaron con sabiduría y experiencia en esta etapa de mi vida universitaria, a mi Universidad San Marcos por abrirme sus puertas y darme la oportunidad de crecer personal y profesionalmente y a esa persona especial por brindarme todo su amor, apoyo, ánimo y compañía incondicional.

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS
FRAMEWORK USANDO UN COMPONENTE DE PRE-PROCESAMIENTO
Y MARS PARA LA PREDICCIÓN DE VENTAS EN
LAS PYMES DEL PERÚ

Autor: Felix Ricardo Berto Castro
Asesor de Tesis: León Fernández Cayo Víctor
Titulo: Tesis para optar el Título Profesional en
Ingeniero de Sistemas
Fecha: 20 de septiembre de 2018

RESUMEN

La predicción de Ventas se basa en predecir eventos futuros, lo cual nos permite tomar decisiones acerca de cuántos productos vamos a producir, cuanto necesitamos de insumos, cuanto personal vamos a requerir, cuanto vamos a invertir, etc. Por esta razón la predicción de ventas resulta de vital importancia para tomar decisiones y mantener una gestión eficiente del negocio.

Las técnicas o métodos de predicción de ventas han tenido importantes avances, lo cual les permite predecir con mayor precisión las condiciones y tendencias del mercado, pero a pesar de ello la mayoría de las PYMES en el Perú no utilizan métodos de predicción de ventas, esto debido a que desconocen cómo aplicar una metodología adecuada en su negocio.

Por otro lado existen PYMES que sí utilizan métodos de predicción de ventas, los cuales aunque son de bajo costo y tiempo, no brindan una precisión adecuada. En el presente trabajo se propone un Framework utilizando un componente de Pre-Procesamiento y MARS para obtener una predicción de ventas de forma precisa y confiable reduciendo el tiempo y costos.

Para ello lo primero que se realizó fue obtener las ventas mensuales de los productos “Resorte Toy Corolla” y “Resorte Toy Corolla Gasol” de una empresa llamada “Franco” dedicada a la venta de autopartes, luego se procesó esta información con un componente de Pre-Procesamiento en donde se detectó y eliminó valores atípicos, Posteriormente se seleccionó las variables de predicción y la variable objetivo, por último se construyó los modelos de predicción y se calculó las medidas de error con la ayuda del software "Salford Predictive Modeler 8.0".

Luego de realizar las pruebas se llegó a la conclusión que el Framework propuesto se ajusta mejor a las necesidades de una PYME ya que se obtiene predicciones de ventas más precisas. Por otro lado no es necesario que el usuario sea un experto ni que tenga conocimientos muy avanzados con respecto a las técnicas o métodos de Predicción, porque cuando usamos el software “Salford Predictive Modeler 8.0”, esta misma selecciona las variables de Predicción más importantes y crea el modelo de predicción con sus respectivas funciones base además también se encarga del cálculo de las medidas de error.

Palabras clave: MARS, Componente de Pre-Procesamiento, Framework, Predicción de ventas, Medidas de Error.

MAJOR NATIONAL UNIVERSITY OF SAN MARCOS
FACULTY OF SYSTEMS ENGINEERING AND INFORMATICA
PROFESSIONAL SCHOOL OF SYSTEMS ENGINEERING
FRAMEWORK USING A PRE-PROCESSING COMPONENT AND MARS
FOR THE PREDICTION OF SALES
IN PYMES OF PERU

Author: Felix Ricardo Berto Castro
Advisor: León Fernández Cayo Víctor
Title: Thesis, to choose the Professional Title
Systems Engineer
Date: 20 September 2018

ABSTRACT

The prediction of Sales is based on predicting future events, which allows us to make decisions about how many products we will produce, how much we need supplies, how much staff we will require, how much we will invest, etc. For this reason, the prediction of sales is of vital importance to make decisions and maintain an efficient management of the business.

Sales prediction techniques or methods have made important advances, which allows them to more accurately predict market conditions and trends, but despite this, most PYMES in Peru do not use sales prediction methods. because they do not know how to apply an adequate methodology in their business. On the other hand, there are PYMES that do use sales prediction methods, which, although they are of low cost and

time, do not provide adequate accuracy. In the present work a Framework is proposed using a component of Pre-Processing and MARS to obtain a prediction of sales in a precise and reliable way reducing the time and costs.

For this, the first thing that was done was to obtain the monthly sales of the products "Toy Corolla Spring" and "Toy Corolla Gasol Spring" from a company called "Franco" dedicated to the sale of auto parts, then this information was processed with a component of Pre-Processing where atypical values were detected and eliminated, then the prediction variables and the objective variable were selected, finally the prediction models were constructed and the error measures were calculated with the help of the software "Salford Predictive Modeler 8.0" .

After conducting the tests, it was concluded that the proposed Framework is better suited to the needs of an PYME since more accurate sales predictions are obtained. On the other hand it is not necessary for the user to be an expert or have very advanced knowledge regarding the techniques or methods of Prediction, because when we use the software "Salford Predictive Modeler 8.0", it selects the most important Prediction variables and it creates the prediction model with its respective base functions and also calculates the error measures.

Keywords: MARS, Pre-Processing Component, Framework, Sales Prediction, Error Measurements.

INDICE

RESUMEN.....	2
ABSTRACT.....	4
INDICE DE GRAFICOS.....	8
INDICE DE CUADROS.....	9
INTRODUCCION.....	11
1. CAPITULO 1.....	1
1.1. Antecedentes.....	1
1.2. Descripción del Problema.....	3
1.2.1. Problema General.....	3
1.2.2. Problemas Específicos.....	5
1.3. Objetivo.....	5
1.3.1. Objetivo General.....	5
1.3.2. Objetivos Específicos.....	7
1.4. Alcance.....	7
1.5. Justificación.....	8
2. CAPITULO 2 MARCO TEORICO Y ESTADO DEL ARTE.....	9
2.1. MARCO TEÓRICO.....	9
2.1.1. Multivariate Adaptive Regression Splines (Splines de regresión adaptativa multivariante ó MARS).....	9
2.1.2. Medidas de Error.....	14
2.1.3. Componente de Pre-Procesamiento.....	19
2.2. Estado del Arte.....	22
2.2.1. Un Modelo Inteligente Híbrido para la Predicción de Ventas a mediano plazo en las cadenas de suministro de moda de venta al por	

menor utilizando una máquina de aprendizaje extremo y un algoritmo de búsqueda armonía	22
2.2.2. Predicción de ventas para comerciantes informáticos: una comparación de Adaptación Multivariado Splines de Regresión y Redes Neuronales Artificiales.....	27
2.2.3. Rendimiento de los Modelos de Espacio de Estado y ARIMA para la predicción de venta del consumo al por menor	31
2.2.4. Una aplicación mejorada del modelo Lotka-Volterra para pronosticar las ventas de dos competidores minoristas.	36
2.2.5. Predicción de Ventas de Productos utilizando las revisiones en línea y los datos históricos de ventas: Un método que combina el modelo de BASS y el análisis de sentimientos	44
2.3. Cuadro Comparativo	50
2.3.1. Factores de evaluación	50
2.3.2. Trabajos Evaluados	52
2.3.3. Cuadro Comparativo	52
2.3.4. Análisis del cuadro comparativo	56
3. CAPITULO 3: METODOLOGIA	58
3.1. Definición de la Metodología Propuesta.....	58
3.2. Aporte Teórico	60
3.2.1. Contribución Teórica	60
3.3. Aporte Práctico	74
3.3.1. Introducción.....	74
3.3.2. Experimentos Realizados	75
4. CAPITULO 5: CONCLUSIONES Y RECOMENDACIONES.....	95
4.1. Conclusiones	95
4.2. Recomendaciones	97
REFERENCIAS BIBLIOGRAFICAS	3

INDICE DE GRAFICOS

Grafico 1. Árbol de Problemas.....	4
Grafico 2: Árbol de Objetivos.....	6
Grafico 3: Un ejemplo simple de MARS.....	11
Grafico 4: Componente de Pre-Procesamiento.....	20
Grafico 5: Framework del modelo Inteligente Híbrido.....	23
Grafico 6: Diagrama de Flujo del Pronosticador HI.....	25
Grafico 7: Framework de Predicción propuesto del modelo LV.....	39
Grafico 8: Pasos detallados del Framework de Predicción Propuesto.....	41
Grafico 9: Framework Propuesto que combina el modelo de Bass/Norton y el análisis de Sentimientos	47
Grafico 10: Metodología Propuesta.....	59
Grafico 11: Componentes de la Contribución Teórica.....	60
Gráfico 12: Framework propuesto.....	64

Grafico 13: Archivo csv con las ventas mensuales y las variables de Predicción de un determinado producto	70
Gráfico 14: Pantalla de Ingreso del historial de ventas.....	71
Gráfico 15: Pantalla para la selección de la metodología y las variables.....	72
Gráfico 16: Pantalla de Testing.....	73

INDICE DE CUADROS

Cuadro 1: Medidas de Error de Predicción Absoluta.....	15
Cuadro 2: Medidas de Error basadas en el error Porcentual.....	17
Cuadro 3: Evaluación de rendimiento de los modelos de predicción debajo de diferentes proporciones relativas.....	30
Cuadro 4: Predicciones de RMSE.....	33
Cuadro 5: Predicciones de MAE.....	34
Cuadro 6: Predicciones de MAPE.....	35
Cuadro 7: Cuadro Comparativo de los trabajos analizados.....	56
Cuadro 8: Cuadro resumen de Factores Deseables para el Framework de Predicción de Ventas.....	56
Cuadro 9: Media y Desviación Estándar para agosto.....	66
Cuadro 10: Valores Atípicos para el mes de Agosto.....	66
Cuadro 11: Variable Objetivo.....	68
Cuadro 12: Variables de Predicción para el Framework.....	69
Cuadro 13: Meses procesados en los experimentos realizados.....	75
Cuadro 14: Valor de las Funciones base de los modelos de Predicción correspondientes a los meses de enero a mayo del 2014 usando únicamente MARS sin Pre-Procesamiento.....	77

Cuadro 15: Tabla Resumen de los valores Predichos calculados a partir de los modelos de Predicción usando únicamente MARS sin Pre-Procesamiento.....	78
Cuadro 16: Valor de las Funciones base de los modelos de Predicción correspondientes a los meses de enero a mayo del 2014 usando únicamente MARS sin Pre-Procesamiento.....	80
Cuadro 17: Tabla Resumen de los valores Predichos calculados a partir de los modelos de Predicción usando únicamente MARS sin Pre-Procesamiento.....	81
Cuadro 18: Valor de las Funciones base de los modelos de predicción correspondientes a los meses de enero a mayo del 2014 usando el Framework Propuesto.....	83
Cuadro 19: Tabla Resumen de los valores Predichos calculados a partir de los modelos de Predicción usando el Framework Propuesto.....	84
Cuadro 20: Valor de las Funciones base de los modelos de Predicción correspondientes a los meses de enero a mayo del 2014 usando el Framework Propuesto.....	86
Cuadro 21: Tabla Resumen de los valores Predichos calculados a partir de los modelos de Predicción usando el Framework Propuesto.....	87
Cuadro 22: Tabla Comparativa de las Predicciones de ventas del Producto “Resorte Toy Corolla”.....	88
Cuadro 23: Tabla Comparativa de las Predicciones de ventas del Producto “Resorte Toy Corolla Gasol”.....	89

Cuadro 24: Cuadro Comparativo de las medidas de error del producto "Resorte Toy Corolla"	91
Cuadro 25: Cuadro Comparativo de las medidas de error del producto "Resorte Toy Corolla Gasol"	92

INTRODUCCION

Las PYMES en la actualidad buscan tener un buen rendimiento de ventas, por lo cual necesitan mantener un adecuado equilibrio con la demanda del cliente y el inventario que gestionan, ya que tener en exceso cualquiera de estos dos factores, traería consecuencias negativas para la PYME, por ejemplo si tenemos un inventario muy grande se podría satisfacer la demanda del cliente en todo momento, pero se correría el riesgo de quedarse con su capital inmovilizado o perder sus productos si estos son perecibles.

Por otro lado si es que la PYME tiene un nivel de inventario pequeño, esto reduciría su costo de inventario, pero conduce a otros problemas como costos de oportunidad derivados de las oportunidades de ventas perdidas y la reducción de la satisfacción del cliente.

Para conseguir el equilibrio entre la demanda del cliente y el inventario, las PYMES tienen la opción de utilizar técnicas o métodos de predicción de ventas las cuales pueden reflejar con mayor precisión las condiciones y tendencias del mercado, pero a pesar de todos estos beneficios, la gran mayoría de las PYMES en el Perú no las utilizan, esto debido a que desconocen como ejecutarlas de una forma adecuada en su negocio.

Por todo lo expuesto anteriormente el problema que se plantea en esta tesis se resume en una pregunta la cual es “¿De qué manera las PYMES del Perú pueden predecir sus ventas de forma precisa y confiable reduciendo el tiempo y costos para obtener un mayor margen de ganancias?”.

Para dar solución al problema que se plantea, esta tesis desarrollo un Framework utilizando un componente de Pre-procesamiento y MARS para obtener una predicción de ventas de forma precisa y confiable reduciendo tiempos y costos.

El Framework utilizo las ventas mensuales de los productos “Resorte Toy Corolla” y “Resorte Toy Corolla Gasol” de una empresa llamada "Franco" la cual se dedica a la venta de auto-partes. Para dar una solución eficiente al problema planteado en primer lugar se utilizó un componente de Pre-Procesamiento el cual tuvo un impacto significativo en el rendimiento del Framework porque elimino las muestras no fiables o influenciadas por diversos factores externos impredecibles. Luego se utilizó el software "SALFORD PREDICTIVE MODELER 8.0" el cual usa MARS para crear los modelos de predicción y calcular las medidas de error para cada producto seleccionado

1. CAPITULO 1

1.1. Antecedentes

Según (Yalta Maicelo, Juan José, 2013) “A lo largo de la historia de la humanidad el hombre siempre ha buscado predecir su entorno para obtener alguna ventaja sobre el resto, es así que desde la prehistoria realizaban incisiones en los huesos de animales para representar las fases de la luna buscando predecir los cambios de estación. Luego en la edad antigua muchos reyes y altos jefes militares griegos recurrían al Oráculo de Delfos para consultar como sería el resultado de las guerras o cuando sería el advenimiento de plagas y enfermedades. En la Edad Media los reyes católicos contaban con astrólogos para ayudarlos a tomar decisiones. En la Edad Moderna se contaba con técnicas de predicción de eclipses. En la Edad Contemporánea se desarrollan la mayoría de técnicas estadísticas de predicción comúnmente utilizados en la actualidad”.

(Nojek, Britos, Rossi y Garcia, 2003) nos indica que en la actualidad las predicciones son de gran importancia para las empresas ya que se utilizan en diversas áreas como el área de compras, marketing y ventas. Cada una de las técnicas para la predicción se pueden seleccionar tomando en cuenta diversos factores, como por ejemplo, la validez y disponibilidad de los datos históricos, la precisión deseada de la predicción, el costo del procedimiento, los beneficios

del resultado, los periodos futuros que se desee predecir y el tiempo disponible para hacer el estudio entre otros.

(Mentzer J. T, Moon M. A, 2005) realizo una encuesta a 207 ejecutivos sobre sus prácticas en la predicción de ventas. Un hallazgo particular indico una mayor confianza y satisfacción con las técnicas de predicción cuantitativa actuales frente a las de hace diez años atrás. Otro hallazgo indicó que la familiaridad y el uso de varias técnicas de predicción sofisticados han aumentado, esto lo demuestra el estudio ya que el 92% de los encuestados estaban familiarizados con las técnicas de media móvil, 90% con la técnica de suavizado exponencial, 85% con la técnica de proyección de línea recta y 77% con la técnica de regresión, Sin embargo la mayoría no estaban familiarizados con las redes neuronales.

De acuerdo con (Chi-Jie Lu, Tian-Shyug Lee y Chia-Mei Lian, 2012) “En el mundo de los negocios de hoy en día, las empresas buscan un buen rendimiento de ventas por lo cual necesitan mantener un equilibrio entre la demanda del cliente y el inventario, ya que llevar un inventario muy grande permite satisfacer la demanda del cliente en todo momento, pero puede resultar muy desfavorable para la empresa ya que esto puede conducir a tener capital inmovilizado, la reducción de márgenes de ganancia y la pérdida de los productos si estos son perecibles, por otro lado tener niveles bajos de inventario pueden reducir el costo del inventario, pero pueden conducir a otros problemas como costos de oportunidad derivados de las oportunidades de venta perdidas y reducción de la satisfacción del cliente”.

Se puede llegar a la conclusión de que con el paso del tiempo se han ido perfeccionando las técnicas o métodos de predicción de ventas, haciendo de estos cada vez más elaborados y complejos, brindando una mayor precisión en las predicciones de ventas, pero a pesar del gran avance muchas PYMES en el Perú no los implementan, es por ello que en la siguiente sección se tratara el problema que afrontan muchas PYMES en el Perú, que hacen muy limitada el uso de estos métodos o técnicas.

1.2. Descripción del Problema

1.2.1. Problema General

¿De qué manera las PYMES del Perú pueden predecir sus ventas de forma precisa y confiable reduciendo el tiempo y costos para obtener un mayor margen de ganancias?

Para mayor comprensión del problema en el grafico 1 se muestra el árbol de problemas.

Árbol de Problemas

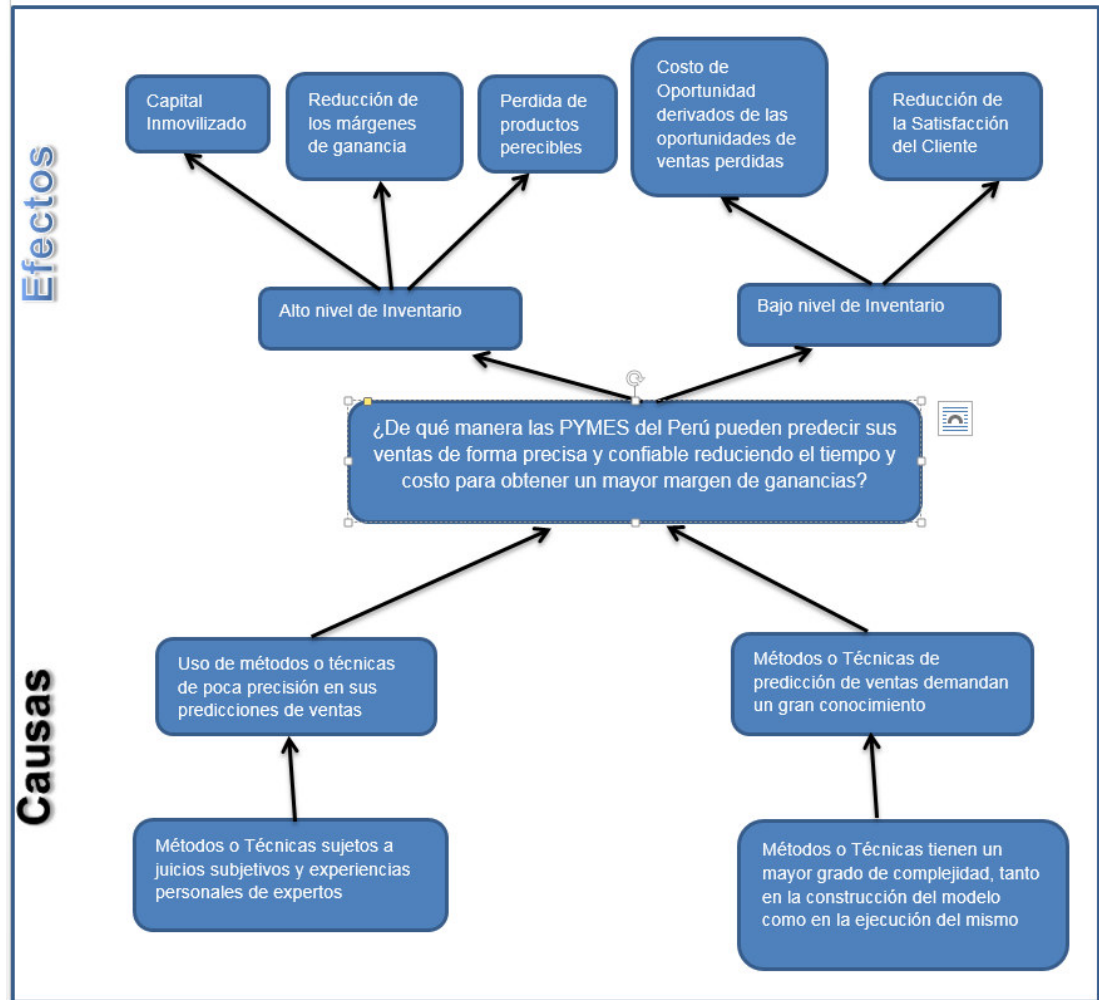


Grafico 1, Árbol de Problemas, (Fuente: elaboración propia)

1.2.2. Problemas Específicos

- ¿Cómo influye la precisión de las predicciones de ventas en la productividad de las PYMES del Perú?
- ¿Cómo influye las medidas de error en la confiabilidad de las predicciones de Ventas de las PYMES del Perú?
- ¿Cómo influye el uso de los métodos de predicción con deficiente precisión en los márgenes de ganancia de las PYMES del Perú?
- ¿Cómo influye el uso de los métodos de predicción de ventas muy complejos en su implementación en las PYMES del Perú?

