

Dedicatoria

A Dios por darme siempre sus bendiciones y guiar e iluminar cada paso en mi vida. En memoria de mis queridas Madre Rosa y Madrina Nidia a quienes debo todo. A Vilma mi compañera de vida por su amor y comprensión y familia por brindarme siempre su incondicional apoyo en todo momento. A todos mi familiares que me brindaron su ayuda.

Alberto Garcete

Doy gracias a Dios por haber llegado a esta etapa de mi vida y haber concluído con éxito este objetivo trazado hace tiempo. Gracias a mi familia, mi papá César y mi madre Filomena que siempre me dieron su apoyo y estuvieron para alentarme en todo momento. Agradezco a mi novia Fátima por hacerme el aguante en todo este período de tesis, por su apoyo constante y comprensión.

Raúl Benítez

Agradecimiento

A los tutores:

D.Sc. Diego Pinto Roa y MSc. Aditardo Vázquez, por su orientación y el acompañamiento en la realización de este trabajo, nuestra gratitud y respeto para ambos.

A nuestras familias por la ayuda incondicional en todo momento, estaremos eternamente agradecidos.

A todos los docentes de la Facultad Politécnica quienes transmitieron sus conocimientos con mucha vocación y sabiduría.

A los compañeros de la Facultad Politécnica, por su apoyo y colaboración.

Resumen

Pronosticar ciertos eventos constituye una actividad por la cual el ser humano siempre sintió fascinación y necesidad de realizarlo. En la actualidad, uno de esos eventos se relacionan con las empresas y consiste en pronosticar la demanda de ventas para un período futuro, a su vez representa uno de los más importantes retos con que se enfrenta una organización. Este pronóstico de demanda disminuirá la incertidumbre del Gerente de Compras en el momento de tomar decisiones acerca del volumen de productos a adquirir para la reposición de stock. Esta tesis propone una nueva técnica de pronósticos basada en la integración de herramientas de Business Intelligence y Machine Learning. Los experimentos indican que la técnica propuesta alcanza resultados prometedores y esta nueva técnica puede transformarse en una sólida herramienta de apoyo para la toma de decisiones.

Palabras claves: Pronóstico, Retail, Compras, Business Intelligence, KPI, Machine Learning

Abstract

Predicting certain events is an activity for which the human being always felt fascination and need to do it. Currently, one of these events is related to companies and is to forecast the demand for sales for a future period, which in turn represents one of the most important challenges facing an organization. This demand forecast will reduce the uncertainty of the Purchasing Manager at the time of making decisions about the volume of products to acquire for the replacement of stock. This thesis proposes a new forecasting technique based on the integration of Business Intelligence and Machine Learning tools. The experiments indicate that the proposed technique achieves promising results and this new technique can become a solid support tool for decision making.

Keywords: Forecasting, Retail, Purchasing, Business Intelligence, KPI, Machine Learning

Índice General

Dedicatoria	I
Agradecimientos	II
Resumen	III
Abstract	IV
Índice de Figuras	IX
Índice de Tablas	X
Índice de Algoritmos	XI
Lista de Siglas y Abreviaturas	XIII
1. Introducción	1
1.1. Motivación	2
1.2. Planteamiento del Problema	3
1.3. Objetivos	3
1.3.1. Objetivo General	3
1.3.2. Objetivos Específicos	3
1.4. Organización del Trabajo Final de Grado	4
2. Pronóstico de la Demanda	5
2.1. Administración de compras	5
2.2. Métodos de pronósticos cualitativos	7
2.3. Métodos de pronósticos cuantitativos	8
2.3.1. Pronósticos de series de tiempo	8
2.3.2. Pronósticos causales	13
2.3.3. Precisión de los métodos cuantitativos	14
2.4. Revisión Literaria acerca del pronóstico de la demanda	14
2.5. Resumen	16
3. Business Intelligence	17
3.1. Definición	17
3.2. Objetivos	18
3.3. Componentes de Business Intelligence	18
3.3.1. Fuentes de Información	19
3.3.2. Extracción, Transformación y Carga	20

3.3.3. Datawarehouse	22
3.3.4. OLAP	24
3.3.5. Herramientas de BI	25
3.4. Indicadores Clave de Rendimiento	26
3.5. Aplicaciones de BI	26
3.6. Resumen	29
4. Machine Learning	30
4.1. Definición	33
4.2. Formas de Aprendizaje	34
4.2.1. Aprendizaje supervisado	34
4.2.2. Aprendizaje no supervisado	35
4.2.3. Aprendizaje por refuerzo	35
4.3. Algoritmos de aprendizaje supervisado y no supervisado	36
4.3.1. Algoritmos de regresión	37
4.3.2. Algoritmos basados en instancia	38
4.3.3. Algoritmos de regularización	38
4.3.4. Algoritmos de árbol de decisiones	38
4.3.5. Algoritmos bayesianos	38
4.3.6. Algoritmos de agrupación	38
4.3.7. Algoritmos de reglas de asociación	39
4.3.8. Algoritmos de redes neurales artificiales	39
4.3.9. Algoritmos de aprendizaje profundo	39
4.3.10. Algoritmos de reducción de dimensionalidad	39
4.3.11. Algoritmos combinados	39
4.3.12. Algoritmos máquinas de soporte vectorial	40
4.4. Algoritmos de aprendizaje por refuerzo	40
4.4.1. Programación dinámica	40
4.4.2. Método de Monte Carlo	40
4.4.3. Aprendizaje por Diferencias Temporales	40
4.5. Problemas de clasificación y regresión	41
4.5.1. Clasificación	41
4.5.1.1. Clasificación binaria	41
4.5.1.2. Clasificación multiclase	41
4.5.2. Regresión	42
4.6. Algoritmos de clasificación en Weka	42
4.7. Técnicas de evaluación para aprendizaje supervisado	45
4.7.1. Método de retención	45
4.7.2. Validación cruzada de k iteraciones	46
4.8. Métricas de desempeño para problemas de clasificación	46

4.8.1.	Porcentaje de acierto	47
4.8.2.	Matriz de confusión	47
4.8.2.1.	Precisión (accuracy)	49
4.8.2.2.	Exactitud (precision)	49
4.8.2.3.	Sensibilidad (sensitivity)	49
4.8.2.4.	Especificidad (specificity)	49
4.8.2.5.	Medida F (f-measure)	50
4.8.2.6.	Estadística kappa (kappa Statistic)	50
4.9.	Componentes de los errores de predicción: Ruido, Sesgo y Varianza	51
4.9.1.	Ruido	51
4.9.2.	Sesgo	51
4.9.3.	Varianza	52
4.10.	Complejidad del modelo y curvas de aprendizaje	52
4.11.	Trabajos relacionados a Pronóstico - Business Intelligence - Machine Learning	54
5.	Propuesta de solución	57
5.1.	Esquema general de solución	57
5.2.	Componente de solución “Business Intelligence”	57
5.2.1.	Fuente de Información	58
5.2.2.	Proceso ETL.	59
5.2.3.	Datawarehouse	59
5.3.	Definición de los Indicadores Claves de Rendimiento	61
5.4.	Obtención de los valores para los Indicadores Clave de Rendimiento.	63
5.5.	Asignación de etiquetas	67
5.6.	Resumen	70
6.	Experimentación	71
6.1.	Componente de solución “Machine Learning”	71
6.2.	Datos de entrada de Machine Learning	71
6.3.	Entrenamiento, testeo y evaluación	72
6.4.	Análisis del desempeño	73
6.4.1.	Comparación con línea base ZeroR	73
6.4.2.	Recuentos por tipo de clasificador	75
6.4.3.	Porcentajes de acierto por período	76
6.5.	Componente de solución “Técnica de solución propuesta”	77
6.6.	Discusión	78
6.6.1.	Impacto del período de análisis	78
6.6.2.	Impacto del etiquetado	78

7. Conclusiones	79
7.1. Conclusiones Generales	79
7.2. Aporte de este Trabajo	80
7.3. Trabajos Futuros	80
Bibliografía	81

Índice de Figuras

2.1. Componentes de tendencia en las series de tiempo [ASW ⁺ 11].	10
2.2. Ejemplo del componente cíclico y de tendencia en las series de tiempo [ASW ⁺ 11].	10
3.1. Componentes de Business Intelligence [Can07]	19
3.2. Cubo OLAP	25
4.1. Proyección de crecimiento de datos del 2005 al 2020.	32
4.2. Modelo general para agentes que aprenden [RN04].	34
4.3. Agrupación de algoritmos según Jason Brownlee [REF]	37
4.4. Esquema general de iteraciones para 10-fold Cross-Validation [REF]	46
4.5. Matriz de confusión para clasificación binaria.	48
4.6. Matriz de confusión para clasificación multiclase.	48
4.7. Complejidad del modelo [FHT01]	53
4.8. Curvas de aprendizaje	54
5.1. Esquema general de solución	57
5.2. Tabla de hechos Detalles	61
5.3. Ejemplo de valores KPI para un periodo semanal	64
5.4. Ejemplo de valores KPI para un periodo quincenal	65
5.5. Ejemplo de valores KPI para un periodo mensual	66
5.6. Ejemplo de etiquetado para periodo semanal.	69
6.1. Esquema de entrenamiento y prueba.	72
6.2. Entrenamiento vs Test vs ZeroR para períodos semanales.	74
6.3. Entrenamiento vs Test vs ZeroR para períodos quincenales.	74
6.4. Entrenamiento vs Test vs ZeroR para períodos mensuales.	75
6.5. Conteo de clasificadores para períodos semanales.	75
6.6. Recuento de clasificadores para períodos quincenales.	76
6.7. Recuento de clasificadores para períodos mensuales.	76
6.8. Promedio de porcentaje de aciertos para los tres períodos de análisis.	77

Índice de Tablas

3.1. Áreas de aplicación de BI	27
4.1. Clasificadores Weka	43
5.1. Rango KPI Ticket Medio	67
5.2. Rango KPI Cifra Ventas (%)	67
5.3. Rango KPI Margen Comercial (%)	67
5.4. Rango KPI Rotación Stock	68
5.5. Rango KPI Cobertura Stock	68
5.6. Tabla de etiquetado por el experto	68

Algoritmos

6.1. Pseudocódigo para el proceso de clasificación.	72
6.2. Pseudocódigo para el proceso de pronóstico de la demanda.	77

Lista de Siglas y Abreviaturas

AR	Auto Regressive
ARIMA	Auto Regressive Integrated Moving Average
ARMA	Auto Regressive Moving Average
BI	Business Intelligence
CERN	Conseil Europeen pour la Recherche Nucleaire
CRM	Customer Relationship Management
CSV	Comma Separated Values
DP	Dynamic Programming
DW	Datawarehouse
ERP	Entreprise Resource Planning
ETL	Extract, Transform and Load
FN	Falsos Negativos
FP	Falsos Positivos
GPS	General Problem Solver
HOLAP	Hybrid Online Analytical Processing
HTML	HyperText Markup Language
IA	Inteligencia Artificial
IBM	International Business Machines
IDC	International Data Corporation
IVA	Impuesto Valor Agregado
KPI	Key Performance Indicators

LASSO	Least Absolute Shrinkage and Selection Operator
LMT	Logistic Model Trees
LWL	Locally Weighted Learning
MA	Moving Average
MAD	Median Absolute Desviation
MIT	Massachusetts Institute of Technology
ML	Machine Learning
MOLAP	Multidimensional Online Analytical Processing
MSE	Mean Squared Error
ODS	Operational Data Store
OLAP	Online Analytical Processing
PC	Personal Computer
PCA	Principal Component Analysis
PDF	Portable Document Format
ROLAP	Relational Online Analytical Processing
SaaS	Software as a Service
SARIMA	Seasonal Autoregressive Integrated Moving Average
SCM	Supply Chain Management
SMO	Sequential Minimal Optimization
SNARC	Stochastic Neural Analog Reinforcement Calculator
SQL	Structured Query Language
SVM	Support Vector Machines
TD	Temporal-Difference Learning
VN	Verdaderos Negativos
VP	Verdaderos Positivos
Weka	Waikato Environment for Knowledge Analysis

Capítulo 1

Introducción

En *empresas retail*¹ o de ventas minoristas uno de los principales problemas con que se enfrentan es el manejo eficiente de stock de tal manera a evitar tener productos en exceso en los depósitos que incurran en sobrecostos. En el otro extremo, la falta de dichos productos o ruptura de stock conlleva pérdidas de oportunidades de ventas por no disponer del producto lo cual ocasiona insatisfacción de clientes, a su vez ésto repercute sobre las utilidades de la empresa. Uno de los mayores desafíos en empresas de este sector es pronosticar de manera eficiente la demanda para el próximo período comercial [Gio17].

Actualmente en el proceso de gestión de compras se utilizan ciertas técnicas de pronósticos para determinar cantidades de las órdenes de compra, dichas técnicas pueden ser cuantitativas o cualitativas [ASW⁺11]. Independientemente de la técnica elegida, el problema real de los pronósticos es su falta de confiabilidad, ya que por lo general no son precisos, entonces, la interrogante es si los pronósticos serán superiores o inferiores a la demanda real y en que medida [JLF12].

A raíz de lo anterior en el presente trabajo se plantea una nueva técnica para estimar volúmenes de demanda para el próximo período comercial. Luego, esta estimación servirá de apoyo para la toma de decisión de la cantidad establecida en las órdenes de compras para la reposición de stock. En este nuevo enfoque se integran técnicas de Business Intelligence (conocida por sus siglas BI) y Machine Learning (conocida por las siglas ML) [CC10][SV08].