1.3. Objetivo

1.3.1. Objetivo General

Desarrollar un Framework utilizando un componente de Pre-procesamiento y MARS para obtener una predicción de ventas precisa y confiable reduciendo tiempo y costos. Para que las PYMES puedan obtener mayores márgenes de ganancias. A continuación en el Grafico 2 se muestra el árbol de Objetivos.

ARBOL DE OBJETIVOS

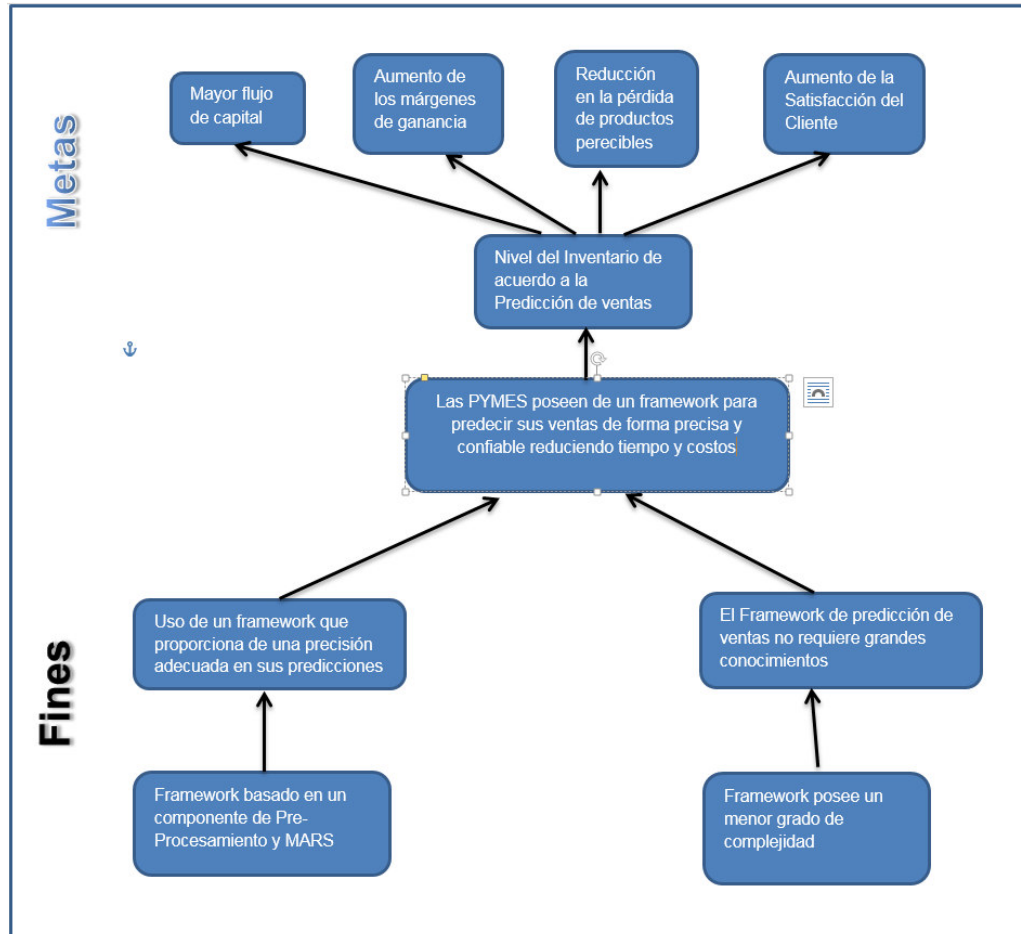


Grafico 2, Árbol de Objetivos, (Fuente: elaboración Propia)

1.3.2. Objetivos Específicos

- Determinar la influencia de la precisión de las predicciones de ventas en la productividad de las PYMES del Perú.
- Determinar la influencia de las medidas de error en la confiabilidad de las predicciones de ventas de las PYMES del Perú.
- Obtener el historial de ventas de una PYME del Perú, para realizar pruebas y demostrar la precisión del Framework que se desarrollara.
- Determinar la mejor técnica o metodología para obtener Predicciones de Ventas Precisas.
- Desarrollar el Framework de forma entendible y poco complejo para facilitar su implementación en las PYMES del Perú.

1.4. Alcance

En esta investigación se diseñó un Framework de Predicción de Ventas que permite a las PYMES del Perú obtener predicciones de Ventas precisas y confiables reduciendo tiempo y costos. Este Framework está conformado por un componente de Pre-Procesamiento y la metodología MARS. Se debe mencionar que no se tomó en cuenta a las grandes empresas, tampoco se tomó en cuenta factores externos como las decisiones estratégicas por parte de los competidores de la PYME en estudio.

1.5. Justificación

Esta investigación es necesaria para las PYMES del Perú, porque la mayoría de ellas no utilizan algún método de Predicción de Ventas y las PYMES que si las usan no obtienen resultados positivos esto debido a que utilizan métodos que aunque son de bajo costo y tiempo, no brindan una precisión adecuada, pues en su mayoría estos métodos están sujetos a juicios subjetivos y experiencias personales de los expertos. Es por ello que el desarrollo de la solución planteada en esta tesis contribuirá de forma positiva en el desempeño y crecimiento de las PYMES del Perú brindándoles de un Framework de Predicción de Ventas confiable y preciso a un menor tiempo y costo.

Si las PYMES utilizan el Framework propuesto en este estudio de forma correcta, podrán obtener predicciones de ventas más precisas, lo cual les permitirá mantener un equilibrio entre la demanda del cliente y su inventario, ya que llevar un inventario muy grande permite satisfacer la demanda del cliente en todo momento, pero puede resultar desfavorable para la empresa ya que esto puede conducir a tener capital inmovilizado, la reducción de márgenes de ganancia y la pérdida de los productos si estos son perecibles. Por otro lado tener niveles bajos de inventario pueden reducir el costo del inventario, pero pueden conducir a otros problemas como costos de oportunidad derivados de las oportunidades de venta perdidas y reducción de la satisfacción del cliente”.

2. CAPITULO 2 MARCO TEORICO Y ESTADO DEL ARTE

2.1. MARCO TEÓRICO

En esta sección se desarrollara el marco teórico el cual tiene como propósito situar el problema de investigación dentro de un conjunto de conocimientos que nos permitirá delimitar teóricamente los conceptos planteados, es decir el marco teórico es el conjunto de conocimientos que respaldan el proceso de investigación y crea las condiciones para llegar a la solución del problema, es por ello que a continuación se abordaran conceptos como la metodología Multivariate Adaptive Regression Splines (MARS), Medidas de Error, Pre-Procesamientos de datos, Algoritmo de Búsqueda de Armonía, Redes Neuronales, Modelos Espaciales de Estado y Modelos ARIMA.

2.1.1. Multivariate Adaptive Regression Splines (Splines de regresión adaptativa multivariante ó MARS)

De acuerdo con (J.H. Friedman, 1991), MARS es una metodología de regresión no lineal y no paramétrico como un procedimiento flexible en el que las relaciones del modelo son casi aditivas con un menor número de variables. El procedimiento de modelado MARS se basa en la estrategia divide y conquista en la que los conjuntos de datos de entrenamiento se dividen en regiones separadas, cada uno de los cuales se le asigna su propia ecuación de regresión.

MARS sobresale no solo en la búsqueda variables optimas e interacciones, sino también en la identificación de las estructuras de datos complejas a menudo oculta en los datos de gran dimensión. Esto hace particularmente adecuado a MARS si hay grandes cantidades de datos de entrada.

MARS se basa esencialmente en la construcción de modelos flexibles por regresiones lineales a trozos; es decir la no linealidad de un modelo se aproxima mediante el uso de las pendientes de regresiones lineales separadas en intervalos distintos de espacio de variable independiente. Por lo tanto, la pendiente de la regresión lineal permite cambiar de un intervalo a otro, como dos puntos ('nudo') son cruzados.

Las variables que se utilizarán y los puntos finales de los intervalos para cada variable se encuentran a través de un rápido pero intensivo procedimiento de búsqueda. Además de la búsqueda de las variables uno por uno, MARS también busca interacciones entre las variables, lo que permite cualquier grado de interacción a considerar siempre que el modelo construido proporcione un mejor ajuste con los datos.

El Grafico 3 representa un ejemplo sencillo de cómo MARS podría intentar encajar los datos en un espacio de dos dimensiones con la regresión lineal a trozos. Tenga en cuenta que 'Y' y 'X' son variables dependientes e independientes respectivamente. Se puede

observar que K_1 y K_2 son puntos nudos que delimitan tres intervalos donde diferentes relaciones lineales son identificadas.

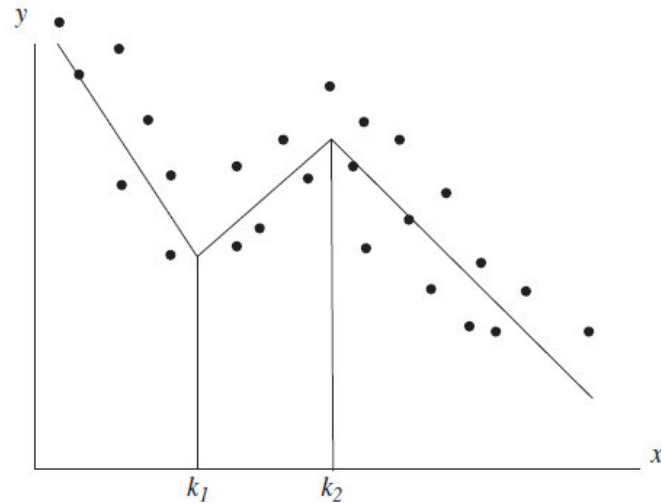


Grafico 3, Un ejemplo simple de MARS, (Chi-Jie Lu, Tian-Shyug Lee y Chia-Mei Lian, 2012)

La función general de MARS puede representarse mediante la siguiente ecuación (Ec.1).

$$f(x) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^K [s_{k,m}(x(k,m) - t_{k,m})] \quad \dots \text{Ec. 1}$$

Donde a_0 es una constante; a_m son los coeficientes del modelo, que son estimados para dar el mejor ajuste a los datos; M es el número de la función base; K es el número de "divisiones" que generan la

m-ésimo función base; $s_{k,m}$ la cual toma valores de 1 o -1 e indica el sentido (derecha /izquierda) de la función escalón asociado; Las $x_{(k, m)}$ es la etiqueta de la variable independiente; y $t_{k, m}$ indica la localización del nudo.

El modelo MARS óptimo es seleccionado en un proceso de dos etapas, en la primera etapa MARS construye un gran número de funciones base para ajustar los datos, donde se permite variables de entrada continuas, categóricas u ordinales. En esta etapa, MARS encuentra el par de funciones base que da la reducción máxima en el error residual de suma de cuadrados.

Una función de bisagra se define por una variable y un nudo, es decir, $\max(0, x - \text{const})$ o $\max(0, \text{const} - x)$ donde const es el nudo. MARS utiliza un algoritmo voraz para considerar si se agrega una función base al modelo MARS buscando todas las combinaciones de todos los valores de todas las variables.

Las funciones base son seleccionadas inicialmente en función de la media de los valores de las variables independientes. Este proceso de agregar funciones base continúa hasta que el cambio en el error residual es demasiado pequeño para continuar o hasta que el número máximo de funciones base es alcanzado.

En la segunda etapa, las funciones base se suprimen en el orden de las que menos contribuciones aportan utilizando el criterio (GCV) de validación cruzada generalizada. Una medida de Importancia variable puede entonces ser evaluada observando la disminución en el GCV calculado cuando se elimina una variable del modelo.

Este proceso continúa hasta que las funciones base restantes cumplen todos los requisitos predeterminados. El GCV se puede expresar como en la siguiente ecuación (Ec.2).

$$GCV(M) = \frac{1}{N} \sum_{i=1}^N \frac{[y_i - f_M(x_i)]^2}{\left[1 - \frac{C(M)}{N}\right]^2} \dots \text{Ec. 2}$$

Donde N son las observaciones y $C(M)$ es la medida de penalización de costos de un modelo que contiene M funciones base (Por lo tanto, el numerador mide la falta de ajuste en el modelo de función base M la cual es $f_M(x_i)$ y el denominador denota la penalidad para la complejidad del modelo $C(M)$). Dicho de otro modo, el propósito de $C(M)$ es penalizar la complejidad del modelo, evitar la sobrecarga y promover la parsimonia del modelo. Para ello, $C(M)$ introduce un costo incurrido por función base para el modelo que es similar al R^2 ajustado en la regresión de mínimos cuadrados.

Después de crear un modelo MARS, se puede estimar la importancia relativa de una variable basada en su contribución al ajuste del modelo en una escala de 0 a 100. Para ello, MARS borra todos los términos que contienen la variable seleccionada, vuelve a montar el modelo y luego calcula la reducción del ajuste. El puntaje más alto y la variable más importante es la que más reduce el ajuste del modelo después de ser eliminada.

Las variables menos importantes reciben puntuaciones más bajas, que corresponden a la proporción de reducción en el ajuste producido por estas variables que de la variable más importante.

MARS es capaz de rastrear estructuras de datos muy complejas, a menudo ocultas en datos de alta dimensión.

2.1.2. Medidas de Error

De acuerdo con (Maxim Vladimirovich Shcherbakov, Adriaan Brebels, Nataliya Lvovna Shcherbakova, Anton Pavlovich Tyukov, Timur Alexandrovich Janovsky y Valeriy Anatol'evich Kamaev, 2013) Las medidas de error de predicción son aplicadas para estimar la calidad de los métodos de predicción y para elegir el mejor mecanismo de predicción. En esta sección se realizó un análisis de las medidas de error que son usadas en la predicción. Las medidas se dividen en grupos de acuerdo al método de cálculo.

2.1.2.1. Absolute Forecasting Error (Error de predicción absoluto):

El primer grupo se basa en el cálculo del error absoluto. Incluye estimaciones basadas en el cálculo del valor e_t , como se muestra en la siguiente ecuación (Ec. 3).

$$e_t = (y_t - f_t^{(m)}) \dots \text{Ec. 3}$$

Donde y_t es el valor medido en el tiempo t , $f_t^{(m)}$ valor predicho en el tiempo t , Obtenido del uso del modelo (m) de predicción. A continuación se mostrara el cuadro 1, con las medidas de error

pertenecientes a este primer grupo, donde n es el horizonte de predicción.

Medida de Error	Calculo
MAE o MAD (Mean Absolute Error)	$MAE = \frac{1}{n} \sum_{i=1}^n e_i $
MSE (Mean Square Error)	$MSE = \frac{1}{n} \sum_{i=1}^n (e_i^2)$
RMSE (Root Mean Square Error)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i^2)}$

Cuadro 1, Medidas de Error de predicción absoluta, (Maxim Vladimirovich Shcherbakov, Adriaan Brebels, Nataliya Lvovna Shcherbakova, Anton Pavlovich Tyukov, Timur Alexandrovich Janovsky y Valeriy Anatol'evich Kamaev, 2013)

Estas medidas de error son los más populares en diversos ámbitos. Sin embargo, las medidas de error absoluto tienen las siguientes deficiencias:

- El principal inconveniente es la dependencia de la escala. Por lo tanto, si la tarea de predicción incluye objetos con diferentes escalas o magnitudes, entonces no se podrían aplicar medidas de error absoluto.
- El siguiente inconveniente es la fuerte influencia de valores atípicos en los datos de la evaluación del rendimiento de predicción.
- RMSE tiene una fiabilidad baja, los resultados podrían ser diferentes dependiendo de fracciones diferentes de datos.

2.1.2.2. Measures Based on Percentage Errors (Medidas basadas en porcentajes de errores):

El Segundo grupo son los errores porcentuales, los cuales se calculan sobre la base del valor P_t , como se muestra en la siguiente ecuación (Ec. 4)

$$p_t = \frac{|e_t|}{y_t} \dots \text{Ec. 4}$$

Además estos errores son los más comunes en el ámbito de las predicciones. A continuación se mostrara el cuadro 2 con las medidas de error pertenecientes a este segundo grupo, donde n es el horizonte de predicción.

Medida de Error	Calculo
MAPE (Measures Based on Percentage Errors)	$MAPE = \frac{1}{n} \sum_{i=1}^n 100 \cdot p_i $
RMSPE (Root Mean Square Percentage Error)	$RMSPE = \sqrt{\frac{mean}{i=1,n} (100 \cdot p_i)^2}$ <p>Donde mean es una operación para el cálculo de la media</p>
RMdSPE (Median percentage error of the quadratic)	$RMdSPE = \sqrt{\frac{median}{i=1,n} (100 \cdot p_i)^2}$ <p>Donde median es una operación para el cálculo de la mediana</p>

Cuadro 2, Medidas de Error basadas en el error porcentual, (Maxim Vladimirovich Shcherbakov, Adriaan Brebels, Nataliya Lvovna Shcherbakova, Anton Pavlovich Tyukov, Timur Alexandrovich Janovsky y Valeriy Anatol'evich Kamaev, 2013)

Para este segundo grupo, se puede observar las siguientes deficiencias:

- División por cero cuando el valor real es igual a cero.
- Los valores atípicos tienen un impacto significativo en el resultado, especialmente si el valor atípico tiene un valor mucho mayor que el valor máximo de los casos "normales".
- Las medidas de error son sesgadas, esto puede conducir a una evaluación incorrecta del rendimiento de los modelos de predicción.

Esta sección se revisó las principales medidas de error para modelos de predicción de series temporales. Todas estas medidas se agruparon en 2 grupos los cuales son el error de pronóstico absoluto y el error de predicción porcentual. Para cada medida de error se presentaron la forma del cálculo y también se definieron los inconvenientes para cada grupo.

2.1.3. Componente de Pre-Procesamiento

(W.K. Wong, Z.X.Guo, 2010) realizaron un estudio para afrontar el problema de las predicciones de ventas de artículos de moda a mediano plazo. Para lo cual diseñaron un modelo híbrido inteligente, que comprende un componente de pre-procesamiento de datos y un pronosticador híbrido inteligente (HI). En esta sección nos centraremos en explicar cómo funciona el componente de Pre-Procesamiento.

El Pre-procesamiento de datos tiene un impacto significativo en el rendimiento de los modelos de aprendizaje controlado debido a que las muestras poco fiables probablemente conducen a predicciones erróneas. Aunque los datos de ventas de artículos de moda suelen ser ruidosos e influenciados por varios factores externos impredecibles, los estudios previos no consideraron el pre-procesamiento de los datos de ventas.

Las series de tiempo son usadas para observar los cambios de tendencia de la venta de artículos de moda, lo cual se puede lograr observando el historial de ventas de los mismos periodos de años pasados. $S_{i,j}$ denota la cantidad de ventas en el mes 'j' del año 'i'. La siguiente secuencia representa la serie de tiempo mensual original del primer mes del año 'i' hasta el doceavo mes del año 'i+k'.

$$S_{i,1}, S_{i,2}, \dots, S_{i,2}, S_{i,12}, S_{i+1,2}, \dots, S_{i+1,12}, \dots, S_{i+k,1}, S_{i+k,2}, \dots, S_{i+k,12}.$$

La serie temporal S comprende 12 mensualidades del mismo período de la serie de tiempo. Sea S_j denota la serie de tiempo del mismo período

en el mes j ($j= 1,2,3,\dots,12$). S_j se puede representar de la siguiente manera:

$$S_{i,j}, S_{i+1,j}, \dots, S_{i+k,j},$$

Tal y como muestra el Grafico 4, el componente de Pre-Procesamiento de datos implica tres procesos, los cuales incluyen la detección y eliminación de valores atípicos, interpolación de datos faltantes y la normalización de los datos, a continuación se detalla cada uno de estos procesos.



Grafico 4, Componente de Pre-Procesamiento, (W.K. Wong, Z.X.Guo, 2010)

2.1.3.1. Detección y Eliminación de Valores Atípicos

Un valor atípico es una observación que se desvía mucho del resto de las observaciones de tal manera que se sospeche que el valor atípico fue generado por un mecanismo diferente al resto de valores. Se considera la observación $S_{i,j}$ en el mismo periodo de la serie de tiempo S_j de una categoría de artículo como un valor atípico si cumple la siguiente condición:

$$abs(S_{i,j} - mean(S_j)) > n . std(S_j) \dots \text{Ec. 5}$$

Donde $mean(.)$ denota la función de la media, $std(.)$ denota la función de desviación estándar y $abs(.)$ denota la función de valor de absoluto. En esta investigación, n se fija en 3. Los valores atípicos deben ser removidos y luego ser manejados como una observación faltante.

2.1.3.2. Interpolación de datos Faltantes

Los datos incompletos constituyen un problema inevitable en la fuente de datos en estudio. Los datos que faltan deben ser interpolados Para mantener la integridad y los cambios de tendencia de las series de tiempo. Las observaciones faltantes se completan mediante el uso de la media de sus dos últimos vecinos en el mismo periodo.

2.1.3.3. Normalización y des-normalización

La normalización de los datos puede acelerar el tiempo de formación de las Redes Neuronales (NNs) para iniciar el proceso de entrenamiento para cada característica dentro de la misma escala.

2.2. Estado del Arte

En este capítulo se describen los conocimientos existentes acerca de investigaciones que proponen una solución al problema planteado, esto servirá como referencia para asumir una postura crítica frente a lo que se ha hecho y lo que falta por hacer, para evitar duplicar esfuerzos o repetir lo que ya se ha dicho, además localizar errores que ya fueron superados. A continuación se mostraran las investigaciones de acuerdo a su orden cronológico desde la más antigua a la más reciente.

2.2.1. Un Modelo Inteligente Híbrido para la Predicción de Ventas a mediano plazo en las cadenas de suministro de moda de venta al por menor utilizando una máquina de aprendizaje extremo y un algoritmo de búsqueda armonía

De acuerdo con (W.K. Wong, Z.X. Guo, 2010) la Predicción de ventas es la base para la planificación de varias fases de las operaciones de una empresa, lo cual es una tarea crucial en la gestión de la cadena de suministro. Sin las predicciones de ventas, las operaciones sólo pueden responder de manera retroactiva, lo que lleva a una mala planificación de la producción, pedidos perdidos, servicio inadecuado al cliente y recursos mal utilizados.

En este artículo se investigó las ventas de accesorios de moda a mediano plazo, ya que estos se caracterizan por el corto ciclo de vida de sus productos, las volátiles demandas de los clientes y las enormes variedades de productos. La mayoría de las ventas de

artículos de moda son de fuerte estacionalidad, la incertidumbre de la demanda de los clientes es frecuente en un entorno de mercado cambiante. Tal característica aumenta la complejidad de la predicción de ventas. Se planteó como solución un modelo Híbrido Inteligente el cual se compone de un componente de pre-procesamiento de datos y un pronosticador del HI. El Grafico 5 muestra el Framework del modelo Híbrido Inteligente.

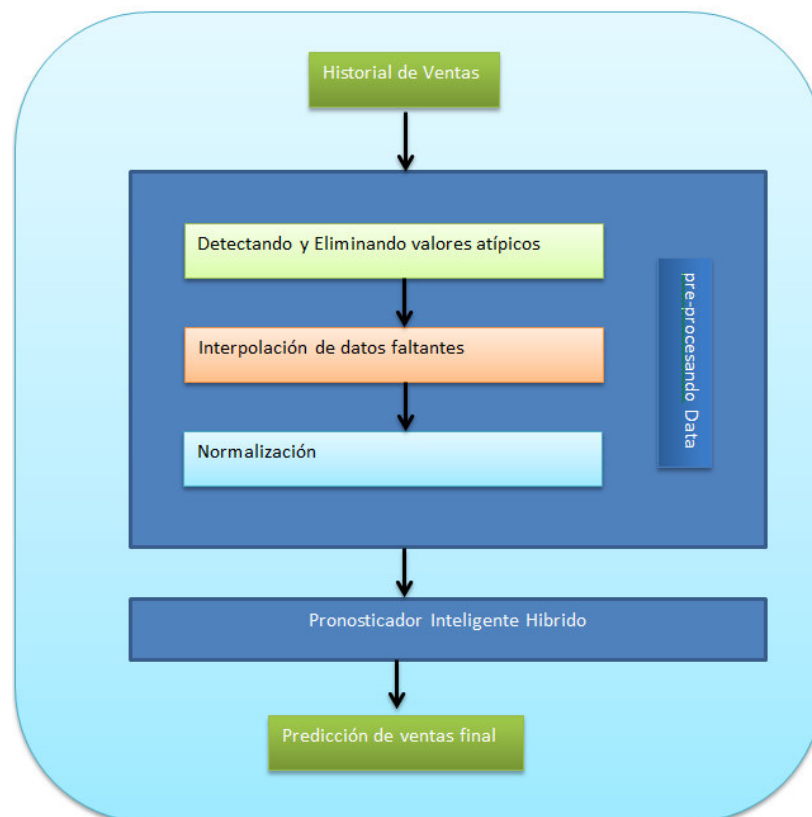


Grafico 5, Framework del modelo Inteligente Híbrido, (W.K. Wongn, Z.X. Guo, 2010)

El componente de pre-procesamiento de datos implica tres procesos, los cuales incluyen la detección y eliminación de valores atípicos, la interpolación de los datos faltantes y la normalización de los datos. La data del historial de ventas pre-procesada se utiliza como muestra de entrenamiento del pronosticador Inteligente Híbrido (HI).

El Pronosticador Inteligente Híbrido (HI) se utiliza después de que las series de tiempo de ventas son pre-procesadas, un pronosticador HI es aplicado para generar predicciones de ventas a mediano plazo. Lo esencial del pronosticador HI es un novedoso algoritmo de aprendizaje basado en predicciones de una NN (Red Neuronal). El algoritmo de aprendizaje HS-ELM es desarrollado para mejorar la capacidad de generalización de la NN (Red Neuronal) mediante la integración del algoritmo HS (Técnica Meta heurística) con ELM (Extreme Learning Machine).

Para disminuir la aleatoriedad de los resultados de la NN (Red Neuronal), el pronosticador HI considera las salidas de predicción de múltiples NNs (Redes Neuronales) con diferente número de neuronas ocultas. En el pronosticador HI, los resultados obtenidos mediante HS-ELM generan múltiples resultados de predicción ejecutando repetidamente la red con un número distinto de neuronas ocultas de 1 a $N_{\max \text{ hl}}$. Luego sigue un proceso de des-normalización de los resultados propuestos, por último se utiliza un proceso de ajuste heurístico para analizar estos resultados y generar la predicción de ventas final. El diagrama de flujo del pronosticador HI se muestra a continuación en el Grafico 6.

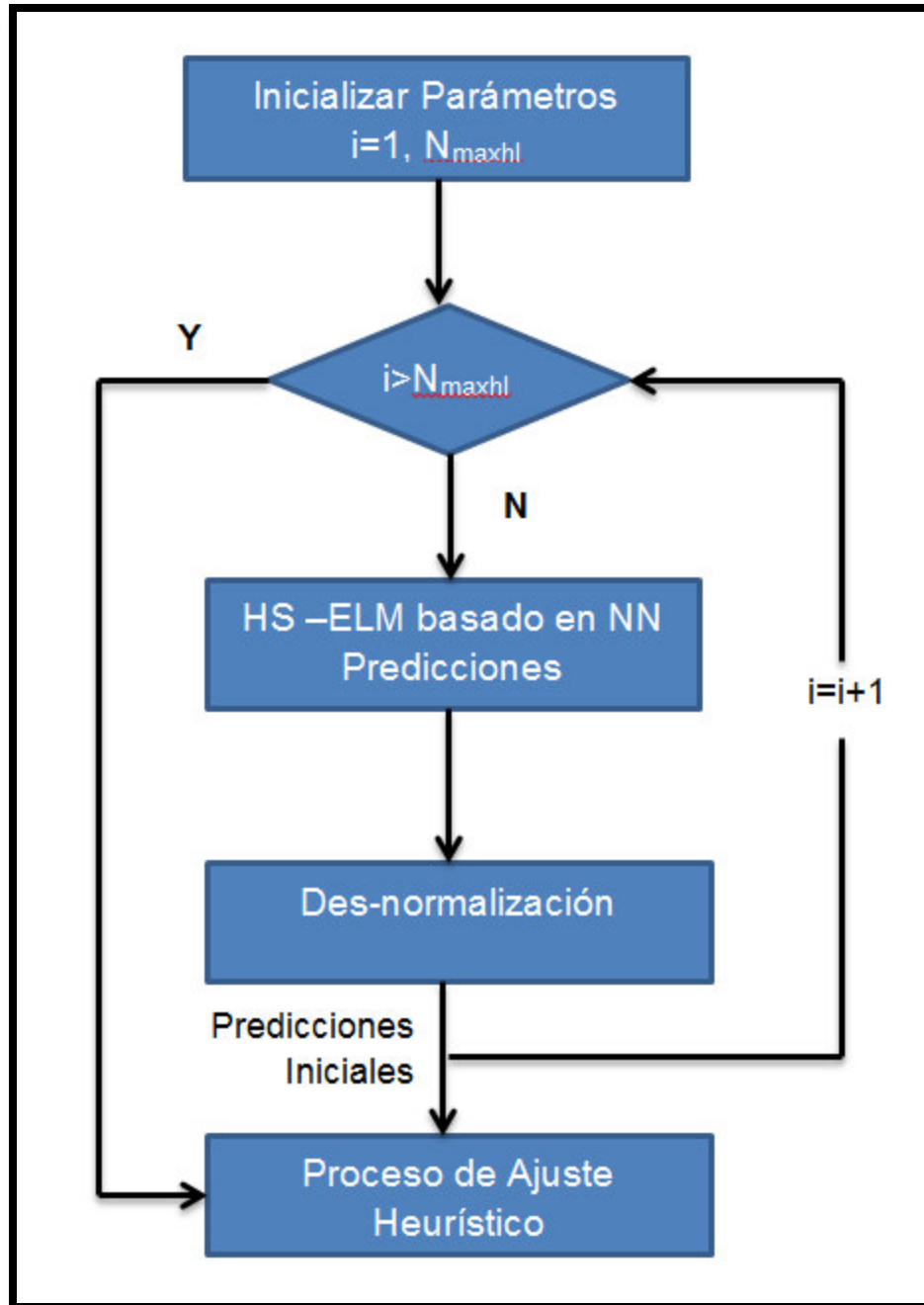


Grafico 6, Diagrama de Flujo del Pronosticador HI, (W.K. Wong, Z.X. Guo, 2010)

Se llevaron a cabo numerosos experimentos para validar el modelo HI propuesto en términos de datos de venta al por menor de artículos de moda real. Los resultados experimentales han demostrado que el modelo HI puede abordar el problema de la predicción de ventas a medio plazo con eficacia, lo que también demuestra que el modelo propuesto puede proporcionar un rendimiento muy superior sobre los modelos tradicionales ARIMA y dos modelos NN (Redes Neuronales) de predicción de ventas desarrolladas recientemente.

Además, ya que las series temporales que se abordaron en este trabajo poseen patrones tales como la irregularidad y la estacionalidad, el modelo propuesto es ampliamente aplicable y puede ser fácilmente ampliado para resolver otros problemas de predicción con patrones de series de tiempo similares. El modelo propuesto proporciona predicciones solo basándose en el historial de ventas, que no puede reflejar los efectos de los factores exógenos, tales como el clima y los índices económicos en las ventas de los artículos de moda.

2.2.2. Predicción de ventas para comerciantes informáticos: una comparación de Adaptación Multivariado Splines de Regresión y Redes Neuronales Artificiales

Según (Chi-Jie Lu a, Tian-Shyug Lee b, Chia-Mei Lian, 2012) las empresas que buscan tener un buen rendimiento de ventas por lo regular necesitan mantener un equilibrio entre la satisfacción de la demanda del cliente y los costos del control de inventario. Llevar un inventario muy grande permite satisfacer la demanda del cliente en todo momento, pero puede resultar perjudicial para la empresa ya que podría derivar en problemas como tener capital inmovilizado y la reducción de los márgenes de ganancias. Por otro lado menores niveles de inventario, pueden reducir los costos del inventario, pero pueden resultar en costos de oportunidad derivados de la pérdida de la oportunidad de venta y la reducción de la satisfacción del cliente. Es por ello que la predicción de ventas puede ser utilizada para determinar el nivel de inventario requerido y así evitar los problemas que provocan tener un alto o bajo nivel de inventario.

Para lograr obtener una predicción de ventas precisa, este estudio aplico cinco metodologías diferentes las cuales son la Multivariate Adaptive Regression Splines (MARS), Backpropagation neural network (BPN), Support vector regression (RVS), extreme learning machine (ELM) y Cerebellar model articulation controller neural network (CMACNN), y cuatro esquemas de predicción de dos etapas. En primer lugar las metodologías MARS, BPN, SVR, y CMACNN se aplican para la construcción de modelos de predicción de ventas utilizando todas las variables de predicción. En el caso de la

predicción de dos etapas, primero se utiliza la metodología MARS como una herramienta para la detección de las variables de predicción, después de lo cual se utilizan las variables de predicción obtenidos como variables de entrada para los modelos de predicción utilizando BPN, RVS, ELM, y CMACNN. Esto da lugar a cuatro modelos de predicción de dos etapas respectivamente llamados MARS-BPN, MARS-SVR, MARS-ELM, y MARS-CMACNN. Por otra parte, para evaluar plenamente el rendimiento del modelo MARS se compara su rendimiento con los métodos tradicionales como el modelo autorregresivo integrado de promedio móvil (ARIMA) y multivariate linear regression (MLR). Por último, para comparar los resultados de los modelos propuestos en este estudio, los experimentos se llevaron a cabo utilizando los datos de ventas mensuales de dos comerciantes informáticos taiwaneses.

El rendimiento de los 11 modelos de predicción comparado en este estudio es probado usando diferentes índices (ratios) de entrenamiento y tamaño de muestra para evaluar la solidez de la propuesta del método MARS. El plan de prueba se basa en el ratio relativo calculado como el tamaño de conjunto de datos de entrenamiento para el tamaño del conjunto de datos completa. Tres ratios relativos son considerados: 70%, 80%, y 90%. Los resultados de la predicción de datos de prueba para los 11 métodos examinados en este punto son resumidos en el Cuadro 3, en términos del error porcentual absoluto medio (MAPE) el cual es uno de los criterios de

rendimiento para evaluar las metodologías propuestas en este estudio.

Ratio relativo (%)	Modelo	Pruebas MAPE (%)
90	MARS	0.07
	SVR	0.18
	BPN	0.44
	CMACNN	0.24
	ELM	0.25
	ARIMA	0.53
	MLR	0.48
	MARS-SVR	0.14
	MARS-BPN	0.3
	MARS-CMACNN	0.17
	MARS-ELM	0.21
80	MARS	0.11
	SVR	0.32
	BPN	0.41
	CMACNN	0.36
	ELM	0.35
	ARIMA	0.56
	MLR	0.64
	MARS-SVR	0.25
	MARS-BPN	0.38
	MARS-CMACNN	0.36
	MARS-ELM	0.33

70	MARS	0.19
	SVR	0.37
	BPN	0.45
	CMACNN	0.4
	ELM	0.38
	ARIMA	0.54
	MLR	0.68
	MARS-SVR	0.3
	MARS-BPN	0.42
	MARS-CMACNN	0.33
	MARS-ELM	0.35

Cuadro 3, Evaluación de rendimiento de los modelos de predicción debajo de diferentes proporciones relativas, (Chi-Jie Lu a, Tian-Shyug Lee b, Chia-Mei Lian, 2012)

Del Cuadro 3, se puede ver que el modelo MARS tiene la MAPE más bajo para todos los ratios de prueba. Sus resultados de predicción son también mucho mejores que los producidos usando la otros métodos, indicando que la metodología MARS efectivamente proporciona una mejor precisión en la predicción que los otros diez enfoques y es adecuado para predicciones de venta de comerciantes informáticos.

2.2.3. Rendimiento de los Modelos de Espacio de Estado y ARIMA para la predicción de venta del consumo al por menor

De acuerdo con (Patrícia Ramos, Nicolau Santos, Rui Rebelo, 2015) la Predicción de las ventas es una de las cuestiones más importantes que está más allá de toda decisión estratégica o de planificación. Para las empresas de ventas rentables, la precisión de la predicción de la demanda es crucial en la organización. Las ventas de negocios minoristas pertenecen a un tipo especial de series temporales que normalmente contienen patrones de tendencia y estacionalidad, lo que presenta retos en el desarrollo de modelos de predicción eficaces.

En este estudio se compara el rendimiento de los modelos de predicción ARIMA y el modelo de Espacio de Estados. La capacidad de predicción es demostrado a través de un caso de estudio de las ventas al por menor de cinco categorías diferentes de calzado para mujeres de la marca Foreva las cuales son botas, botines, flats, sandalias y zapatos.

Los resultados de los cuadros 4, 5 y 6 muestran que en general el desempeño de las predicciones de los modelos ARIMA y ETS evaluados a través de medidas de precisión RMSE, MAE y MAPE son bastante similares tanto en las predicciones de un solo paso como en las predicciones de varios pasos. En ambos enfoques de modelado, las predicciones de varios pasos son generalmente mejor que las predicciones de un solo paso, lo cual es gracias a que

las predicciones de varios pasos incorporan información más actualizada.