En la etapa de Business Intelligence el objetivo principal es calcular los Indicadores Claves de Rendimiento abreviado por las siglas KPI (*del inglés “Key Performance Indicators”*), de los productos en base a datos históricos obtenidos de las bases de datos transaccionales. Luego, cada serie de KPI obtenidos pasan por un proceso de etiquetado, donde el experto en compras analiza y determina qué nivel de compra conviene para cada serie de KPI, para el proceso de aprendizaje [Alv13].

En la etapa de Machine Learning se utilizan como entrada las series de KPI con sus etiquetas obtenidas de la etapa anterior y constituyen las instancias que alimentan los distintos algoritmos

¹Una empresa retail es cualquier comercio que vende sus productos al consumidor final, desde un supermercado a una tienda de barrio, desde un negocio de electrodomésticos a una franquicia textil, ya sea con cientos de puntos de venta o con un solo establecimiento.

de clasificación. Nótese que el enfoque corresponde a un proceso de aprendizaje supervisado. Posteriormente tienen lugar los procesos de entrenamiento y testeo para finalmente evaluar los distintos desempeños.

1.1. Motivación

En general las técnicas de pronóstico de demanda cuantitativas dependen en gran medida de la variable: *cantidad de ventas*. Mientras que los pronósticos de demanda cualitativos dependen en gran medida de la opinión y experiencia del experto; que si bien pueden ser válidos, se utilizan principalmente en casos donde la proyección es a largo plazo o por carencia de datos históricos [ASW⁺11].

Otra debilidad de los pronósticos es que no se adaptan, es decir, las herramientas de pronósticos existentes en su mayoría utilizan algoritmos genéricos con la suposición de que la demanda puede predecirse de igual forma para todas las empresas, industrias, productos y para cualquier ubicación geográfica. Este enfoque de algoritmos genéricos obtiene pronósticos que no reflejan el impacto de los diferentes factores que afectan la demanda; de la misma forma no se adapta a la evolución del mercado y al comportamiento de los consumidores [Gio17].

La oportunidad que surge es poder implementar una técnica de pronóstico de demanda automática o con una mínima intervención manual. Esta técnica permitiría paliar las debilidades de los modelos actuales, y a su vez integraría características esenciales de las técnicas cualitativas y cuantitativas existentes.

Como ya se mencionó, cualquiera sea la técnica aplicada el problema con los pronósticos es que generalmente no son precisos. En [SB98a], los autores mencionan que las desviaciones normales del pronóstico en general con respecto a los valores reales oscilan entre el 5 y el 40 %. Para el sector retail, la mayoría de las empresas puede pronosticar la demanda total de todos los productos, como un grupo con errores menores al 5 %, no obstante el pronóstico de la demanda de un producto puede generar errores considerablemente mayores [ASW⁺11].

En otros ámbitos, la importancia del pronóstico es de igual relevancia. Según un estudio realizado sobre pronósticos de tráfico en proyectos de infraestructura de transporte con una muestra importante que cubrió 210 proyectos en 14 países con una inversión de U\$\$ 59.000 millones demostró que los pronosticadores realizaron un trabajo pobre de estimación de la demanda, en 9 de cada 10 proyectos ferroviarios las previsiones de pasajeros fueron sobrestimadas con un promedio del 106 %, mientras que para la mitad de los proyectos de carreteras la diferencia entre el tráfico real y el previsto fue ± 20 %. El resultado fue un riesgo financiero sustancial que los tomadores de decisiones suelen ignorar o minimizar en detrimento del bienestar social y económico [FHB05].

1.2. Planteamiento del Problema

Dado un conjunto de productos que la empresa retail “*MARKETPLUS S.A*” ofrece en venta, este trabajo aborda el problema del pronóstico de la demanda para la reposición de stock. Se considera que la empresa “*MARKETPLUS S.A*” opera un stock con períodos de reposición regulares y no se analizan los productos estacionales.

En este contexto, este trabajo apuesta por una solución que automatice la toma de decisiones de reposición de stock, una técnica de pronóstico de la demanda basada en Business Intelligence y Machine Learning aplicadas en forma secuencial; que busca a través de los KPI, la opinión de un experto en compras y mediante algoritmos de clasificación prever volúmenes eficientes de productos para la reposición de stock. En la primera etapa, a partir del histórico de movimientos de stock y los detalles de ventas, el módulo de Business Intelligence obtiene los KPI asociados por producto y período. En la segunda etapa, estos datos son utilizados como entradas al módulo de Machine Learning para determinar los volúmenes de compras para el siguiente periodo de ventas.

1.3. Objetivos

A continuación se dan a conocer el objetivo general y los objetivos específicos que se pretenden alcanzar con este trabajo.

1.3.1. Objetivo General

El objetivo general y aporte principal es:

- Desarrollar una nueva herramienta de pronóstico de la demanda para la toma de decisión en la reposición de stock integrando sistemas y conceptos de Business Intelligence y Machine Learning.

1.3.2. Objetivos Específicos

Los objetivos particulares para el logro del objetivo principal son:

- Reportar las principales soluciones que abordan el problema.
- Aplicar el proceso de Business Intelligence para obtener los KPI.
- Realizar el proceso de etiquetado y obtener las instancias de entrenamiento.
- Realizar el proceso de entrenamiento con los distintos clasificadores.
- Evaluar el rendimiento de los distintos clasificadores entrenados.

1.4. Organización del Trabajo Final de Grado

El libro se estructura de la siguiente manera:

- El Capítulo 2 presenta el concepto de Pronóstico de la Demanda y las principales técnicas de pronóstico utilizadas en la actualidad.
- El Capítulo 3 aborda los conceptos de Business Intelligence.
- El Capítulo 4 se enfoca en los conceptos referentes a Machine Learning.
- En el Capítulo 5 se explica el proceso de Business Intelligence para obtener los KPI junto con el etiquetado de los datos.
- En el Capítulo 6 se desarrolla el proceso de entrenamiento y selección de los algoritmos clasificadores, se muestran los resultados experimentales y se analiza el desempeño de los algoritmos.
- En el Capítulo 7 se extraen las conclusiones generales del trabajo, se contrastan los objetivos propuestos con los resultados obtenidos. Se describen las propuestas de trabajos futuros que complementarían la técnica propuesta.

Capítulo 2

Pronóstico de la Demanda

El pronóstico de la demanda es una técnica utilizada para estimar la cantidad de bienes o servicios que los consumidores demandarán en el futuro, basadas en el conocimiento del pasado, como son los datos históricos [Cha06].

La elaboración de pronósticos de demanda precisos es uno de los retos importantes en empresas del tipo retail, determinar qué sucederá en el futuro con el fin de tomar decisiones adecuadas es un problema que se presenta con frecuencia [SB98a].

En un escenario inicial se tienen depósitos llenos de productos listos para ser llevados a los mostradores. A medida que pasa el tiempo la cantidad en depósito va decreciendo por la demanda de los clientes y llegado un momento crítico hay que tomar la decisión de reponer el stock. Si bien la reposición de stock se lleva a cabo dentro de un proceso empresarial llamado *Administración de Compras*, hay un componente vital dentro de este proceso que es la estimación de la cantidad o volumen de productos a adquirir para reponer el stock. Es ahí donde entra en juego el pronóstico de la demanda [AE91].

A continuación se explica el proceso de *Administración de Compras*, para luego analizar las principales técnicas de pronósticos de demanda que están vigentes en el mundo empresarial.

2.1. Administración de compras

Los términos *compras*, *adquisiciones*, *administración de materiales*, *logística*, *abastecimiento*, *administración del suministro* y *administración de la cadena de suministro* se utilizan de manera indistinta ya que no existe un consenso general sobre la terminología. El proceso de adquisición es el eje central de la actividad empresarial de administración de compras o suministros. Cualquier organización requiere de proveedores por lo que es importante incorporarlos con efectividad al entorno organizacional, y que las decisiones de compras no contradigan las estrategias de la empresa. Las empresas centran sus esfuerzos en aumentar ingresos, disminuir costos, o una combinación de ambos a fin de obtener ganancias de forma más eficiente [JLF12].

Este trabajo busca contribuir en lograr decisiones eficientes de compras basadas en pronósticos de demanda precisos. Se considera como un componente importante dentro de la planeación

estratégica y operacional la estimación eficiente de la cantidad o volumen de productos para la reposición del stock del siguiente período de ventas [PMAR07].

El stock de una empresa son los materiales y artículos almacenados, tanto los que son utilizados para la producción como aquellos destinados para la venta al consumidor final [DDM15]. Las funciones que desempeña el stock o existencia en una empresa son [PB10]:

- Evitar la escasez, ante la incertidumbre de la demanda o ante un posible retraso en la reposición o suministro de los pedidos.
- Aprovechar la disminución de los costes a medida que aumenta el volumen de compras o de fabricación.
- Lograr un equilibrio entre las compras y las ventas para alcanzar la máxima competitividad.

El proceso de compras o adquisiciones se trata de un conjunto de etapas:

1. Detectar la necesidad.
2. Traducir la necesidad en una especificación comercial.
3. Buscar potenciales proveedores.
4. Seleccionar el proveedor adecuado.
5. Detallar la orden de compra y pactar el suministro.
6. Recibir los productos.
7. Pagar a los proveedores.

En el punto 5, en el detalle de la orden se ven reflejadas las estimaciones de las cantidades a comprar de los productos, en este proceso el caso ideal por supuesto sería poder adivinar por cada producto la cantidad exacta que se va a vender en el siguiente periodo de venta. De este modo al finalizar cada periodo de ventas se dispondría de un stock de cero unidades, con lo cual se llegaría a una máxima eficiencia en la cantidad o volumen de productos comprados. Sin embargo adivinar es una tarea imposible, lo que si se puede hacer es predecir eficientemente la demanda futura.

Del porqué la importancia de estimar de forma correcta esta cantidad o volumen, los expertos en negocios explican que los productos parados en stock mientras no se venden es capital inmovilizado, que generan sobrecostos por mantenimiento, seguros, personal encargado, fecha de vencimiento de los productos, etc [PMAR07]. Otro hecho no deseado a causa de una mala estimación es la ruptura de stock, es decir, no disponer de un producto cuando existan clientes interesados en comprarlo, lo cual se considera una pérdida para la empresa. En este punto, con una estimación acertada lo que se desea es mantener un nivel de stock eficiente, es decir, tener

suficiente cantidad para satisfacer la demanda sin roturas de stock y por otra parte evitar que haya un exceso inútil.

Si bien el presente trabajo no toma en cuenta los costos asociados al almacenamiento, lo que se busca es pronosticar la cantidad o volumen de demanda de productos de forma eficiente utilizando las herramientas de Business Intelligence y Machine Learning para el siguiente periodo.

Una buena gestión de las compras contribuye significativamente al éxito de una organización. La función del suministro evoluciona a medida que la tecnología y el ambiente competitivo mundial requieren enfoques innovadores [JLF12].

Antes de realizar una compra surgen las siguientes preguntas:

- ¿Cuándo debemos realizar un pedido?
- ¿Qué cantidad debemos solicitar en cada pedido?
- ¿Cuántas unidades de cada artículo debemos mantener en stock?

Para responder estas preguntas una de las herramientas que ayudan son las técnicas de pronósticos de demanda, entre las que se destacan los *Métodos de Pronósticos Cualitativos* y los *Métodos de Pronósticos Cuantitativos* [ASW⁺11] [HH08].

2.2. Métodos de pronósticos cualitativos

Son pronósticos subjetivos porque intervienen factores como la experiencia, intuición o emociones del que toma la decisión.

Las principales técnicas cualitativas de pronóstico de demanda son [VAH11][HH08]:

- Opinión del Gerente: El pronóstico se basa en la opinión, experiencia o el conocimiento técnico de las condiciones de un solo gerente. Pueden haber datos en los cuales el gerente apoya su decisión.
- Junta de opinión ejecutiva: Similar al método anterior, la diferencia está en que se basa en un grupo de ejecutivos que intercambian opiniones, perspectivas y conocimientos, luego formulan y componen ideas comunes que sirven de base para emitir un pronóstico unificado, compartiendo de este modo la responsabilidad.
- Consulta a la fuerza de ventas: Esta técnica se basa en la experiencia del personal más cercano al cliente que es el cuerpo de vendedores de la empresa. Cada vendedor realiza una estimación de la demanda en su zona de influencia. Luego las estimaciones son revisadas por los mandos superiores, para obtener un pronóstico corporativo final.
- Encuesta en el mercado de consumo: Se encuesta a los clientes acerca de sus planes de compras, sus intereses por determinados productos o posibles nuevas características. La

estimación se extrae de los resultados de las encuestas. Son útiles para elaborar planes de marketing, lanzamiento de nuevos productos, etc.

- Método Delphi: Se basa en identificar un panel de expertos que pueden ser gerentes, empleados, o competentes en el área. Se tiene un cuestionario donde cada uno de ellos lo completa de forma aislada. Se integran todas las respuestas, luego cada experto tiene acceso al set de respuestas y puede ajustar su respuesta conforme le parezca conveniente. Este proceso se repite de forma iterativa hasta alcanzar un cierto nivel de consenso. Finalmente los resultados de este panel de expertos sirven de base para las decisiones de pronóstico de los gerentes.
- Analogía de productos similares: Se basa en el comportamiento de las ventas de un producto similar o modelo. Técnica útil para nuevos productos que se quieren introducir en el mercado y de los cuales no se dispone de información histórica de ventas, entonces se puede pronosticar haciendo analogía con productos sustitutos o complementarios.