Serie al por menor	Modelo	Pronostico de un paso	Pronostico de multiples pasos												
			1	2	3	4	5	6	7	8	9	10	11	12	
Botas	ETS	1263.63	1772.99	2424.99	2823.25	3629.93	3629.93	4575.80	4913.38	3532.30	2012.93	812.90	203.37	181.60	193.94
	ARIMA	1886.38	2315.46	2570.81	2866.74	2176.30	2307.08	2466.34	2269.76	2485.37	2657.74	1648.04	721.31	13.00	
Botines	ETS	1151.46	604.69	1056.24	1381.72	1733.74	2013.21	2074.64	1944.03	1418.13	665.95	384.70	59.27	87.77	
	ARIMA	843.68	538.86	719.85	943.74	1400.74	1594.39	1564.41	1141.17	671.83	131.48	53.90	19.49	14.57	
Flats	ETS	757.45	448.94	564.52	589.74	837.14	1048.20	1075.22	1218.73	1460.33	1682.60	1771.42	1909.65	1731.85	
	ARIMA	797.07	511.76	542.19	390.76	746.12	1049.51	1064.89	1152.75	1444.94	1719.01	1833.64	1944.21	1700.07	
Sandalias	ETS	1201.01	1279.47	804.48	894.17	280.57	1413.19	2587.91	1200.82	1032.87	1434.66	1920.06	1722.84	1857.38	
	ARIMA	1414.50	1506.71	1475.93	1536.06	1608.02	1608.02	1788.28	1925.44	2110.13	2359.24	2724.48	3335.86	4657.98	
Zapatos	ETS	651.72	624.38	738.61	628.96	732.80	904.56	962.67	827.87	949.67	1097.71	1220.35	1113.71	1178.26	
	ARIMA	798.62	876.86	1098.66	1103.78	1128.61	1121.56	1182.66	908.78	1058.72	1132.51	1212.91	1280.02	1236.24	

Cuadro 4, Predicciones de RMSE, (Chi-Jie Lu a, Tian-Shyug Lee b, Chia-Mei Lian, 2012)

Serie al por menor	Modelo	Pronostico de un paso	Pronostico de multiples pasos											
			1	2	3	4	5	6	7	8	9	10	11	12
Botas	ETS	690.63	948.26	1342.31	1854.88	2356.78	3019.70	3584.63	2488.65	1650.09	560.31	148.10	167.65	193.94
	ARIMA	1159.49	1257.29	1695.41	2115.84	1513.89	1676.37	1911.57	1660.83	1961.00	2045.75	1233.00	516.50	13.00
Botines	ETS	749.78	357.31	756.01	1011.60	1318.99	1464.28	1500.68	1484.99	1087.02	539.72	286.58	49.74	87.77
	ARIMA	539.75	324.15	526.36	677.33	972.66	1088.10	1018.14	836.41	481.61	106.55	47.65	19.48	14.57
Flats	ETS	513.90	323.46	428.50	472.94	588.44	683.03	769.54	938.54	1152.01	1323.08	1489.38	1894.98	1731.85
	ARIMA	601.65	369.83	448.94	337.28	503.29	715.21	782.96	916.03	1151.53	1376.13	1558.57	1928.18	1700.07
Sandalias	ETS	745.65	860.59	555.52	583.84	202.12	744.23	1042.48	548.40	527.79	747.01	1141.96	1350.99	1857.38
	ARIMA	713.32	750.49	744.50	761.65	788.56	845.32	836.24	924.47	1108.91	1380.73	1836.12	2702.99	4657.98
Zapatos	ETS	547.97	475.31	553.71	447.48	496.22	669.36	762.33	618.28	664.29	889.93	1137.13	994.69	1178.26
	ARIMA	683.03	666.59	831.30	854.60	838.40	835.83	868.51	634.15	804.72	876.07	1114.09	1129.78	1236.24

**Cuadro 5, Predicciones de MAE, (Chi-Jie Lu a, Tian-Shyug
Lee b, Chia-Mei Lian, 2012)**

Serie al por menor	Modelo	Pronostico de un paso	Pronostico de multiples pasos											
			1	2	3	4	5	6	7	8	9	10	11	12
Botas	ETS	211.03	130.02	199.10	100.39	67.97	70.70	74.97	61.63	63.42	32.13	12.96	103.58	340.24
	ARIMA	169.87	176.39	198.09	63.76	53.31	53.77	61.13	54.30	64.78	77.86	82.96	87.64	22.81
Botines	ETS	72.86	55.79	91.86	71.04	91.12	87.07	72.17	68.70	73.77	59.32	34.92	25.07	151.32
	ARIMA	85.63	66.09	98.38	45.64	50.19	46.05	42.43	44.98	42.51	36.25	14.00	19.66	25.12
Flats	ETS	49.26	31.28	43.12	46.75	43.27	42.45	48.15	61.59	63.09	58.49	53.49	62.32	54.15
	ARIMA	62.37	33.43	49.39	45.49	44.04	46.10	50.87	63.78	64.82	63.15	57.09	63.40	53.16
Sandalias	ETS	165.87	449.07	253.55	231.13	246.04	281.16	298.73	321.29	114.01	26.76	40.64	44.24	54.33
	ARIMA	68.55	1149.57	62.23	62.00	72.72	229.45	74.09	112.21	78.31	76.71	76.90	91.60	136.24
Zapatos	ETS	18.32	13.85	16.53	12.52	12.58	16.73	20.26	16.51	15.26	22.02	29.18	19.48	22.40
	ARIMA	23.79	20.91	25.57	25.85	22.36	21.54	22.08	16.03	19.55	20.86	27.20	22.09	23.50

Cuadro 6, Predicciones de MAPE, (Chi-Jie Lu a, Tian-Shyug Lee b, Chia-Mei Lian, 2012)

2.2.4. Una aplicación mejorada del modelo Lotka-Volterra para pronosticar las ventas de dos competidores minoristas.

Según (Hui-Chih Hung, Yu-Chih Chiu, Huang-Chen Huang, Muh-Cherng Wu, 2017) existen varios tipos de minoristas que compiten entre sí ofreciendo diferentes propuestas de valor a los consumidores. El estudio que realizaron (Hui-Chih Hung, Yu-Chih Chiu, Huang-Chen Huang, Muh-Cherng Wu, 2017) abordó dos paradigmas principales de los minoristas en Taiwán. El primer paradigma es el orientado a la conveniencia, el cual atrae a los consumidores ahorrando tiempo a través de un fácil acceso a los bienes o servicios. El segundo paradigma es el orientado al presupuesto, el cual atrae a los consumidores al permitirles ahorrar dinero a través de la compra de un alto volumen de bienes o servicios.

Tradicionalmente, los clientes tienden a comprar con precios bajos. Sin embargo, una de las tendencias notables en la industria minorista reciente es que más personas están cambiando sus hábitos de compra desde un comportamiento orientado al presupuesto hasta un comportamiento orientado a la comodidad debido a la rápida urbanización y al ritmo de vida acelerado. En Taiwán las principales tiendas orientadas a la conveniencia son 7-Eleven™ y FamilyMart™. Ellos operan 24 horas al día, 7 días a la semana para permitir a los clientes acceder a los bienes o servicios a poca distancia en cualquier momento. Sus Servicios incluyen impresión, fax, venta de boletos, cobranza de facturas, Entregas, solicitudes de licencia de conducir, lavandería, etc. El paradigma orientado al presupuesto en Taiwán

comprende dos canales: Hipermercados y supermercados. Los hipermercados suelen estar ubicados en zonas sub-urbanas con amplio aparcamiento. Venden grandes cantidades de mercancías a bajo precio a los clientes durante una visita de compras. Dos hipermercados de representación en Taiwán son CostcoTM y RT-MartTM.

El objetivo del estudio que realizaron (Hui-Chih Hung, Yu-Chih Chiu, Huang-Chen Huang, Muh-Cherng Wu, 2017) fue de desarrollar un modelo de predicción de ventas que pueda analizar los efectos de interacción de los dos tipos de competidores minoristas (paradigma orientado a la conveniencia y orientado al presupuesto). Los ejecutivos de la industria minorista deben predecir con exactitud los ingresos y comprender los efectos de interacción entre los tipos de competidores minoristas para tomar las decisiones de inversión y operativas apropiadas.

Un enfoque típico para realizar una predicción de ventas y analizar los efectos de interacción de dos especies competidoras en un ecosistema empresarial se basa en las ecuaciones de Lotka-Volterra. Las ecuaciones de Lotka-Volterra (denotadas como el modelo LV de aquí en adelante) pueden aplicarse a un ecosistema empresarial, en el que la población representa la cantidad de ventas o los ingresos por ventas. Dos especies pueden ser dos tecnologías competidoras, dos productos competidores, dos canales competidores, o dos especies competidoras en un ecosistema empresarial. El modelo LV, que formula la dinámica de un ecosistema en el que compiten dos o múltiples especies, se utiliza para predecir la población de cada especie y revelar los efectos de

interacción entre diferentes especies competidoras. Sin embargo las ventas mensuales de los minoristas suelen estar incluidos con variaciones estacionales y tendencias. Es por ello que (Hui-Chih Hung, Yu-Chih Chiu, Huang-Chen Huang, Muh-Cherng Wu, 2017) proponen un framework de predicción de ventas (una aplicación mejorada del modelo LV), en donde se considera que el historial de ventas son datos compuestos que se descomponen en tres componentes individuales los cuales son:

- Agregado
- Competencia
- Componentes estacionales

El modelo LV se utiliza para predecir el componente de la competencia; Los otros dos componentes predicen mediante métodos de series de tiempo típicos; Y finalmente los 3 componentes se combinan en uno, tal y como se muestra en el grafico 7.

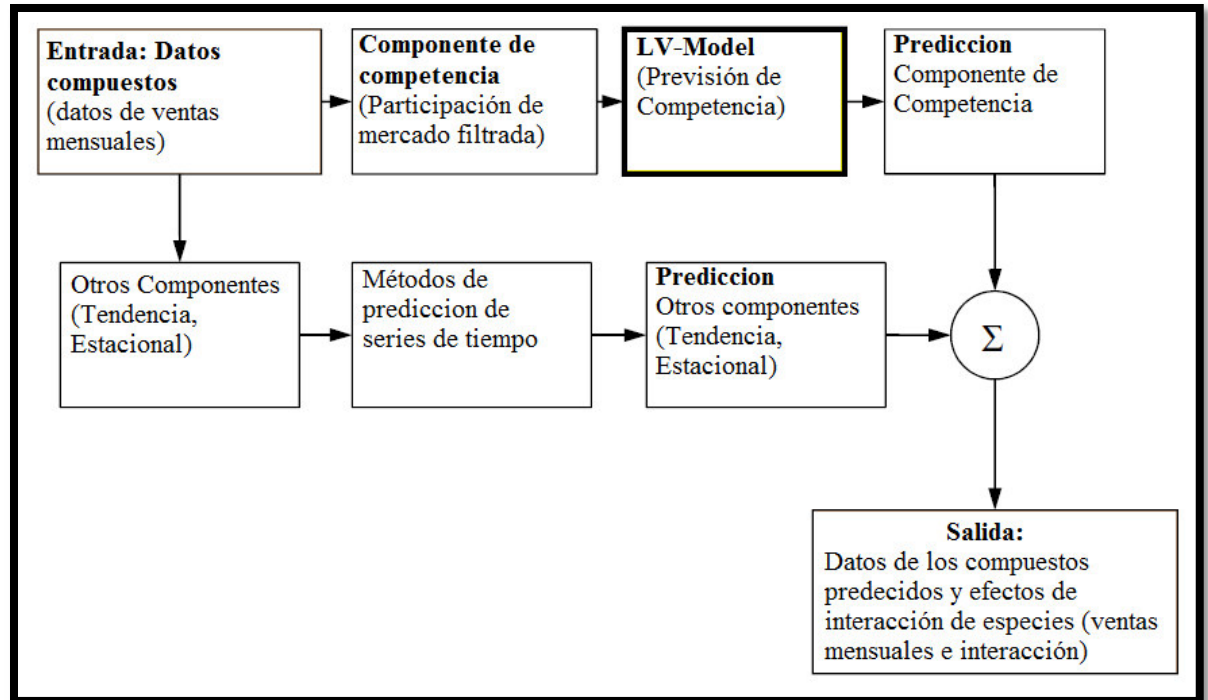


Grafico 7, Framework de predicción propuesto del modelo LV, (Hui-Chih Hung, Yu-Chih Chiu, Huang-Chen Huang, Muh-Cherng Wu, 2017)

Para estudiar los efectos de interacción entre especies competidoras, se comparó los resultados del enfoque propuesto con el modelo tradicional. Los errores de predicción y la racionalidad de los efectos de interacción entre las dos especies competidoras se toman como criterios de desempeño para comparar el método propuesto (lo que conlleva predecir tomando los datos del componente de competencia como entrada al modelo LV) frente al enfoque tradicional tomando los datos compuestos como entrada al modelo LV).

Se realizó un estudio empírico basado en los datos de ventas mensuales de los dos tipos de minoristas competidores en Taiwán, de enero de 1999 a diciembre de 2014 (Ministerio de Economía, Taiwán). Los errores de predicción y la racionalidad de los efectos de interacción entre las dos especies competidoras se toman como criterios de desempeño para comparar el Framework propuesto (lo que conlleva predecir tomando los datos del componente de competencia como entrada al modelo LV) frente al enfoque tradicional tomando los datos compuestos como entrada al modelo LV).

A continuación se muestra el Framework propuesto de forma detallada en el gráfico 8

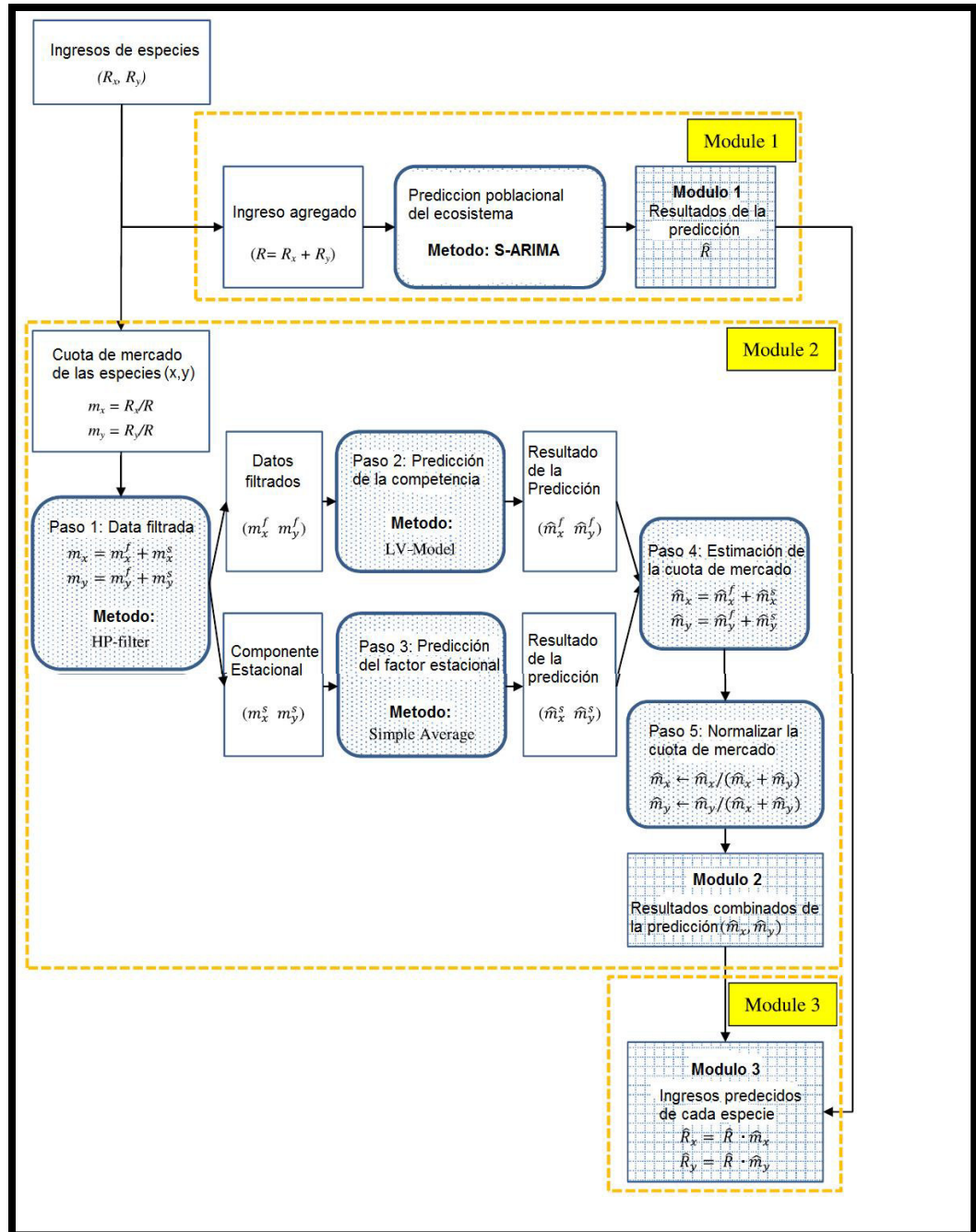


Grafico 8, Pasos detallados del Framework de predicción propuesto, (Hui-Chih Hung, Yu-Chih Chiu, Huang-Chen Huang, Muh-Cherng Wu, 2017)

Los detalles del framework de predicción propuesto por (Hui-Chih Hung, Yu-Chih Chiu, Huang-Chen Huang, Muh-Cherng Wu, 2017) se muestran en el gráfico 8, en el cual se incluyen tres módulos. Las dos especies competidoras son denotadas por X (formato orientado al presupuesto) y Y (formato de venta al por menor orientado a la conveniencia); x denota el ingreso de X e y denota el ingreso de Y. De los tres módulos, el primer módulo predice la población del ecosistema (\hat{R}), el segundo predice la cuota de mercado de los dos formatos al por menor (\hat{m}_x y \hat{m}_y) y el tercero combina las salidas de los dos primeros módulos y obtiene la predicción de ingresos de cada formato minorista. A continuación se describen los detalles de cada módulo.

Módulo 1: Pronóstico de la población del ecosistema

La entrada del Módulo 1 son los ingresos agregados de los dos formatos minoristas competidores ($R = R_x + R_y$) también llamado población del ecosistema. El resultado es la predicción de los ingresos agregados (\hat{R}). En el módulo 1 se adopta el modelo S-ARIMA.

Módulo 2: Predicción de cuota de mercado

Las cuotas de mercado de los dos formatos de venta al por menor (m_x y m_y) se utilizan como entrada del módulo 2, mientras que las cuotas de mercado predichas de los dos formatos de venta al por menor (\hat{m}_x y \hat{m}_y) sirven como salida del módulo.

Módulo 3: Predicción de la población de especies y métricas de error

El módulo 3 obtiene la predicción de los ingresos de cada formato de venta al por menor

$(\widehat{R}_x = \widehat{m}_x \cdot \widehat{R}$ y $\widehat{R}_y = \widehat{m}_y \cdot \widehat{R})$ donde \widehat{R} es la salida del módulo 1 y $(\widehat{m}_x$ y $\widehat{m}_y)$ son las salidas del módulo 2. La métrica para medir el error de predicción es el error de porcentaje absoluto medio (MAPE).

El estudio empírico indicó que el método propuesto por (Hui-Chih Hung, Yu-Chih Chiu, Huang-Chen Huang, Muh-Cherng Wu, 2017) supera sustancialmente el enfoque tradicional en términos de errores de predicción (4.4% frente a 16.7% para el paradigma orientado a la conveniencia y 5.8% frente 16.2% para el paradigma orientado al presupuesto).

Esta investigación realizada por (Hui-Chih Hung, Yu-Chih Chiu, Huang-Chen Huang, Muh-Cherng Wu, 2017) hace una contribución en cómo aplicar apropiadamente el modelo LV para predecir la población de especies competidoras, particularmente cuando la población está sustancialmente afectada por factores estacionales y ambientales.

2.2.5. Predicción de Ventas de Productos utilizando las revisiones en línea y los datos históricos de ventas: Un método que combina el modelo de BASS y el análisis de sentimientos

De acuerdo con (Zhi-Ping Fan, Yu-Jie Che, Zhen-Yu Chen, 2017) las revisiones en línea proporcionan a los consumidores información rica que puede reducir su incertidumbre con respecto a las compras. Como tales, estas revisiones tienen una influencia significativa en las ventas del producto. Según (Marshall Dockendorff, Ibáñez, 2013) las empresas utilizan las predicciones de ventas de productos como una base para estimar los ingresos por ventas y tomar decisiones con respecto a las estrategias de producción, operación y marketing. Conforme a (Mentzer J.T, Bienstock C.C, 1998) a través de la predicción de ventas de productos, las empresas pueden crear un plan de comercialización, gestión de ventas, producción, adquisición, logística y así sucesivamente para mejorar sus beneficios económicos y reducir las pérdidas causadas por las debilidades del plan de producción. Dos factores principales influyen en las decisiones de compra de los consumidores las cuales son:

- La influencia de otros consumidores que han comprado el producto y lo han recomendado a través de la comunicación verbal.
- La influencia de los anuncios y los medios de comunicación, entre otros factores.

Varios investigadores han estudiado las predicciones de ventas de productos y han desarrollado modelos de predicción eficaces que tienen en cuenta factores relevantes. Entre ellos están el modelo de Bass (Bass F. M, 1969), el cual toma en cuenta los factores que influyen en las

decisiones de compra mencionados anteriormente (influencia de otros consumidores, influencia de anuncios) como coeficientes externos e internos.

Según (Herr Kardes, Kim, 1991) el boca a boca (WOM) es considerado uno de los factores más importantes que influyen en las decisiones de compra de los consumidores, especialmente con respecto a los imitadores. Además de acuerdo con (Wysocki P.D, 2000; Ryu K, Han H, 2010) las revisiones en línea y micro blogs, se han popularizado con el desarrollo de las tecnologías de Internet, una serie de e-commerce websites como Amazon y Taobao han establecido sistemas de revisión en línea para animar a los consumidores a publicar revisiones de productos y como han cambiado gradualmente los patrones de comportamiento de los consumidores y afectado las decisiones de compra de los consumidores. Por ejemplo, los consumidores están prestando cada vez más atención a las opiniones en línea al decidir qué películas mirar, en qué acciones deben invertir, etc. Además, muchas comunidades en línea, como Facebook y Douban, proporcionan plataformas para discusiones de consumidores. Estas revisiones revelan a menudo emociones personales, tales como felicidad, cólera, dolor, crítica y alabanza, y los consumidores potenciales pueden hojear las opiniones públicas sobre un producto para informar sus decisiones de compra. En consecuencia, en la última década se han utilizado técnicas de análisis del sentimiento para medir los sentimientos transmitidos a través del contenido de las reseñas en línea.

Como se ha indicado por (Yu X, Liu Y, Huang An 2012), el índice de sentimiento extraído del contenido de las revisiones en línea por técnicas


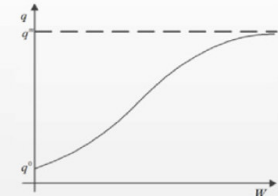
de análisis de sentimientos puede utilizarse para predecir muchos fenómenos económicos sociales, incluyendo las cuotas de mercado de productos, información o enfermedades.

Varios investigadores han explorado la relación entre las revisiones en línea y las ventas de productos (Chevalier J. A, Mayzlin D, 2006; Dellarocas C, Awad N. F, Zhang X, 2007) y han desarrollado métodos de predicción que combinan el modelo de Bass y las puntuaciones de WOM en línea (Wu M, Wang L, Li M, 2015), sin embargo estos modelos han utilizado sólo las ventas históricas y los datos de calificación para predecir las ventas del producto. Pocos estudios han desarrollado versiones mejoradas de los modelos de Bass para considerar los sentimientos expresados en el contenido de las revisiones en línea, es por esta razón que (Zhi-Ping Fan, Yu-Jie Che, Zhen-Yu Chen, 2017) desarrollaron un nuevo Framework que combina el modelo de Bass/Norton y el análisis de sentimientos al utilizar los datos históricos de ventas y los datos de revisión en línea, dicho Framework se ilustra en el grafico 9:

Paso 1 Revisar la recopilación de datos y el preprocesamiento

<p>Datos de revisión del producto - La estructura de los datos revisados</p>	<p>Preprocesamiento de datos - Segmentación de palabras - Estadísticas de frecuencia de palabras</p>
	

Paso 2 Extracción de sentimientos y construcción de modelos

<p>Análisis de los sentimientos - Algoritmo Naive Bayes (NB) - Diccionario de sentimientos CNKI</p>	<p>Modelo de predicción - Modelo de Bass-Emotion - Modelo de Norton-Emotion</p>
	

Paso 3 Modelo de Ajuste y Evaluación


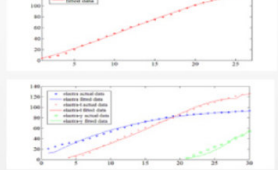
<p>Recolección de datos y ajuste del modelo - www.pcauto.com.cn - R² & RMSE</p>	<p>Comparación y evaluación de modelos - Modelo Bass y Modelo Norton - MAPE y PE</p>
	

Grafico 9, Framework Propuesto que combina el modelo de Bass/Norton y el análisis de sentimientos, (Chevalier J. A, Mayzlin D, 2006, Dellarocas C, Awad N. F, Zhang X, 2007)

A continuación se detallan pasos del Framework que se muestran en el gráfico 9:

- Recolección de datos y pre-procesamiento. En este paso, productos con generaciones tecnológicas múltiples son seleccionados y se recogen los datos de ventas junto con sus revisiones en línea. Se extrajeron los siguientes atributos para cada revisión en línea
 - o ID de revisor
 - o Título
 - o Clasificación
 - o Hora
 - o Número de veces que los usuarios navegaron por la revisión.
 - o Número de usuarios que estuvieron de acuerdo con la revisión.
 - o Número de usuarios que se opusieron a la revisión.
 - o Contenido de la revisión.

Los métodos de segmentación de palabras y las estadísticas de frecuencia de palabras se utilizan para el pre-procesamiento de datos.

- Extracción de índices de sentimiento y construcción de modelos de predicción. En este paso, se utiliza el método de análisis de sentimiento llamado Naive Bayes(NB), para analizar los datos de revisión y calcular el índice de sentimiento basado en el punto de tiempo. Se desarrolla un nuevo Framework de predicción que

combina el modelo de Bass y Norton con el índice de sentimiento extraído para mejorar la precisión de la predicción.

- Validación del rendimiento. En este paso se ajusta el Framework propuesto y se evalúa el desempeño de la predicción usando medidas específicas como el error de porcentaje absoluto medio (MAPE) y el porcentaje de error (PE). Al mismo tiempo los resultados se comparan con los del modelo estándar de Bass y Norton.

De acuerdo a (Wang F. K, Chang K. K, Hsiao Y. Y, 2013) el modelo Bass es adecuado para predecir las ventas de un producto para el cual el número de usuarios es aproximadamente igual a las ventas del producto. En cuanto al modelo Norton, sin embargo, se debe seleccionar un producto con múltiples generaciones. Por lo tanto, en este estudio se eligió los automóviles de la empresa automotriz Bitauto (Plataforma de auto-trading de China) como objeto de investigación, para los cuales se recogieron los datos de ventas y las revisiones en línea de cada generación. Se tomaron en cuenta tres generaciones de la marca Beijing Hyundai Elantra las cuales llamaron Elantra, Elantra-y y Elantra-I. Para asegurar su consistencia, los datos de ventas y revisión en línea fueron elegidos del mismo período de tiempo: datos de Elantra de julio de 2007 a diciembre de 2014, del Elantra-y de abril de 2008 a diciembre de 2014 y del Elantra-I de agosto de 2012 a diciembre de 2014. Se recopilaron 1407, 2524 y 368 revisiones de Elantra, Elantra - y, Elantra - I respectivamente.

Luego del experimento, los resultados computacionales indican que la combinación del modelo de Bass / Norton y el análisis de sentimiento tiene mayor precisión de predicción que el modelo estándar de Bass / Norton y algunos otros modelos de predicción de ventas. En base a lo desarrollado por (Zhi-Ping Fan, Yu-Jie Che, Zhen-Yu Chen, 2017) se puede llegar a la conclusión de que se puede obtener mayor precisión en las predicciones de ventas si se usa adecuadamente el índice de sentimiento en los modelos Bass/Norton

2.3. Cuadro Comparativo

En éste apartado se seleccionara y describirá los factores de evaluación o atributos de comparación, los cuales nos ayudaran a evaluar de forma precisa y exhaustiva todas las soluciones descritas anteriormente, esta evaluación se llevara a cabo gracias a un cuadro comparativo y posteriormente se llegara a una conclusión de la evaluación de los trabajos expuestos. A continuación se describirá los factores de evaluación que fueron seleccionados para esta evaluación:

2.3.1. Factores de evaluación

- a. Sector de enfoque: este factor de evaluación nos indica cual es el sector de mercado donde se enfoca la solución.
- b. Rendimiento en las predicciones de venta de productos de corto ciclo de vida: este factor de evaluación nos indica cual sería el rendimiento de la solución cuando se estudian productos de corta vida, por ejemplo artículos de moda.

- c. Rendimiento en las predicciones de venta de productos de gran estacionalidad: este factor de evaluación nos indica cual sería el rendimiento de la solución cuando se estudian productos que tengan una fuerte estacionalidad por ejemplo la ropa, ya que se venden más chompas en invierno que en verano.
- d. Refinamiento de las predicciones de ventas: este factor de evaluación nos indica si la solución tiene algún mecanismo de refinado de su predicción, una vez que se haya terminado de procesar la solución.
- e. Componente de Pre-Procesamiento: este factor de evaluación nos indica si la solución tiene algún componente de Pre-Procesamiento.
- f. Revisiones en Línea: Uso de datos de revisiones en Línea.
- g. Objetivo: este factor de evaluación nos indica cual es el objetivo que se persigue la solución.
- h. Validación: este factor de evaluación nos indica cuales fueron las pruebas que se realizaron a las soluciones para garantizar su efectividad.
- i. Medidas de Precisión: este factor de evaluación nos indica el rendimiento de cada metodología de Predicción de Ventas.

2.3.2. Trabajos Evaluados

- A. Un Modelo Inteligente Híbrido para la Predicción de Ventas a mediano plazo en las cadenas de suministro de moda de venta al por menor utilizando una máquina de aprendizaje extremo y un algoritmo de búsqueda armonía.
- B. Predicción de ventas para comerciantes informáticos: una comparación de Adaptación Multivariado Splines de Regresión y Redes Neuronales Artificiales.
- C. Rendimiento de los Modelos de Espacio de Estado y ARIMA para la predicción de venta del consumo al por menor.
- D. Una aplicación mejorada del modelo Lotka-Volterra para pronosticar las ventas de dos competidores minoristas.
- E. Predicción de Ventas de Productos utilizando las revisiones en línea y los datos históricos de ventas: Un método que combina el modelo de BASS y el análisis de sentimientos

2.3.3. Cuadro Comparativo

En esta sección se realizara un análisis de las ventajas, desventajas y beneficios de cada uno de los trabajos mencionados anteriormente los cuales dan solución al problema planteado en esta tesis. A continuación se presenta el Cuadro 7, donde se muestra un cuadro comparativo tomando en cuenta los criterios de evaluación antes definidos:

Solución Factores	Solución del Trabajo A	Solución del Trabajo B	Solución del Trabajo C	Solución del Trabajo D	Solución del Trabajo E	Factor Deseable
Sector de enfoque	PYMES que comercialicen artículos de moda	Comerciantes de Productos Informáticos	PYMES	Comerciantes minoristas en Taiwán	Comerciantes Minoristas	'C'
Rendimiento en las predicciones de venta de productos de corto ciclo de vida.	La solución presenta un gran rendimiento en este aspecto ya que por lo general los artículos de moda tienen corto ciclo de vida.	La solución proporciona un gran rendimiento en este aspecto ya que los productos informáticos se renuevan constantemente.	No aplica	No aplica	No aplica	'A'
Rendimiento en las predicciones de venta de productos de gran estacionalidad	La solución presenta un gran rendimiento en este aspecto ya que por lo general los artículos de moda son de gran estacionalidad.	No aplica	Presenta un rendimiento aceptable ya que el desarrollo del modelo se realizó en base a las ventas de calzado para mujeres.	El framework propuesto considera que el historial de ventas son datos compuestos que se descomponen en "Agregado", "Competencia" y "Estacionalidad", por lo que se puede obtener buen rendimiento en productos de	No aplica	'A'

				gran estacionalidad.		
Refinamiento de las predicciones de ventas	Presenta gran precisión en sus predicciones ya que al finalizar del proceso, se utiliza un ajuste heurístico para obtener predicciones más precisas.	No aplica	No aplica	No aplica	En el tercer paso del Framework propuesto se realiza un ajuste del modelo, para obtener predicciones de ventas más precisas.	'A', 'E'
Componente de Pre-Procesamiento	Presenta un componente de Pre-Procesamiento para eliminar los valores atípicos que pudieran afectar de forma negativa la predicción de ventas	No aplica	No aplica	No aplica	En el primer paso del Framework propuesto se realiza un pre-procesamiento de datos donde se usan los métodos de segmentación de palabras y las estadísticas de frecuencia de palabras	'A'
Uso de datos de revisiones en Línea	No aplica	No aplica	No aplica	No aplica	El Framework propuesto usa los datos de revisiones en línea para obtener el índice de sentimiento	-
Objetivo	Proporcionar a las pymes de un modelo de	Proporcionar a los comerciantes de un modelo de	Proporcionar a las PYMES de un modelo de	Desarrollar un Framework de predicción de	Desarrollar un Framework de Predicción de	

	predicción de ventas que sea flexible y lo suficientemente robusto, para las distintas características de los artículos de moda.	predicción de ventas que le ayuden a mejorar su rendimiento en las ventas.	predicción de ventas que le permita mejorar su gestión en los distintos departamentos de la organización.	ventas que pueda analizar los efectos de interacción de los dos tipos de competidores minoristas (Orientado a la conveniencia y al presupuesto)	ventas que combine el modelo de Bass/Norton y el análisis de sentimientos al utilizar los datos históricos de ventas y los datos de revisión en línea	'A', 'B', 'C'
Validación	Se realizaron pruebas donde la predicción de ventas del modelo propuesto demostró ser superior en comparación a las predicciones de ventas del modelo ARIMA y otros dos modelos basados en redes neuronales.	Se realizaron pruebas con distinto número de ratios donde el modelo de predicción de ventas propuesto presento mejores resultados.	Se realizaron pruebas donde el modelo propuesto presento un rendimiento superior en casi todos los productos y medidas de precisión.	Se realizaron pruebas empíricas que indicaron que el Framework propuesto supera sustancialment e el enfoque tradicional (ecuaciones de Lotka-Volterra) en términos de errores de predicción (4.4% frente a 16.7% para el paradigma orientado a la conveniencia y 5.8% frente 16.2% para el paradigma	Se realizaron pruebas en donde se obtuvo como resultados que la combinación del modelo de Bass / Norton y el análisis de sentimiento tienen mayor precisión de predicción que el modelo estándar de Bass / Norton y algunos otros modelos de predicción de ventas.	-

				orientado al presupuesto)		
Medidas de Precisión	RMSE, MAPE y MASE	RMSE, MAD, MAPE y RMSPE	ME, MAE y RMSE	MAPE	MAPE y PE	'A'

Cuadro 7, Cuadro Comparativo de los trabajos analizados, (Fuente: Elaboración Propia)

2.3.4. Análisis del cuadro comparativo

Como se puede ver en el cuadro 7, las distintas soluciones que se utilizan actualmente nos proporcionan predicciones de ventas con gran exactitud, de las cuales se seleccionaron factores deseables de cada una de ellas y se resumen en el cuadro 8.

Factores Deseables	Trabajos Seleccionados
Sector de enfoque	Trabajo 'C'
Rendimiento en las predicciones de venta de productos de corto ciclo de vida.	Trabajo 'A'
Rendimiento en las predicciones de venta de productos de gran estacionalidad	Trabajo 'A'
Refinamiento de las predicciones de ventas	Trabajo 'A' y 'E'
Componente de Pre-Procesamiento	Trabajo 'A'
Uso de datos de revisiones en Línea	-
Objetivo	Trabajo 'A', 'B' y 'C'
Medidas de Precisión	Trabajo 'A'

Cuadro 8, Cuadro resumen de Factores Deseables para el Framework de Predicción de Ventas, (Fuente: Elaboración Propia)

A pesar de que las soluciones de predicciones de ventas presentadas tienen una buena precisión en el ámbito en el que se desarrollan, estas no funcionan eficientemente en todos los casos, es así que se puede llegar a la conclusión que si la solución funciona para un determinado tipo de producto no es garantía que funcione con otros tipos de productos.

En la mayoría de soluciones no se toma en cuenta los productos que tengan corto tiempo de vida o que tengan gran estacionalidad, por otro lado las soluciones no cuentan o tienen un deficiente desarrollo cuando se trata de refinar los resultados de las predicciones de ventas de sus respectivos modelos propuestos, es por ello que en el siguiente capítulo se planteó un Framework de predicción de ventas que considere los factores presentados en el cuadro 8, además que tome en cuenta características propias de las series temporales como la periodicidad, tendencia y estacionalidad de los productos.

3. CAPITULO 3: METODOLOGIA

En este capítulo se desarrolló la solución al problema planteado, para ello en primer lugar se describirá la metodología, luego en base a dicha metodología se utilizara las ventas mensuales de 2 productos de una empresa dedicada a la venta de auto-partes llamada “Franco” y se generaran sus respectivos modelos de predicción. Luego para demostrar que usando el Framework propuesto se pueden obtener mejores resultados en las predicciones de ventas, se compararan las medidas de error y las ventas predichas previamente con los resultados de los modelos de predicción de los productos seleccionados únicamente usando MARS.