2.3. Métodos de pronósticos cuantitativos

Estos modelos se basan en métodos de pronósticos estadísticos que a partir de los datos históricos de ventas y suponiendo que las tendencias históricas continuarán, son capaces de anticipar la demanda futura [HH08].

En general, para modelar cuantitativamente se debe disponer de información sobre la variable a pronosticar, la información se debe cuantificar y el patrón histórico de cierto modo debe repetirse en el futuro [ASW⁺11].

El pronóstico de la demanda de productos es sólo una aplicación importante de estos métodos. En otros casos, los pronósticos se podrían utilizar para evaluar los requerimientos de cantidades variadas como partes de repuestos, rendimiento de la producción y necesidades de personal. Las técnicas de pronóstico se usan también para anticipar las tendencias económicas a nivel regional, nacional e incluso internacional [HH08].

En general, los métodos cuantitativos se clasifican en técnicas de series de tiempo y en pronósticos causales.

2.3.1. Pronósticos de series de tiempo

Una serie de tiempo consiste es un serie de registros observados de la variable a pronosticar, tomados en períodos progresivos del tiempo pasado [HH08]. El histórico de ventas de un producto donde se observan valores diarios de las cantidades vendidas constituye un buen ejemplo de serie de tiempo. Los datos históricos de la variable a predecir están limitados a sus valores pasados.

El objetivo del método es obtener una buena predicción del valor futuro de la variable a pronosticar, enmarcado en la serie de tiempo. Para lograr el objetivo, el método debe descubrir

el patrón dentro de la serie y luego ser capaz de proyectarlo hacia el futuro [ASW⁺11]. De cierta manera, hay una suposición intrínseca del modelo de que los factores que influyen en las ventas pasadas y presentes continuarán a futuro.

Si bien el volumen de ventas es un buen indicador de la historia de la demanda, no toma en cuenta muchos aspectos del proceso entero de las ventas, como pueden ser la ruptura de stock, plazos de reposición de stock, precio del producto, la incidencia del marketing u otros. De igual modo se pueden descubrir tendencias, estacionalidad, ciclos, etc., en la historia de la demanda para luego extrapolarlo a un tiempo futuro. También hay que destacar que el intervalo de muestreo tiene mucha influencia en el pronóstico y por ende en los resultados obtenidos [PMAR07].

En el sentido estricto de la interpretación, es erróneo hablar de pronosticar el siguiente valor de la observación en una serie de tiempo. Como este valor puede ser cualquiera y dependerá de circunstancias futuras que son ajenas al control humano, entonces es imposible predecirlo exactamente. El siguiente valor de una serie de tiempo es una variable al azar y tiene alguna distribución de probabilidades. Si ese siguiente valor es la media de la distribución de probabilidades acertaría el problema, pero se desconoce su distribución de probabilidades así como también su media.

Lo mejor que se puede realizar es una estimación de la media tan cerca como sea posible, utilizando todos los datos disponibles. La meta de los métodos de pronóstico de series de tiempo es estimar la media de la distribución de probabilidades subyacente del siguiente valor de la serie de tiempo. Para una serie de tiempo que tiene exactamente la misma distribución para todos y cada uno de los periodos, el método de pronóstico de promedios proporciona la mejor estimación de la media, pero en general se usan otros métodos de pronósticos porque la distribución cambia con el paso del tiempo. Si la distribución de probabilidad de una serie sigue siendo la misma en el siguiente periodo entonces se dice que es estable (puede haber cambios en la distribución pero deben ser pequeños). Si la distribución de probabilidad presenta cambios grandes y frecuentes entonces se dice que es inestable [HH08].

El rol del analista es capturar los componentes del patrón de la demanda y luego traducirlo a un valor de pronóstico. Para Johnson [JLF12] este patrón tiene seis componentes básicos: valor constante (la fluctuación de los datos alrededor de una media constante), tendencia (el incremento o decremento sistemático de la media a lo largo del tiempo), variaciones estacionales, cíclicas, aleatorias y puntos críticos. Para Anderson [ASW⁺11] el patrón de los datos en una serie de tiempo tienen cuatro componentes separados: tendencia, cíclico, estacional e irregular, y que luego se combinan para generar los valores de la serie de tiempo.

Se analizan estos cuatro componentes:

1. *Componente de tendencia*: Los valores de la serie de tiempo pueden ir cambiando continuamente, tendiendo hacia valores que incrementan o que disminuyen. Cuando estos incrementos o disminuciones se dan por periodos de tiempo prolongados se dice que la serie tiene un componente de tendencia. La figura 2.1.a muestra una tendencia no li-

neal; en este caso la serie de tiempo indica poco crecimiento inicial, luego un periodo de rápido crecimiento y por último una estabilización. En la figura 2.1.b la tendencia lineal decreciente de la gráfica es útil para las series de tiempo que muestran una declinación constante en el tiempo. En la figura 2.1.c la línea horizontal en la gráfica representa una serie de tiempo que no tiene un aumento o disminución constante en el tiempo, y por tanto no muestra tendencia. Generalmente el componente tendencia se debe a factores a largo plazo.

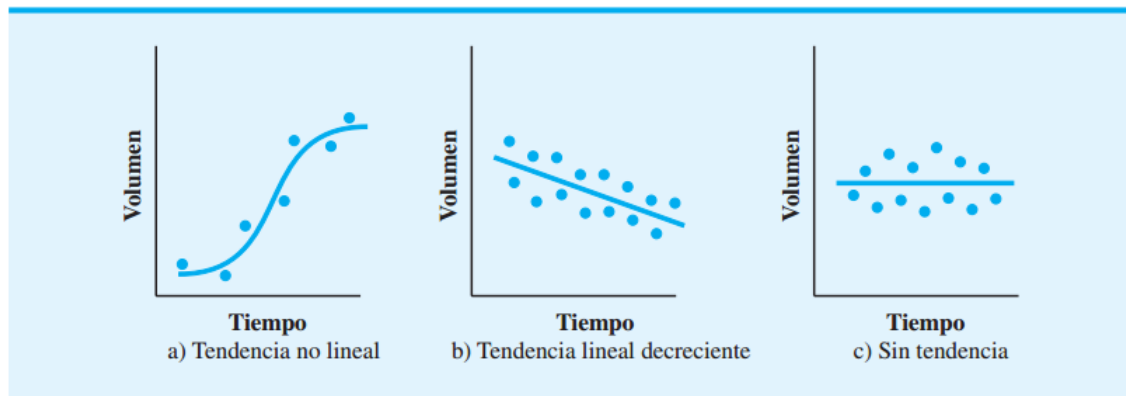


Figura 2.1: Componentes de tendencia en las series de tiempo [ASW⁺11].

2. *Componente cíclico*: es habitual que los puntos de la serie de tiempo se encuentren por encima o por debajo de la línea de tendencia. Cuando el patrón de puntos está de forma alterna por encima y por debajo de la línea de tendencia durante períodos mayor a un año, entonces estamos ante presencia del componente cíclico de la serie de tiempo. Esto se observa en la figura 2.2

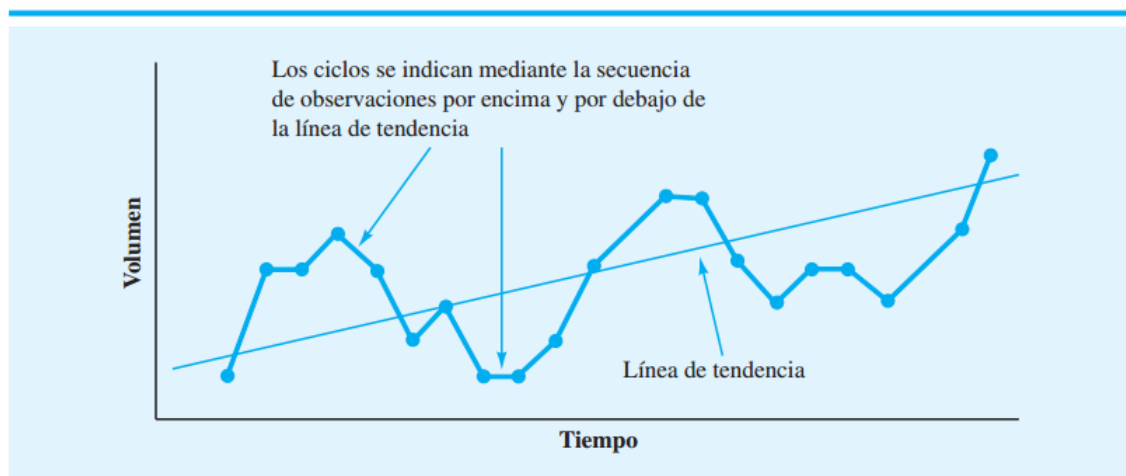


Figura 2.2: Ejemplo del componente cíclico y de tendencia en las series de tiempo [ASW⁺11].

3. *Componente estacional*: Cuando el patrón de puntos está por encima o por debajo de la línea de tendencia durante períodos alternos menores o igual a un año, entonces estamos ante presencia del componente estacional de la serie de tiempo. Las altas cantidades de

ventas de abrigos durante el otoño e invierno son patrones que se repite debido a la influencia estacional. En ciertos tipos de productos, las ventas que se anticipan en un mes particular están influidas por la temporada del año. Por ejemplo, un producto que es popular en Navidad, podría tener ventas en diciembre que son dos veces mayores que las ventas de enero [HH08].

4. *Componente irregular*: Se refiere a la presencia de variabilidad aleatoria en la serie de tiempo. Corresponde a aquellos puntos que aparecieron desviados en relación a lo esperado de los efectos del componente de tendencia, cíclico y estacional. Como este componente es impredecible no se puede cuantificar el impacto que tiene en la serie de tiempo. Generalmente es debido a factores a corto plazo o a circunstancias casuales.

En general, los métodos de series de tiempo se clasifican en [HH08]:

- **El método de pronóstico del último valor**: Este método utiliza solamente el último valor de la serie de tiempo como pronóstico del valor futuro. También es conocido como método ingenuo, porque sin mucho análisis aparentemente resulta ingenuo elegir un solo valor de toda la serie. Pero en ocasiones sí es una buena aproximación, como por ejemplo cuando hay demasiada fluctuación en la serie y entonces el último valor se convierte en el más fiable. Recomendable para series de tiempo inestables.

$$\text{Pronóstico} = \text{último valor} \quad (2.1)$$

- **El método de pronóstico por promedios**: En este caso se utilizan todos los valores de la serie y luego se promedia para obtener el valor de pronóstico de la serie. Recomendable para series de tiempo estables, razón por la cual todos los valores tienen el mismo peso y son considerados relevantes.

$$\text{Pronóstico} = \text{promedio de todos los valores hasta la fecha} \quad (2.2)$$

- **El método de pronóstico de promedio móvil**: Consiste en considerar solamente los últimos n períodos y luego promediar para así obtener el valor de pronóstico de la serie. Recomendable para series de tiempo medianamente estables, razón por la cual se toman en cuenta únicamente n valores que tienen el mismo peso y que son considerados importantes.

$$\text{Pronóstico} = \text{promedio de los últimos } n \text{ valores} \quad (2.3)$$

donde:

$$n = \text{número de periodos más recientes}$$

- **El método de pronóstico por suavizamiento exponencial**: Con este método se asignan pesos diferentes a los valores de la serie. El último período es el de mayor peso y así paulatinamente se van asignando pesos cada vez menores a los valores más antiguos

de la serie. Este resultado se puede obtener de forma simple y sintética mediante una combinación del último valor de la serie y del último pronóstico correspondiente a dicho último valor.

$$\text{Pronóstico} = \alpha \times (\text{último valor}) + (1-\alpha) \times (\text{último pronóstico}) \quad (2.4)$$

donde:

α es una constante entre 0 y 1 llamada “constante de suavizamiento”.

Para series de tiempo estables es recomendable valores de α pequeños como 0,1, y para series de tiempo inestables valores mayores. En general en aplicaciones de hoy en día se utilizan valores entre 0,1 y 0,3.

- **El método de suavizamiento exponencial con tendencia:** El inconveniente del método de pronóstico por suavizamiento exponencial sin tendencia es que justamente se retrasa respecto de la tendencia ya que no lo toma en cuenta. Al considerar la tendencia se obtienen pronósticos más precisos. Se calcula la pendiente actual de la línea de tendencia para luego ajustar el nuevo pronóstico a la pendiente obtenida. Los valores más recientes de la serie de tiempo se utilizan para obtener la línea de tendencia actual y pueden tener dirección ascendente, descendente u horizontal.

$$\text{Pronóstico} = \alpha \times (\text{último valor}) + (1-\alpha) \times (\text{último pronóstico}) + \text{tendencia estimada} \quad (2.5)$$

donde:

$$\text{tendencia estimada} = \beta \times (\text{tendencia actual}) \times (1-\beta) \times (\text{tendencia estimada anterior}) \quad (2.6)$$

$$\begin{aligned} \text{tendencia actual} = & \alpha \times (\text{último valor} - \text{penúltimo valor}) + \\ & (1-\alpha) \times (\text{último pronóstico} - \text{penúltimo pronóstico}) \end{aligned} \quad (2.7)$$

con β (beta) es una constante de suavizamiento de tendencia entre 0 y 1. La elección del valor y rango de β tienen igual significado que α .