3.1. Definición de la Metodología Propuesta

El modelo de Predicción, las medidas de error y las Predicciones mensuales se obtuvieron usando la metodología descrita en el Grafico 10:



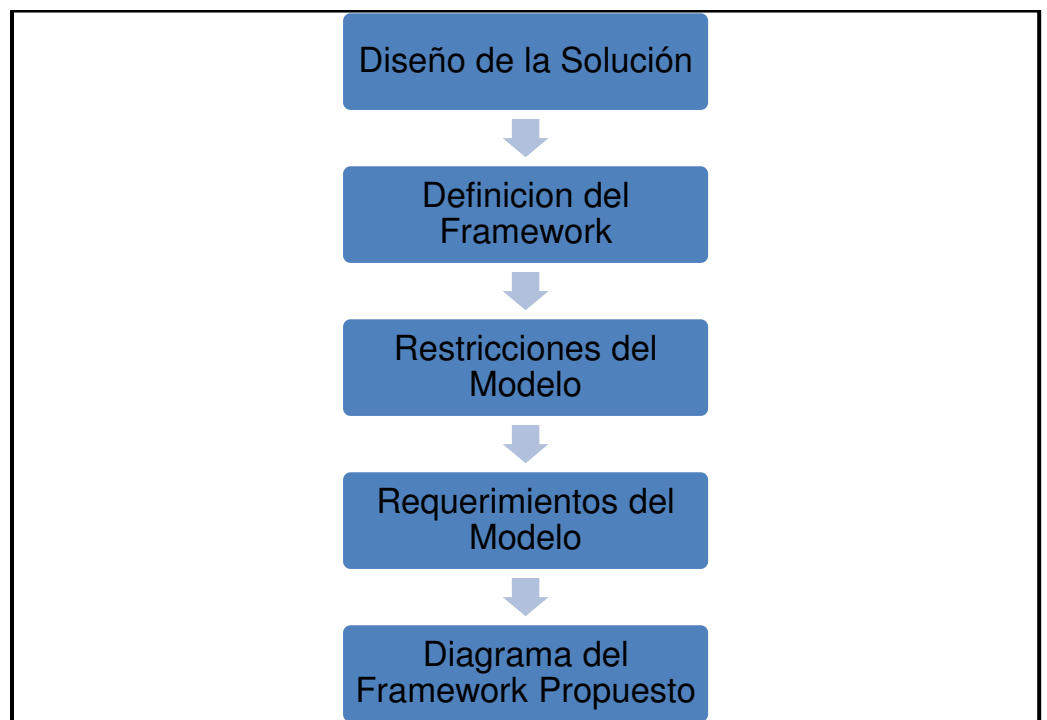
Grafico 10, Metodología Propuesta, (Fuente: Elaboración Propia)

3.2. Aporte Teórico

3.2.1. Contribución Teórica

En éste apartado se procederá a describir el Framework que se propuso para dar solución al problema planteado.

En el grafico 11 se muestra los componentes que conforman esta sección, los cuales se describirán con mayor detalle más adelante.



**Grafico 11, Componentes de la Contribución Teórica,
(Fuente: Elaboración Propia)**

3.2.1.1. Diseño de la Solución

Para dar una solución eficiente al problema planteado se utilizó en primer lugar un componente de pre-procesamiento de datos compuesto por 2 etapas el cual tiene un impacto significativo en el rendimiento del Framework, ya que disminuirá los errores en la predicción que conllevaría usar muestras no fiables o influenciados por diversos factores externos impredecibles.

Luego se usara MARS debido a que esta metodología de regresión no lineal incorpora interacciones entre variables, es por eso que puede ser una herramienta flexible capaz de automatizar la construcción de modelos de predicción. Además de acuerdo al estudio realizado por (Chi-Jie Lu a, Tian-Shyug Lee b, Chia-Mei Lian, 2012), se demostró que con la metodología MARS se puede obtener predicciones de ventas más precisas inclusive comparándolas con redes Neuronales, tal y como se mostró en el cuadro 3, en la sección del Estado del Arte.

El software que se usó fue “SALFORD PREDICTIVE MODELER 8.0” debido a que este software puede generar modelos de predicción de forma rápida e intuitiva usando la metodología MARS. Después de generar los modelos de predicción se corrobora la eficiencia de la solución propuesta realizando las siguientes validaciones:

- Que las medidas de error sean menores para los modelos de predicción generados usando el Framework propuesto, en comparación con otros modelos de predicción generados usando únicamente MARS. Las medidas de error que se usaron son “RMSE”, “MAD” y “MAPE”.
- Que las predicciones mes a mes de los modelos de predicción usando el Framework propuesto son más precisas que las predicciones de los modelos generados usando únicamente MARS.

3.2.1.2. Definición del Framework

Para la implementación de la solución propuesta se tomó como premisa que las ventas en un determinado intervalo de tiempo pueden ser predichos a partir de las ventas de meses anteriores. Es por ello que el problema planteado será tratado como una serie temporal y en consecuencia presentara los factores que rigen el comportamiento de las series temporales. En esta tesis se tomó en cuenta los factores periodicidad, tendencia y la estacionalidad.

El factor periodicidad se refiere a cada cuanto tiempo se toman los datos del historial de ventas, por ejemplo pueden ser semanales, mensuales o anuales. Por otro lado el factor tendencia de la serie, se puede obtener de los niveles de ventas históricos para un número determinado de periodos y analizarlos en busca de patrones de tendencia presentes en la

serie. Por ultimo para el factor de estacionalidad, se pueden considerar periodos de tiempo específicos de distintos años. Como se puede ver toda la información que se necesita para describir adecuadamente los factores que afectan la serie temporal, se pueden obtener de la misma serie temporal en estudio.

3.2.1.3. Restricciones del Modelo

La predicción de ventas se realizó para intervalos de tiempos mensuales ya que por lo general las PYMES realizan las predicciones de sus ventas para estos periodos de tiempo.

3.2.1.4. Requerimientos del Modelo

Para el desarrollo del Framework fue necesario disponer de un conjunto de datos de ventas históricos de un considerable número de meses, esto es muy importante ya que la precisión de la predicción de ventas dependerá en gran medida de la cantidad de información que dispongamos.

3.2.1.5. Diagrama del Framework Propuesto

A continuación se muestra en el Grafico 12 un diagrama donde se presentara las etapas que conforman el Framework propuesto.

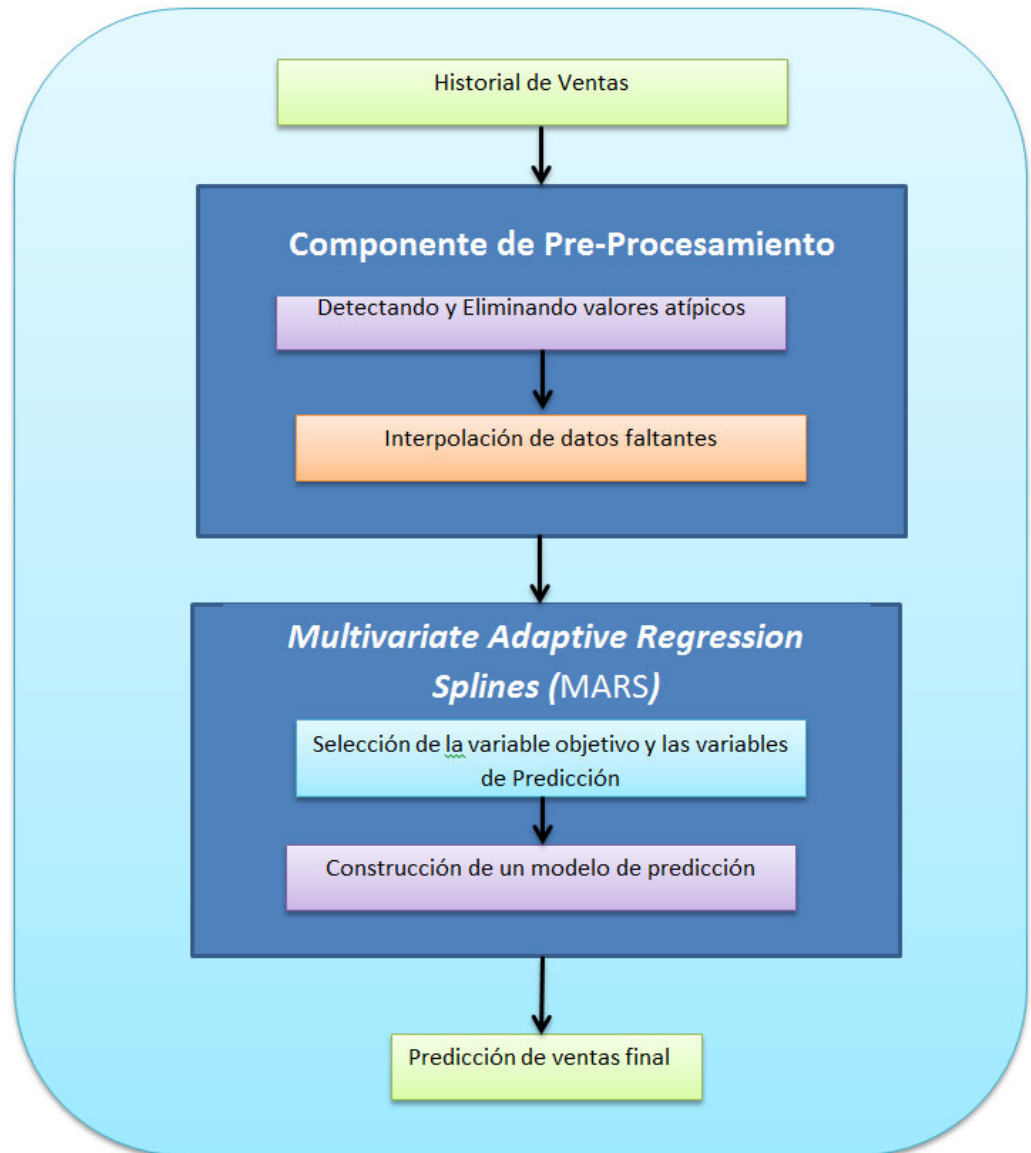


Grafico 12, Framework propuesto, (Fuente: Elaboración Propia)

A continuación se describirá los componentes del Framework propuesto.

3.2.1.5.1. Componente de Pre-Procesamiento

Los datos de ventas suelen estar influenciados por impredecibles factores externos lo cual afecta de forma negativa el rendimiento de los modelos de predicción, es por ello que se propuso usar un componente de Pre-Procesamiento de datos para disminuir los valores atípicos de la muestra de datos.

- **Detectando y Eliminando Valores Atípico**

Un valor atípico es aquel que se desvía mucho del resto de los datos de la muestra de tal manera que se sospeche que fue generado por un factor externo o mecanismo diferente. Para detectar los valores atípicos lo primero que se realizó fue obtener las ventas mensuales de un producto en específico. Luego se tomó un mismo periodo de tiempo de todos los años que se tenga la muestra y se obtuvo la media y la desviación estándar tal y como se ve en el Cuadro 9.

AÑO	MES	CANT_VENDID	MEDIA	DESVIACION ESTÁNDAR
2010	AGOSTO	250	896	1224,33
2011	AGOSTO	544	896	1224,33
2012	AGOSTO	3074	896	1224,33
2013	AGOSTO	392	896	1224,33
2014	AGOSTO	220	896	1224,33

Cuadro 9, Media y Desviación Estándar para agosto, (Fuente: Elaboración Propia)

Después se obtienen las siguientes variables:

X1= el valor absoluto de la cantidad vendida de cada mes restado con la media.

X2= la desviación estándar multiplicado por la constante 1.23

El valor atípico es aquel que cumpla con la condición $X1 > X2$, tal como se muestra en el Cuadro 10.

AÑO	MES	CANT_VENDID	MEDIA	DESVIACION ESTÁNDAR	X1	X2	VALOR ATÍPICO
2010	AGOSTO	250	896	1224,33	646	1616,11	FALSO
2011	AGOSTO	544	896	1224,33	352	1616,11	FALSO
2012	AGOSTO	3074	896	1224,33	2178	1616,11	VERDADERO
2013	AGOSTO	392	896	1224,33	504	1616,11	FALSO
2014	AGOSTO	220	896	1224,33	676	1616,11	FALSO

Cuadro 10, Valores Atípicos para el mes de Agosto, (Fuente: Elaboración Propia)

Como se puede observar en el Cuadro 10 las ventas del mes de agosto del año 2012, representa un valor atípico ya que se desvía mucho del resto de la muestra, por consiguiente debe ser eliminado para posteriormente ser interpolado.

- **Interpolación de Datos Faltantes**

Los datos que se detectaron como valores atípicos son eliminados y son reemplazados por datos interpolados para mantener la integridad y los cambios de tendencia de las series de tiempo. En esta tesis los datos faltantes se completaron mediante el uso de la media de los dos últimos datos en el mismo periodo.

3.2.1.5.2. Multivariate Adaptive Regression Splines (Splines de regresión adaptativa multivariante)

- **Variable Objetivo**

Para generar el modelo de predicción usando MARS lo primero que se hizo fue seleccionar la variable objetivo la cual es la variable que se quiere predecir, tal y como se muestra en el cuadro 11.

Entrada	Descripción	Código
Y1	Cantidad Vendida Mensualmente	CANT_VENDIDA

Cuadro 11, Variable Objetivo, (Fuente: Elaboración Propia)

- **Variables de Predicción**

Luego se seleccionó las 7 variables de predicción representados en el cuadro 12, que fueron elegidos sobre una base de juicio subjetivo y por referencia a las variables de mercado financiero más comúnmente utilizadas, las variables de predicción son de mucha importancia ya que en función de estas MARS generara el modelo de predicción.

Entrada	Descripción	Código
X1	Promedio Móvil de 3 meses	MA3
X2	Promedio Móvil de 6 meses	MA6
X3	Monto de ventas del mes anterior	T-1
X4	Monto de ventas de hace 2 meses	T-2
X5	Monto de ventas de hace 3 meses	T-3
X6	Indicador de fuerza relativa de 3 meses	RSI3
X7	Indicador de fuerza relativa de 6 meses	RSI6

Cuadro 12, Variables de Predicción para el Framework, (Fuente: Elaboración Propia)

- **Construcción de un Modelo de Predicción**

Para la construcción del modelo de predicción usando MARS se utilizó “Salford Predictive Modeler 8.0”, el cual es un software de alta precisión y análisis rápido para la creación de modelos de predicción, descriptivos y analíticos de bases de datos de cualquier tamaño, complejidad u organización.

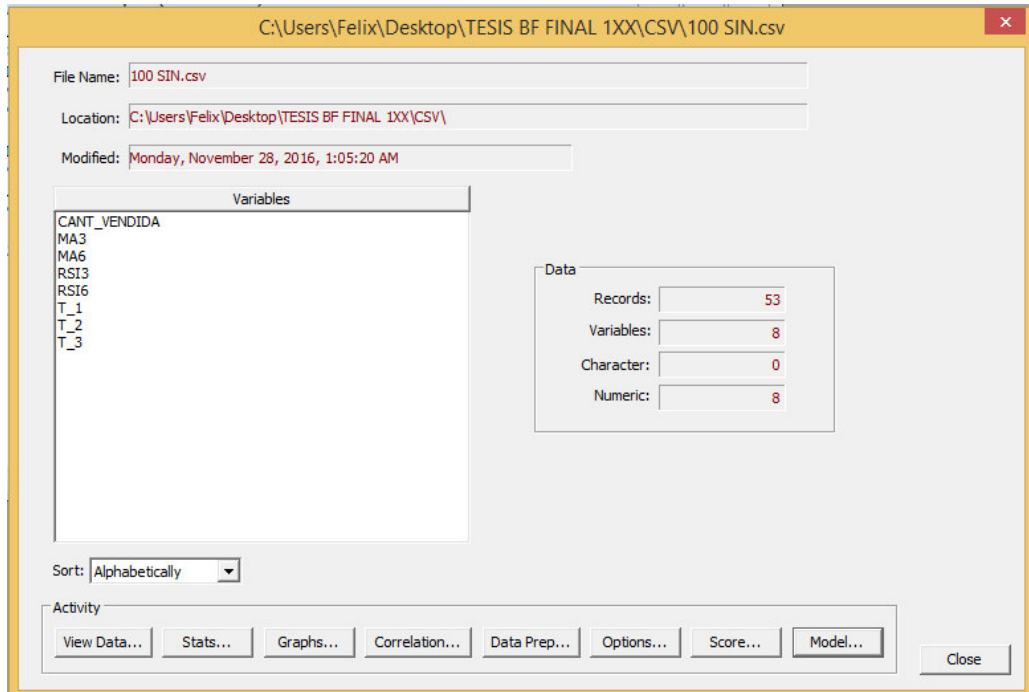
Se siguió los siguientes pasos para crear el modelo de predicción de cada producto seleccionado para esta tesis.

- Seleccionar el archivo csv con las ventas mensuales y las variables de predicción descritas en los cuadros 11 y 12, tal y como se muestra en el Gráfico 13.

CANT_VENDIDA	MA3	MA6	T-1	T-2	T-3	RSI3	RSI6	
412	305,50	261,75	250	212		186	76,74	78,57
254,5	294,83	277,42	412	250		212	42,45	56,24
218	232,17	287,25	254,5	412		250	36,74	52,12
224	249,33	256,42	218	254,5		412	38,77	52,80
306	279,67	250,00	224	218		254,5	63,07	61,74
309	280,67	265,67	306	224		218	63,85	62,06
227	250,67	262,00	309	306		224	34,08	48,85
216	251,67	250,33	227	309		306	31,15	47,23
312	243,33	293,83	216	227		309	67,57	60,82
202	250,00	298,50	312	216		227	35,39	44,91
236	336,00	326,50	202	312		216	47,08	49,79
570	353,67	324,50	236	202		312	85,56	75,41
255	403,00	324,83	570	236		202	42,18	47,80
384	313,00	357,50	255	570		236	55,91	55,76

Grafico 13, Archivo csv con las ventas mensuales y las variables de Predicción de un determinado producto, (Fuente: Elaboración Propia)

- ii. Luego se seleccionó la opción “Model” tal como se muestra en el Grafico 14.



**Grafico 14, Pantalla de Ingreso del historial de ventas,
(Fuente: Elaboración Propia)**

- iii. Luego en la opción Analysis Engine se seleccionó “MARS Regression Splines”, después se seleccionó la variable objetivo y las variables de predicción los cuales eran atributos en el CSV seleccionado en el primer paso, tal y como se muestra en el Grafico 15.

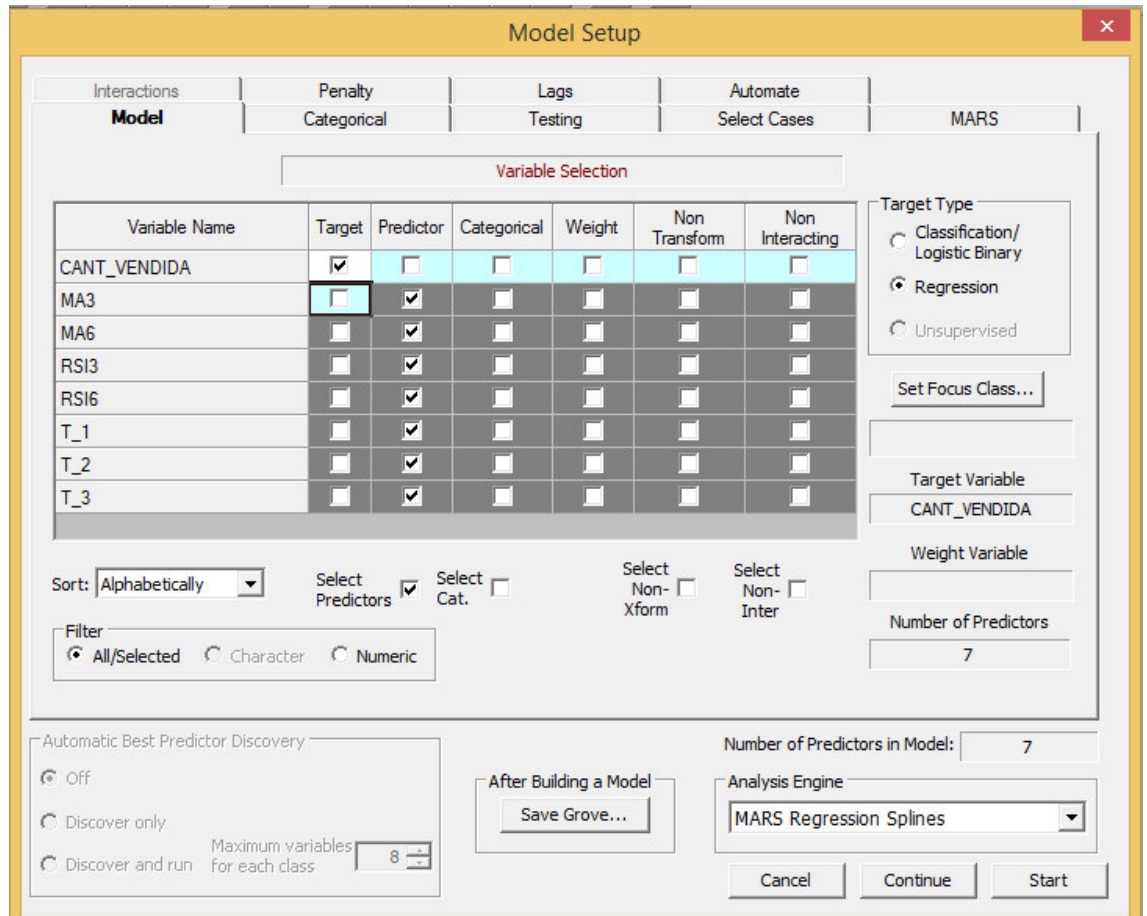


Gráfico 15, Pantalla para la selección de la metodología y las variables, (Fuente: Elaboración Propia)

- iv. A continuación se seleccionó la pestaña “Testing”, luego la opción “V-fold cross-validation” y se colocó el valor de 10, esto para indicar la cantidad de datos que se usaran para el cálculo de las medidas de error, por último se seleccionó la opción “Start” y el software “Salford Predictive Modeler 8.0” creó automáticamente el modelo de predicción y las

medidas de error, tal como se muestra en el Grafico 16.