- **Auto Regressive Integrated Moving Average (ARIMA):** Es una generalización de Auto Regressive Moving Average (ARMA), que combina procesos autorregresivos (Auto-regressive, AR) y procesos de media móvil (Moving Average, MA) creando así un modelo compuesto de series de tiempo. ARIMA (p, d, q) captura los siguientes elementos [SNN18]:

- AR: Autoregression. Un modelo de regresión que utiliza las dependencias entre una observación y un número de observaciones pasadas (p).
- I: Integrated. Para hacer la serie temporal estacionaria midiendo las diferencias de observaciones en diferentes momentos (d).
- MA: Moving Average. Un enfoque que tiene en cuenta la dependencia entre las observaciones y los términos de error residual cuando se usa un modelo de promedio móvil para las observaciones pasadas (q).

En econometría, los pronósticos de series de tiempo se aplican tradicionalmente utilizando modelos ARIMA, generalizado por Box y Jenkins [BJ73]. ARIMA es un método estándar para el pronóstico de series de tiempo y habituales en el modelado de series temporales económicas y financieras.

2.3.2. Pronósticos causales

Ciertamente las series de tiempo se basan en un solo indicador clave, como lo es por ejemplo la variable ventas. Siguiendo el ejemplo, el objetivo de la serie de tiempo es encontrar un valor de pronóstico de la variable ventas a partir de valores pasados. Ahora bien, si tenemos dos variables en relación causa-efecto las series de tiempo no nos sirven.

El pronóstico causal obtiene una proyección de la cantidad de interés (la variable dependiente) relacionándola directamente con una o más cantidades (las variables independientes) que impulsan la cantidad de interés. Por ejemplo, las promociones sobre uno o varios productos pueden ser la causa de una mayor cantidad de ventas en dichos artículos, como tal tenemos una relación causa (promociones)-efecto (mayores ventas), es decir las promociones provocan cambios en los niveles de ventas [HH08].

Una de las técnicas para resolver problemas de pronósticos causales es la *regresión lineal*. El objetivo de este método es encontrar la línea recta que más se aproxime a la relación entre la variable dependiente y la/s variable/s independiente/s. Cuando hay una sola variable independiente la forma de la ecuación es la de la recta:

$$y = a + bx \quad (2.8)$$

donde:

y = variable dependiente,

x = variable independiente,

a = intersección de la línea con el eje y ,

b = pendiente de la línea

Para obtener a y b se utiliza el método llamado de *mínimos cuadrados*, que encuentra el par de valores a y b tal que la suma del cuadrado de los errores de estimación sea el menor posible. Para problemas donde se consideran varios indicadores clave como variables independientes, la ecuación presenta la siguiente forma:

$$y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (2.9)$$

Donde el proceso de obtención de a y b_1, b_2, \dots, b_n es también por el método de *mínimos cuadrados*.

2.3.3. Precisión de los métodos cuantitativos

Para medir la desviación que hay entre el pronóstico y el valor real se utiliza normalmente el valor del *error promedio del pronóstico*, conocido como *Desviación Media Absoluta (MAD - Median Absolute Deviation)* y que se calcula con la siguiente fórmula:

$$MAD = \frac{\text{suma de los errores de pronóstico}}{\text{número de pronósticos}} \quad (2.10)$$

Otra forma significativa de medir la precisión es a través del *Error Cuadrático Promedio (MSE - Mean Squared Error)* y que se calcula de la siguiente forma:

$$MSE = \frac{\text{suma de los cuadrados de los errores de pronóstico}}{\text{número de pronósticos}} \quad (2.11)$$

El resultado de esta última fórmula pone de relieve los errores grandes de pronóstico, así como también destaca si el método de pronóstico es preciso. Se utiliza como complemento informativo a MAD.

2.4. Revisión Literaria acerca del pronóstico de la demanda

Los pronósticos de la demanda son de vital importancia para todas las organizaciones de negocios, así como para la toma de decisiones importantes de los gerentes, es una técnica fundamental utilizada para el logro de los objetivos en la decisiones estratégicas y tácticas de la organización y su gestión. El propósito del pronóstico es estimar la cantidad de bienes y servicios que los clientes consumirán en el futuro [SS17]. En esta sección se hace una revisión acerca de la literatura existente respecto al pronóstico de la demanda, los métodos de pronósticos utilizados y sus áreas de aplicación.

En el trabajo [HGS17], los autores se enfocan en pronosticar la demanda diaria de productos alimenticios perecederos en un punto de venta de una panadería industrializada utilizando el modelo ARIMA multivariante para respaldar las decisiones operativas. Con el modelo utilizado

llegaron a la conclusión de que es posible aumentar la disponibilidad y disminuir la pérdida económica. El enfoque no es óptimo para los minoristas que poseen varios cientos de tiendas, ya que la precisión no es confiable y ocasiona el desabastecimiento regular de productos.

En [AY16], los autores estudian la previsión de la demanda anual de gas natural para la ciudad de Sakarya en Turquía. El pronóstico se basa en datos del consumo de gas natural de los años 2011-2014, recopilados en periodos mensuales de hogares y usuarios comerciales de bajo consumo. Para el análisis utilizaron y compararon tres métodos estadísticos estacionales univariados, descomposición de series de tiempo, suavizamiento exponencial de Holt-Winters y método de promedio móvil integrado autorregresivo estacional (SARIMA). Los autores concluyen que todos los métodos proporcionan resultados satisfactorios y las diferencias entre cada método son muy bajas, a medida que aumenta la complejidad de cómputo del modelo la precisión del pronóstico aumenta y disminuyen las tasas de error.

En [Big12], el autor plantea construir un modelo ARIMA estacional utilizando la metodología de Box-Jenkins para pronosticar a corto plazo la llegada de turistas y pernoctaciones turísticas en Montenegro. El periodo de tiempo para el estudio abarca diez años que va desde 01/2001 a 12/2010, un análisis minucioso de las series temporales no revelaron nada extremo ni inusual sobre los datos. Por lo tanto, solo los impactos económicos fueron afectados por las series de tiempo. Esto fue importante porque se excluyó el análisis econométrico en el diseño y la construcción de los modelos.

Los resultados muestran excelentes resultados en las predicciones. Según la previsiones de producción para el año 2011, Montenegro puede esperar un crecimiento tanto en llegadas de turistas así como pernoctaciones turísticas. El modelo ha mostrado un aumento alrededor del 7.25 % en la llegada de turistas respecto al año anterior. Por otro lado, el aumento calculado de las pernoctaciones es de alrededor del 8.42 % más que el año anterior. Al término del año 2011, con datos reales de llegadas de turistas hubo un crecimiento del 8.74 % mientras que para las pernoctaciones de turistas hubo un crecimiento del 10.16 % respecto al año anterior, esto demuestra que los modelos utilizados son adecuados para las predicciones con una tasa de error mínima aproximado del 1.5 %.