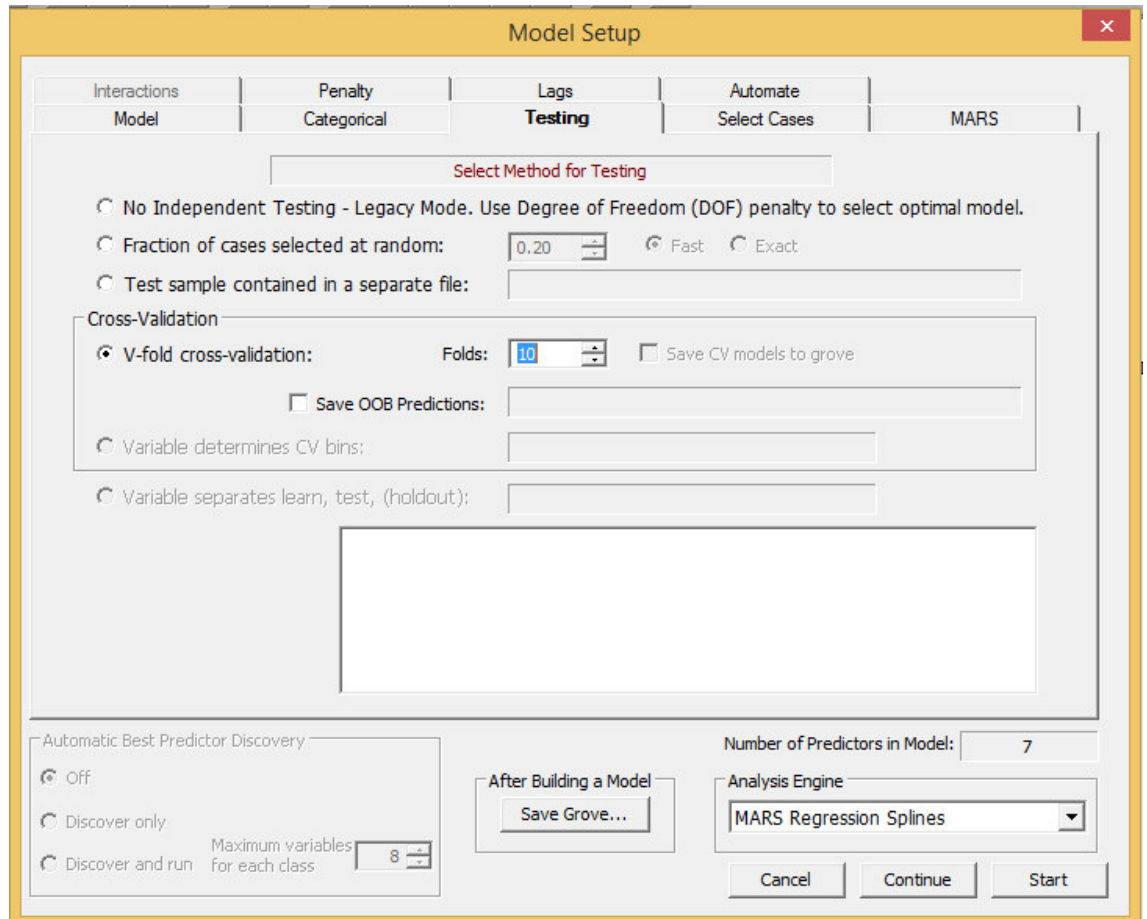


Grafico 16, Pantalla de Testing, (Fuente: Elaboración Propia)

3.3. Aporte Práctico

3.3.1. Introducción

El presente trabajo se traduce en el desarrollo de un Framework para la predicción de ventas, la cual tendrá como finalidad la estimación de las ventas futuras de una PYME, a partir de sus volúmenes de ventas históricas.

Para la generación de los modelos de predicción se usó de entrada el historial de ventas de una empresa dedicada a la venta de autopartes llamada “Franco”, luego para comprobar la eficiencia del Framework propuesto se creó un cuadro comparativo con las medidas de error “RMSE”, “MAD” y “MAPE” obtenidas con la creación de los modelos de predicción usando el Framework propuesto y las medidas de error de los modelos de predicción usando únicamente MARS sin Pre-Procesamiento. Además se usaron los modelos de predicción para predecir las ventas de los meses de enero a mayo del 2014 y compararlos con las ventas reales en dichos meses.

3.3.2. Experimentos Realizados

Para realizar los experimentos se usaron las ventas mensuales de los productos “Resorte Toy Corolla” y “Resorte Toy Corolla Gasol”, estos datos se obtuvieron de una empresa dedicada a la venta de auto-partes llamada “Franco”. Se obtuvieron las ventas mensuales desde noviembre del 2009 hasta mayo del 2014 y se usaron de la forma descrita en el cuadro 13, con el fin de obtener las predicciones de los meses de enero a mayo del 2014 y poder compararlos con las ventas reales de dichos meses.

Meses Procesados	Mes Predicho
Se usaron el historial de ventas desde noviembre del 2009 hasta diciembre del 2013	Enero del 2014
Se usaron el historial de ventas desde noviembre del 2009 hasta enero del 2014	Febrero del 2014
Se usaron el historial de ventas desde noviembre del 2009 hasta febrero del 2014	Marzo del 2014
Se usaron el historial de ventas desde noviembre del 2009 hasta marzo del 2014	Abril del 2014
Se usaron el historial de ventas desde noviembre del 2009 hasta abril del 2014	Mayo del 2014

Cuadro 13, Meses procesados en los experimentos realizados, (Fuente: Elaboración Propia)

Se realizaron dos tipos de experimentos, el primero fue generando los modelos de predicción usando únicamente MARS sin Pre-Procesamiento y la segunda fue usando el Framework propuesto. Después de generados los modelos de predicción se compararon las predicciones mes a mes de acuerdo al cuadro 13, Luego se compararon sus respectivas medidas de error.

3.3.2.1. Modelos de Predicción y valor predicho de cada mes, usando únicamente MARS sin usar Pre-Procesamiento.

A continuación se muestran los modelos de predicción con sus respectivas funciones base y las predicciones de ventas de los meses desde enero a mayo del 2014 por cada producto seleccionado. Para este experimento se usó únicamente MARS sin Pre-Procesamiento.

Producto: Resorte Toy Corolla

Se predijo las ventas del producto “Resorte Toy Corolla” para los meses de enero, febrero, marzo, abril y mayo del año 2014 usando el historial de ventas de acuerdo al cuadro 13, a continuación se muestran las funciones base de cada modelo de Predicción que se obtuvo usando únicamente MARS sin Pre-Procesamiento, tal como se muestra en el cuadro 14.

Mes	Código	Función	Valor
Enero 2014	BF9	$\max(0, \text{RSI6} - 37)$	26.41
Febrero 2014	BF1	$\max(0, \text{RSI3} - 25)$	12.72
	BF2	$\max(0, T_1 - 90)$	392
	BF3	$\max(0, T_3 - 90)$	170
	BF7	$\max(0, \text{MA6} - 329)$	0
Marzo 2014	BF9	$\max(0, \text{RSI6} - 37)$	4.86
Abril 2014	BF1	$\max(0, \text{RSI3} - 25)$	6.90
Mayo 2014	BF1	$\max(0, \text{RSI3} - 25)$	49.42

Cuadro 14, Valor de las Funciones base de los modelos de Predicción correspondientes a los meses de enero a mayo del 2014 usando únicamente MARS sin Pre-Procesamiento, (Fuente: Elaboración Propia)

A continuación se muestra en el cuadro 15, los valores de ventas predichos de cada mes, calculados de los modelos de predicción generados usando únicamente MARS sin Pre-Procesamiento.

Mes	Modelo de Predicción	Predicción	Valor Real	Diferencia
Enero 2014	$184.9 + 8.79551 * BF9$	417.204	466	48.796
Febrero 2014	$70.9276 + 7.19253 * BF1 + 0.118365 * BF2 + 0.115439 * BF3 + 1.32745 * BF7$	228.43	214	14.43
Marzo 2014	$183.943 + 8.82988 * BF9$	226.819	118	108.43
Abril 2014	$137.927 + 7.06432 * BF1$	186.643	263	76.357
Mayo 2014	$138.457 + 7.02723 * BF1$	485.721	200	285.721

Cuadro 15, Tabla Resumen de los valores Predichos calculados a partir de los modelos de Predicción usando únicamente MARS sin Pre-Procesamiento, (Fuente: Elaboración Propia)

Producto: Resorte Toy Corolla Gasol

Se predijo las ventas del producto “Resorte Toy Corolla Gasol” para los meses de enero, febrero, marzo, abril y mayo del año 2014 usando el historial de ventas de acuerdo al cuadro 13, a continuación se muestran las funciones base de cada modelo de Predicción que se obtuvo usando únicamente MARS sin Pre-Procesamiento, tal como se muestra en el cuadro 16.

Mes	Código	Función	Valor
Enero 2014	BF1	$\max(0, \text{RSI6} - 53)$	0
	BF2	$\max(0, 53 - \text{RSI6})$	8.16
	BF3	$\max(0, \text{RSI3} - 56)$	0
	BF6	$\max(0, 202 - T_1)$	0
	BF7	$\max(0, T_3 - 46)$	122
	BF8	$\max(0, \text{RSI6} - 50)$	0
Febrero 2014	BF6	$\max(0, 174 - T_1)$	0
	BF7	$\max(0, \text{RSI3} - 56)$	17.19
	BF8	$\max(0, 56 - \text{RSI3})$	0
	BF10	$\max(0, T_3 - 46)$	252
Marzo 2014	BF1	$\max(0, \text{RSI6} - 55)$	0
	BF2	$\max(0, 55 - \text{RSI6})$	7.83
	BF3	$\max(0, \text{RSI3} - 53)$	0
	BF4	$\max(0, 53 - \text{RSI3})$	10.83
	BF5	$\max(0, T_1 - 196)$	182
	BF6	$\max(0, 196 - T_1)$	0
	BF7	$\max(0, T_3 - 46)$	190
	BF8	$\max(0, \text{RSI6} - 59)$	0
	BF10	$\max(0, \text{RSI6} - 51)$	0
	BF12	$\max(0, \text{RSI3} - 48)$	0
	BF14	$\max(0, \text{MA3} - 227)$	32.33
	BF15	$\max(0, 227 - \text{MA3})$	0
Abril 2014	BF1	$\max(0, \text{RSI6} - 55)$	0
	BF2	$\max(0, 55 - \text{RSI6})$	8.49
	BF3	$\max(0, \text{RSI3} - 52)$	0
	BF6	$\max(0, 190 - T_1)$	0
	BF7	$\max(0, T_3 - 46)$	128
	BF8	$\max(0, \text{RSI6} - 59)$	0

Mayo 2014	BF2	$\max(0, 55 - \text{RSI6})$	11.42
	BF3	$\max(0, \text{RSI3} - 52)$	0
	BF6	$\max(0, 196 - T_1)$	0

Cuadro 16, Valor de las Funciones base de los modelos de Predicción correspondientes a los meses de enero a mayo del 2014 usando únicamente MARS sin Pre-Procesamiento, (Fuente: Elaboración Propia)

A continuación se muestra en el cuadro 17, los valores de ventas predichos de cada mes, calculados de los modelos de predicción generados usando únicamente MARS sin Pre-Procesamiento.

Mes	Modelo de Predicción	Predicción	Valor Real	Diferencia
Enero 2014	209.975 - 6.19869 * BF1 - 6.77769 * BF2 + 5.24894 * BF3 - 0.415476 * BF6 + 0.0945986 * BF7 + 5.55733 * BF8	166.179784	164	2.17978404
Febrero 2014	222.254 - 0.763169 * BF6 + 4.68832 * BF7 - 4.15175 * BF8 + 0.160931 * BF10	343.38	188	155.3846
Marzo 2014	205.742 - 26.4003 * BF1 - 1.72679 * BF2 + 13.2488 * BF3 - 3.3175 * BF4 + 0.0154463 * BF5 - 0.465771 * BF6 + 0.175206 * BF7 + 13.2388 * BF8 + 11.9122 * BF10 - 6.74543 * BF12 + 0.514328 * BF14 + 0.00693996 * BF15	209.0104205	52	157.0104205
Abril 2014	225.817 - 8.9413 * BF1 - 7.15685 * BF2 + 5.48843 * BF3 - 0.506621 * BF6 + 0.126515 * BF7 + 8.67956 * BF8	181.2442	160	21.24
Mayo 2014	241.283 - 6.80067 * BF2 + 4.22304 * BF3 - 0.532618 * BF6	163.6	230	66.40157449

Cuadro 17, Tabla Resumen de los valores Predichos calculados a partir de los modelos de Predicción usando únicamente MARS sin Pre-Procesamiento, (Fuente: Elaboración Propia)

3.3.2.2. Modelos de Predicción y valor predicho de cada mes, usando el Framework Propuesto (Combinación de un Componente de Pre-Procesamiento y MARS).

A continuación se muestran los modelos de predicción con sus respectivas funciones base y las predicciones de ventas de los meses desde enero a mayo del 2014 por cada producto seleccionado. Para este experimento se usó el Framework Propuesto (combinación de un componente de Pre-Procesamiento y MARS).

Producto: Resorte Toy Corolla

Se predijo las ventas del producto “Resorte Toy Corolla” para los meses de enero, febrero, marzo, abril y mayo del año 2014 usando el historial de ventas de acuerdo al cuadro 13, a continuación se muestran las funciones base de cada modelo de Predicción que se obtuvo usando el Framework Propuesto (combinación de un componente de Pre-Procesamiento y MARS), tal como se muestra en el cuadro 18.

Mes	Código	Función	Valor
Enero 2014	BF1	$\max(0, \text{RSI3} - 28)$	14.90
	BF2	$\max(0, \text{MA3} - 209)$	44.17
	BF3	$\max(0, \text{T}_3 - 90)$	115
	BF4	$\max(0, \text{RSI6} - 38)$	8.59
	BF5	$\max(0, \text{T}_2 - 90)$	170
Febrero 2014	BF1	$\max(0, \text{RSI3} - 28)$	15.24
Marzo 2014	BF1	$\max(0, \text{RSI3} - 26)$	0.40
	BF11	$\max(0, \text{T}_3 - 236)$	30
	BF15	$\max(0, \text{MA6} - 235)$	12.5
Abril 2014	BF3	$\max(0, \text{MA3} - 190)$	69.17
	BF4	$\max(0, \text{T}_3 - 90)$	156.5
	BF13	$\max(0, \text{RSI3} - 40)$	0
Mayo 2014	BF7	$\max(0, \text{RSI3} - 40)$	42,30

Cuadro 18, Valor de las Funciones base de los modelos de predicción correspondientes a los meses de enero a mayo del 2014 usando el Framework Propuesto, (Fuente: Elaboración Propia)

A continuación se muestra en el cuadro 19, los valores de ventas predichos de cada mes, calculados de los modelos de predicción generados usando el Framework Propuesto (Componente de Pre-Procesamiento y MARS).

Mes	Modelo de Predicción	Predicción	Valor Real	Diferencia
Enero 2014	$79.4045 + 3.80544 * BF1 + 0.473127 * BF2 + 0.175851 * BF3 + 2.38508 * BF4 + 0.0998504 * BF5$	214.68	466	251.32
Febrero 2014	$163.817 + 5.45306 * BF1$	246.94	214	32.94
Marzo 2014	$134.468 + 5.07096 * BF1 + 0.168088 * BF11 + 0.436369 * BF15$	146.98	118	28.98
Abril 2014	$140.918 + 0.537167 * BF3 + 0.16236 * BF4 + 5.4787 * BF13$	203,48	263	59.52
Mayo 2014	$207.865 + 5.91243 * BF7$	457.96	200	257.96

Cuadro 19, Tabla Resumen de los valores Predichos calculados a partir de los modelos de Predicción usando el Framework Propuesto, (Fuente: Elaboración Propia)

Producto: Resorte Toy Corolla Gasol

Se predijo las ventas del producto “Resorte Toy Corolla Gasol” para los meses de enero, febrero, marzo, abril y mayo del año 2014 usando el historial de ventas de acuerdo al cuadro 13, a continuación se muestran en el cuadro 20 las funciones base de cada modelo de Predicción que se obtuvo usando el Framework Propuesto (Componente de Pre-Procesamiento y MARS).

Mes	Código	Función	Valor
Enero 2014	BF4	$\max(0, \text{RSI3} - 19)$	17.20
Febrero 2014	BF2	$\max(0, T_3 - 68)$	230
	BF3	$\max(0, \text{MA6} - 141)$	67.75
	BF4	$\max(0, \text{RSI3} - 19)$	46.88
Marzo 2014	BF2	$\max(0, T_3 - 68)$	168
	BF3	$\max(0, \text{MA6} - 135)$	65.75
	BF6	$\max(0, \text{RSI3} - 19)$	24.99
Abril 2014	BF1	$\max(0, \text{RSI6} - 37)$	9.96
	BF2	$\max(0, \text{MA6} - 203)$	0
	BF3	$\max(0, 203 - \text{MA6})$	22.58
	BF4	$\max(0, T_3 - 64)$	110
	BF5	$\max(0, \text{RSI3} - 19)$	22.76
	BF6	$\max(0, \text{MA3} - 203)$	0
	BF7	$\max(0, 203 - \text{MA3})$	15
	BF8	$\max(0, T_2 - 64)$	214.5
	BF9	$\max(0, \text{RSI3} - 55)$	0
	BF11	$\max(0, \text{MA3} - 184)$	4

	BF13	$\max(0, \text{RSI3} - 47)$	0
	BF15	$\max(0, T_1 - 64)$	140
Mayo 2014	BF2	$\max(0, T_3 - 64)$	214.5
	BF3	$\max(0, \text{MA6} - 203)$	0
	BF4	$\max(0, 203 - \text{MA6})$	42.33
	BF5	$\max(0, \text{RSI3} - 19)$	13
	BF6	$\max(0, \text{MA3} - 203)$	0
	BF8	$\max(0, T_2 - 64)$	140
	BF9	$\max(0, \text{RSI3} - 54)$	0

Cuadro 20, Valor de las Funciones base de los modelos de Predicción correspondientes a los meses de enero a mayo del 2014 usando el Framework Propuesto, (Fuente: Elaboración Propia)

A continuación se muestra en el cuadro 21 los valores de ventas predichos de cada mes, calculados de los modelos de predicción generados usando el Framework Propuesto (Componente de Pre-Procesamiento y MARS).

Mes	Modelo de Predicción	Predicción	Valor Real	Diferencia
Enero 2014	$119.159 + 2.52876 * BF4$	162.64	164	1.36
Febrero 2014	$45.792 + 0.336945 * BF2 + 0.63049 * BF3 + 2.32674 * BF4$	275.08	188	87.08
Marzo 2014	$46.1596 + 0.358659 * BF2 + 0.602177 * BF3 + 2.16515 * BF6$	200.12	52	148.12
Abril 2014	$96.6895 - 0.0234918 * BF1 - 1.98421 * BF2 - 0.892689 * BF3 + 0.232554 * BF4 + 2.4074 * BF5 + 3.15895 * BF6 - 0.207116 * BF7 + 0.147052 * BF8 - 4.09028 * BF9 - 1.39037 * BF11 + 2.69508 * BF13 - 0.0140275 * BF15$	177,57	160	17.57
Mayo 2014	$68.6247 + 0.251716 * BF2 - 1.93563 * BF3 - 0.817391 * BF4 + 2.89204 * BF5 + 1.32281 * BF6 + 0.167467 * BF8 - 1.27287 * BF9$	149,06	230	80.94

Cuadro 21, Tabla Resumen de los valores Predichos calculados a partir de los modelos de Predicción usando el Framework Propuesto, (Fuente: Elaboración Propia)

3.3.2.3. Cuadro Comparativo de las Predicciones de Ventas

En esta sección se elaboró un cuadro comparativo de las predicciones obtenidas a partir de los modelos de Predicción generados usando MARS sin Pre-Procesamiento y el Framework Propuesto (Componente de Pre-Procesamiento y MARS). A continuación en los cuadros 22 y 23 se muestran los resultados de predicción de los productos “Resorte Toy Corolla” y “Resorte Toy Corolla Gasol” respectivamente.

Mes	Predicción sin Pre-Procesamiento	Predicción con Pre-Procesamiento	Venta Real	Dif (Sin Pre-Proc)	Dif (Con Pre-Proc)
Enero 2014	417.204	214.68	466	48.796	251.32
Febrero 2014	228.43	246.94	214	14.43	32.94
Marzo 2014	226.819	146.98	118	108.819	28.98
Abril 2014	186.643	203.48	263	76.357	59.52
Mayo 2014	485.721	457.96	200	285.721	257.96

Cuadro 22, Tabla Comparativa de las Predicciones de ventas del Producto “Resorte Toy Corolla”, (Fuente: Elaboración Propia)

Mes	Predicción sin Pre-Procesamiento	Predicción con Pre-Procesamiento	Venta Real	Dif (Sin Pre-Proc)	Dif (Con Pre-Proc)
Enero 2014	166.179	162.64	164	2.179	1.36
Febrero 2014	343.38	275.08	188	155.38	87.08
Marzo 2014	209.01	200.12	52	157.01	148.12
Abril 2014	181.244	177.57	160	21.244	17.57
Mayo 2014	163.6	149.06	230	66.4	80.94

Cuadro 23, Tabla Comparativa de las Predicciones de ventas del Producto "Resorte Toy Corolla Gasol", (Fuente: Elaboración Propia)

3.3.2.4. Cuadro Comparativo de las Medidas de Error

A continuación se muestran los cuadros comparativos 24 y 25 de los productos “Resorte Toy Corolla” y “Resorte Toy Corolla Gasol” respectivamente, con los valores de las medidas de error RMSE, MAD y MAPE obtenidos en la generación de los modelos de Predicción usando MARS sin Pre-Procesamiento y el Framework Propuesto (Componente de Pre-Procesamiento y MARS).

Recordar que las medidas de error solo son una forma de medir el rendimiento de los modelos de Predicción en el cual a menor valor, el modelo de Predicción tiene menos error y tiende a ser más preciso, pero se debe tomar en cuenta que dichas medidas de error se generaron con solo un porcentaje del historial de ventas de acuerdo al gráfico 16, el resto del historial de ventas fue usado por MARS para generar los modelos de Predicción, por lo que en algunas ocasiones a pesar de no haber obtenido los mejores resultados en las medidas de error, se puede obtener predicciones de ventas más precisas, por ejemplo: para el producto “Resorte Toy Corolla” en los meses de abril y mayo no se obtuvo buenos resultado en ninguna de las medidas de error para el Framework Propuesto tal y como se muestra en el cuadro 24, sin embargo cuando usamos sus respectivos modelos de Predicción para calcular las predicciones de ventas en dichos meses, pudimos obtener resultados más precisos tal y como se muestra en el cuadro 22.

MES	EXPERIMENTO	RMSE	MAD	MAPE
Enero 2014	MARS Sin Pre-Procesamiento	49.9	38.258	0.128
	Framework Propuesto	43.67	33.23	0.136
Febrero 2014	MARS Sin Pre-Procesamiento	46.9	35.32	0,12
	Framework Propuesto	36,53	27,58	0,11
Marzo 2014	MARS Sin Pre-Procesamiento	50,65	36,87	0,12
	Framework Propuesto	45,58	33,41	0,13
Abril 2014	MARS Sin Pre-Procesamiento	45,09	33,88	0,11
	Framework Propuesto	56,15	40,09	0,16
Mayo 2014	MARS Sin Pre-Procesamiento	45,84	34,11	0,12
	Framework Propuesto	49,08	34,45	0,14

Cuadro 24, Cuadro Comparativo de las medidas de error del producto “Resorte Toy Corolla”, (Fuente: Elaboración Propia)

MES	COMPONENTE	RMSE	MAD	MAPE
Enero 2014	MARS Sin Pre- Procesamiento	32,28	21,79	0,13
	Framework Propuesto	30,48	26,06	0,16
Febrero 2014	MARS Sin Pre- Procesamiento	35,81	24,95	0,15
	Framework Propuesto	28,59	23,38	0,14
Marzo 2014	MARS Sin Pre- Procesamiento	40.52	28.47	0.15
	Framework Propuesto	26.17	21.52	0.12
Abril 2014	MARS Sin Pre- Procesamiento	32.28	22.16	0.11
	Framework Propuesto	27.52	22.88	0.14
Mayo 2014	MARS Sin Pre- Procesamiento	31.88	21.64	0.14
	Framework Propuesto	26.88	20.79	0.12

Cuadro 25, Cuadro Comparativo de las medidas de error del producto “Resorte Toy Corolla Gasol”, (Fuente: Elaboración Propia)

3.3.2.5. Discusión de los resultados

Luego de las pruebas realizadas, los resultados indican que el Framework propuesto (Componente de Pre-Procesamiento y MARS) genera predicciones de ventas con mayor precisión, que usando solo MARS sin Pre-Procesamiento.

Las pruebas se realizaron tomando en cuenta los productos “Resorte Toy Corolla” y “Resorte Toy Corolla Gasol”, para los cuales se realizaron dos tipos de experimentos el primero fue generando los modelos de predicción usando únicamente MARS sin Pre-Procesamiento y la segunda fue usando el Framework propuesto. Después de generados los modelos de predicción se compararon las predicciones mes a mes de acuerdo a los cuadros 22 y 23, luego se compararon sus respectivas medidas de error en los cuadros 24 y 25.

Se calculó las predicciones de ventas para los meses de enero, febrero, marzo, abril y mayo del año 2014 tanto para el producto “Resorte Toy Corolla” como para el producto “Resorte Toy Corolla Gasol”. Para el producto “Resorte Toy Corolla” se obtuvieron resultados más precisos en los meses de marzo, abril y mayo, con lo cual se obtendría una predicción de ventas más precisa en el 60% de los meses predichos tal y como se mostró en el cuadro 22. Para el producto “Resorte Toy Corolla Gasol” se obtuvieron resultados más precisos en los meses de enero, febrero, marzo y abril, con lo cual se obtuvo una predicción de ventas más precisas en el 80% de los meses predichos, tal y como se mostró en el cuadro 23.

En la comparación de las medidas de error RMSE, MAD y MAPE, en la medida de error RMSE se obtuvieron resultados más precisos en el 60% de los meses predichos del producto “Resorte Toy Corolla” y el 100% de meses predichos del producto “Resorte Toy Corolla Gasol”. Para la medida de error MAD se obtuvieron resultados más precisos en el 60% de los meses predichos del producto “Resorte Toy Corolla” y 60% de los meses predichos del producto “Resorte Toy Corolla Gasol”. Para la medida de error MAPE se obtuvieron resultados más precisos en el 20% de los meses predichos del producto “Resorte Toy Corolla” y 60% de los meses predichos del producto “Resorte Toy Corolla Gasol”, tal y como se mostraron en los cuadros 24 y 25.

4. CAPITULO 5: CONCLUSIONES Y RECOMENDACIONES

En el presente capítulo se describe las conclusiones alcanzadas y finalmente se menciona algunas recomendaciones para trabajos futuros.

4.1. Conclusiones

Después de realizar la investigación de la presente Tesis llegamos a la conclusión que con el Framework propuesto (Componente de Pre-Procesamiento y MARS) se pueden obtener mejores resultados en el 70% (60% de los meses predichos del producto “Resorte Toy Corolla” y 80% del producto “Resorte Toy Corolla Gasol” de acuerdo a los cuadros 22 y 23 respectivamente) de los meses predichos, esto debido a que el Framework propuesto tiene un componente de Pre-Procesamiento que permite eliminar e interpolar los datos atípicos que podría existir en el historial de ventas.

De acuerdo a los cuadros 24 y 25 las medidas de error RMSE (en el 80% de los meses predichos) y MAD (en el 60% de los meses predichos) demuestran que las predicciones de ventas calculadas con el Framework propuesto generan resultados más precisos, a diferencia de la medida de error MAPE en el que solo se obtuvo un resultado más preciso en el 40% de los meses predichos.

Con los resultados obtenidos en las pruebas realizadas se puede llegar a la conclusión que se logró el objetivo general de la tesis ya que las PYMES al contar con una predicción de ventas más

precisa y confiable reduciendo tiempo y costos pueden obtener mayores márgenes de ganancia por ejemplo:

- Reduciendo la pérdida de insumos o productos perecibles.
- Reduciendo los costos de almacenaje, al tener un inventario muy grande.
- Aumento de la satisfacción del cliente al tener siempre disponible los insumos o productos que necesitan.

Por otro lado no es necesario que el usuario sea un experto ni que tenga conocimientos muy avanzados con respecto a las técnicas o métodos de Predicción, porque cuando usamos el software “Salford Predictive Modeler 8.0”, esta misma selecciona las variables de Predicción más importantes y crea el modelo de predicción con sus respectivas funciones base.

4.2. Recomendaciones

Como trabajos futuros a partir de la presente investigación se puede considerar lo siguiente:

Adicionar un Proceso de Ajuste Heurístico para refinar los resultados obtenidos por el Framework propuesto en esta tesis, para obtener mejores resultados.

Realizar pruebas para medir el rendimiento cuando usamos una metodología de red neuronal en lugar de MARS para construir el modelo de predicción de ventas.

ANEXOS

ANEXO N° 01: MATRIZ DE CONSISTENCIA

PROBLEMA	OBJETIVOS	VARIABLES E INDICADORES			
<p>Problema General: ¿De qué manera las PYMES del Perú pueden predecir sus ventas de forma precisa y confiable reduciendo el tiempo y costos para obtener un mayor margen de ganancias?</p> <p>Problemas Específicos:</p> <ul style="list-style-type: none"> - ¿Cómo influye la precisión de las predicciones de ventas en la productividad de las PYMES del Perú? - ¿Cómo influye las medidas de error en la confiabilidad de las 	<p>Objetivo General: Desarrollar un Framework utilizando un componente de Pre-procesamiento y MARS para obtener una predicción de ventas precisa y confiable reduciendo tiempo y costos. Para que las PYMES puedan obtener mayores márgenes de ganancias.</p> <p>Objetivos Específicos:</p> <ul style="list-style-type: none"> - Determinar la influencia de la precisión de las predicciones de 	VAR. INDEPENDIENTE: (X): METODOLOGIA DE PREDICCION			
		Dimensiones	Indicadores	Ítems	Índices
		Precisión	Porcentaje de ganancia	2	<ul style="list-style-type: none"> • Siempre(S) • A veces (A) • Nunca (N)
		Medidas de Error	Porcentaje de Error	1	
		VAR. DEPENDIENTE: (Y): RENDIMIENTO DE PREDICCION			
		Dimensiones	Indicadores	Ítems	Índices
		Productividad	Aumento en el porcentaje de ganancias	Comparativa de los productos comprados y predichos mes a mes.	<ul style="list-style-type: none"> • Excelente • Bueno • Normal • Insuficiente • Muy deficiente

<p>predicciones de Ventas de las PYMES del Perú?</p> <ul style="list-style-type: none"> - ¿Cómo influye el uso de los métodos de predicción con deficiente precisión en los márgenes de ganancia de las PYMES del Perú? - ¿Cómo influye el uso de los métodos de predicción de ventas muy complejos en su implementación en las PYMES del Perú? 	<p>ventas en la productividad de las PYMES del Perú.</p> <ul style="list-style-type: none"> - Determinar la influencia de las medidas de error en la confiabilidad de las predicciones de ventas de las PYMES del Perú. - Obtener el historial de ventas de una PYME del Perú, para realizar pruebas y demostrar la precisión del Framework que se desarrollara. - Determinar la mejor técnica o metodología para obtener Predicciones de Ventas Precisas. - Desarrollar el Framework de forma entendible y poco complejo para facilitar su implementación en las PYMES del Perú. 	<p>Confiabilidad</p>	<p>Disminución de las medidas de error de la predicción.</p>	<p>Comparativa de las medidas de error usando el Framework propuesto y sin usarlo.</p>	
---	---	----------------------	--	--	--

Instrucciones:

A continuación, encontraras una serie de preguntas que se refieren a la realidad en la que se encuentran la gran mayoría de PYMES en el Perú respecto a sus Predicciones de Ventas.

- Si lo que se dice en la pregunta es afirmativo con respecto a la realidad de las PYMES del Perú, rodea con un círculo el **SI**.
- Si lo que se dice en la pregunta es falso o incorrecto, rodea con un círculo el **NO**.
- Si **No sabes que contestar** a la pregunta, rodea con un círculo el signo?

Inventario de Preguntas para determinar la Realidad de las PYMES del Perú respecto a las Predicciones de Ventas.

1. ¿Actualmente las PYMES del Perú pueden costear la implementación de un software de Predicción de Ventas que le den resultados Eficientes?
2. ¿Los softwares de Predicción de Ventas que usan las PYMES del Perú requieren mucha capacitación de los usuarios?
3. ¿Los softwares de Predicción de Ventas que usan las PYMES del Perú obtienen resultados precisos y confiables en un tiempo aceptable?
4. ¿Las PYMES del Perú tienen un proceso para evaluar la precisión de sus predicciones de ventas?

ANEXO N° 02: GLOSARIO

Software: Programa o conjunto de programas interrelacionados que permiten al usuario interactuar con los dispositivos sin perderse en la complejidad técnica del equipo.

MARS: Multivariate Adaptive Regression Splines es una metodología de regresión no lineal que incorpora interacciones entre variables, es decir es una herramienta flexible que puede automatizar la construcción de modelos de predicción.

Componente de Pre-Procesamiento: Este es un componente del Framework desarrollado en esta tesis que se usa para disminuir los valores atípicos de la muestra de datos

Framework: es un esquema, un esqueleto ó un patrón para el desarrollo y/o la implementación de una aplicación.

Predicción de ventas: es una estimación de las ventas de un bien o servicio en un determinado periodo en el futuro.

Medidas de Error: son indicadores que permiten estimar la calidad de los métodos de predicción.

Redes Neuronales: El objetivo de las Redes Neuronales es la de crear modelos artificiales que solucionen problemas difíciles de resolver, mediante técnicas algorítmicas convencionales. Las Redes Neuronales están inspiradas en el comportamiento del cerebro humano (en especial a sus neuronas y conexiones).

REFERENCIAS BIBLIOGRAFICAS

[1] Nojek, Predicción de Ventas: Comparación de Predicción basada en Redes Neuronales versus Método Estadístico, 2003, pp 1-2

[2] Yalta Maicelo, Juan José, sistema de información para el Predicción de ventas en pymes del sector comercial en el Perú utilizando redes neuronales artificiales, 2013, pp1-3

[3] Chi-Jie Lu, Tian-Shyug Lee y Chia-Mei Lian, Sales forecasting for computer wholesalers: A comparison of multivariate adaptive regression splines and artificial neural networks, 2012, pp1-2

[4] W.K. Wong, Z.X.Guo, A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm, 2010, pp1-3

[5] Patrícia Ramos, Nicolau Santos, Rui Rebelo, Performance of state space and ARIMA models for consumer retail sales forecasting, 2015, pp1-12

[6] J.H. Friedman, Multivariate adaptive regression splines, The Annals of Statistics 19(1), 1991, pp1-141

[7] Mentzer J. T., Moon M. A., Sales Forecasting Management: A Demand Management Approach, 2005, 368p.

[8] Maxim Vladimirovich Shcherbakov, Adriaan Brebels, Nataliya Lvovna Shcherbakova, Anton Pavlovich Tyukov, Timur Alexandrovich Janovsky, Valeriy Anatol'evich Kamaev, A Survey of Forecast Error Measures, 2013, pp1-2

[9] Yu, X., Liu, Y., Huang, J. X., An, A, Mining online reviews for predicting sales performance: A case study in the movie domain. IEEE Transactions on Knowledge and Data Engineering, 2012, pp720–734.

[10] Ryu, K., Han, H, Influence of the quality of food, service, and physical environment on customer satisfaction and behavioral intention in quick-casual restaurants: Moderating role of perceived price. Journal of Hospitality and Tourism Research, 2010, pp310–329.

[11] Wysocki, P. D, Private information, earnings announcements and trading volume, or stock chat on the internet: A public debate about private information. Working paper. University of Michigan Business School, 2000, pp, 215–227.

[12] Bass, F. M, A new product growth for model consumer durables. Management Science, 1969, pp215–227.

[13] Marshall, P., Dockendorff, M., Ibáñez, S, A forecasting system for movie attendance. Journal of Business Research, 2013, pp1800–1806.

[14] Mentzer, J. T., Bienstock, C. C. Sales forecasting management: Understanding the techniques, systems, and management of the sales forecasting process. Sage Publications, 1998.

[15] Herr, P. M., Kardes, F. R., & Kim, J. Effects of word-of-mouth and product-attribute information on persuasion: An accessibility-diagnostics perspective. *Journal of Consumer Research*, 1991, pp454–462.

[16] Chevalier, J. A., & Mayzlin, D. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 2006, pp345–354.

[17] Dellarocas, C., Awad, N. F., Zhang, X. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 2007, pp23–45.

[18] Wu, M., Wang, L., Li, M. An approach based on the bass model for analyzing the effects of feature fatigue on customer equity. *Computational & Mathematical Organization Theory*, 2015, pp69–89.

[19] Wang, F. K., Chang, K. K., Hsiao, Y. Y. Implementing a diffusion model optimized by a hybrid evolutionary algorithm to forecast notebook shipments. *Applied Soft Computing*, 2013, pp1147–1151.

[20] Hui-Chih Hung, Yu-Chih Chiu, Huang-Chen Huang, Muh-Cherng Wu, An enhanced application of Lotka–Volterra model to forecast the sales of two competing retail formats, 2017, pp1-10

[21] Zhi-Ping Fan, Yu-Jie Che, Zhen-Yu Chen, Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis, 2017, pp1-11