En el trabajo [AAS14], los autores presentan un caso de estudio de pronóstico de la demanda en una cadena farmacéutica minorista (retail) Apollo Pharmacy que cuenta con 70 puntos de venta en la India, donde la gerencia no ha considerado la previsión como una función importante en el proceso de gestión eficiente de la cadena de suministro. Para el caso de estudio se aplicaron las técnicas Promedio Móvil, Suavizamiento Exponencial Simple y Suavizamiento Exponencial de Winters para pronosticar la demanda de productos farmacéuticos.

Se eligieron en forma empírica dos productos para el estudio: Tableta de 10 mg de Okalet (estacional) y Tableta de Stamlo Beta (no estacional). Los autores concluyen que para productos farmacéuticos estacionales la técnica de Suavizamiento Exponencial de Winters tienen un mejor pronóstico, en tanto, que para productos farmacéuticos que tienen una demanda constante el método de Promedio Móvil tiene una mayor precisión en el pronóstico.

En [PBP15], los autores investigan el nivel de previsión de la demanda en los supermercados

minoristas de productos alimenticios en la República Checa. El objetivo principal de la investigación fue descubrir los métodos que se utilizan para el pronóstico de la demanda, que factores afectan la demanda de alimentos y que fuentes de información utilizan para basar el pronóstico, se realizaron consultas a un total de 75 puntos de venta. Finalizada la investigación los autores concluyeron que hay un uso predominante de los métodos cualitativos, entre los más aplicados están el método de Juicio del Experto (utilizado 40 % de las veces), el método de Expectativas del Cliente (utilizado 9 % de las veces) y el método de la Analogía (utilizado 4 % de las veces).

Entre los métodos cuantitativos mas implementados se encuentran el método de Promedio Móvil (21 %) y el método Naive (19 %). Aplicar métodos cuantitativos en tiendas minoristas con una gran variedad de productos es muy costosa y requiere mucho tiempo. Los autores recomiendan de acuerdo a la literatura revisada el uso de métodos de análisis de series de tiempo sobre ventas pasadas.

2.5. Resumen

Este capítulo pretendió dar a conocer las técnicas clásicas de pronósticos de demanda ampliamente conocidas, las cuales son las técnicas cualitativas y cuantitativas. En general las técnicas más utilizadas son los métodos de ARIMA y Suavizamiento Exponencial, que son técnicas cuantitativas. Entre las empresas comerciales que implementan estas técnicas están los negocios minoristas como supermercados, panaderías, cadenas farmacéuticas, proveedores de gas natural y muchos otros [JLF12][VAH11][HH08] [ASW⁺11].

Si bien el presente trabajo también busca encontrar una estimación de pronóstico como lo hacen las técnicas de esta sección, debe quedar claro que la solución propuesta es una alternativa distinta, es decir, no se trata de un método cualitativo ni cuantitativo propiamente dicho, sin embargo toma ciertos aspectos de ambos y se implementa con conceptos, tecnologías y herramientas diferentes.

Capítulo 3

Business Intelligence

En la actualidad Business Intelligence está siendo cada vez más adoptado por las organizaciones debido a la necesidad de los mandos superiores de contar con información rápida y precisa, necesaria para la toma de decisiones y su importancia a nivel estratégico y operativo. En este capítulo se presentan los conceptos y una introducción a Business Intelligence [Sch08][CC10].

3.1. Definición

Business Intelligence no se trata ni de un producto ni de un sistema, es una arquitectura que engloba un conjunto de conceptos, técnicas de computación y herramientas para analizar y transformar los datos empresariales en información significativa y útil, que permite ser de apoyo a las organizaciones en la toma de decisiones y brindarles una visión estratégica, táctica y operativa más efectivas, mediante un acceso fácil a los datos empresariales. Las tecnologías de Business Intelligence ofrecen vistas históricas, actuales y predictivas de las operaciones, son procesos que se extienden en el tiempo, capaces de manejar grandes volúmenes de datos que ayudan a identificar, crear y desarrollar nuevas estrategias de negocios para mejorar la competitividad [Can07].

La era actual de la creciente información genera la exigencia de tener métodos rápidos y eficientes para obtener los datos de una organización, convertirlo en información y distribuirlo a los gerentes y directivos. Business Intelligence responde a dicha necesidad [MA03][CC10].

El primero en adoptar el término fue Howard Dresner en 1989, quién fue consultor de Gartner Group. Dresner utilizó el término para describir un conjunto de conceptos y métodos que mejoran la toma de decisiones, partiendo de la información disponible acerca de los hechos. Del glosario de términos de Gartner se extrae la siguiente definición [Gar06]:

“Business Intelligence es un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un datawarehouse), para descubrir tendencias o patrones, a partir de las cuales derivar ideas y extraer conclusiones. Las áreas incluyen clientes, proveedores, productos, servicios y competidores. El proceso de Business Intelligence incluye

la comunicación de los descubrimientos y efectuar los cambios”.

The Datawarehouse Institute propone una definición más formal [EH05]:

“Business Intelligence es un término paraguas que abarca los procesos, las herramientas y las tecnologías para convertir datos en información, información en conocimiento y planes para conducir de forma eficaz las actividades de los negocios. Business Intelligence abarca las tecnologías de datawarehousing, los procesos en el ‘back end’¹, consultas, informes, análisis y las herramientas para mostrar información (herramientas de Business Intelligence) y los procesos en el ‘front end’”.

3.2. Objetivos

Según lo expuesto en la definición, Business Intelligence tiene los siguientes objetivos principales [Can07]:

- Convertir datos en información, información en conocimiento, y el conocimiento en planes operativos o estratégicos.
- Facilitar la disponibilidad de información a los usuarios de negocios, que les ayude a tomar decisiones más rápidamente.
- Apoyar de forma sostenible y continua a las organizaciones, para mejorar su competitividad ante el entorno de negocios cambiante de forma que puedan adaptarse a él.
- Ante la cantidad de información que va creciendo, disponer de más tiempo para analizarla, en lugar de gastar mucho tiempo en prepararla, organizarla y estructurarla.
- Permitir a las organizaciones dirigir de mejor forma, decidir, medir, gestionar y optimizar el alcance de la eficiencia y los resultados financieros.
- Disminuir sustancialmente la incertidumbre que existe ante la toma de decisiones respecto a un plan estratégico.

3.3. Componentes de Business Intelligence

Implementar un proyecto de Business Intelligence en una organización es un proceso que sigue una serie de pasos, cada paso puede verse como un componente. En la figura 3.1 observamos los distintos componentes que forman parte de Business Intelligence [Can07]:

¹ Los términos “back end” y “front end” comúnmente usados en Sistemas de Información significan, respectivamente, la parte más cercana al área tecnológica y la más cercana a los usuarios. Si hiciéramos un paralelismo con una tienda, serían la “trastienda” y el “mostrador”.

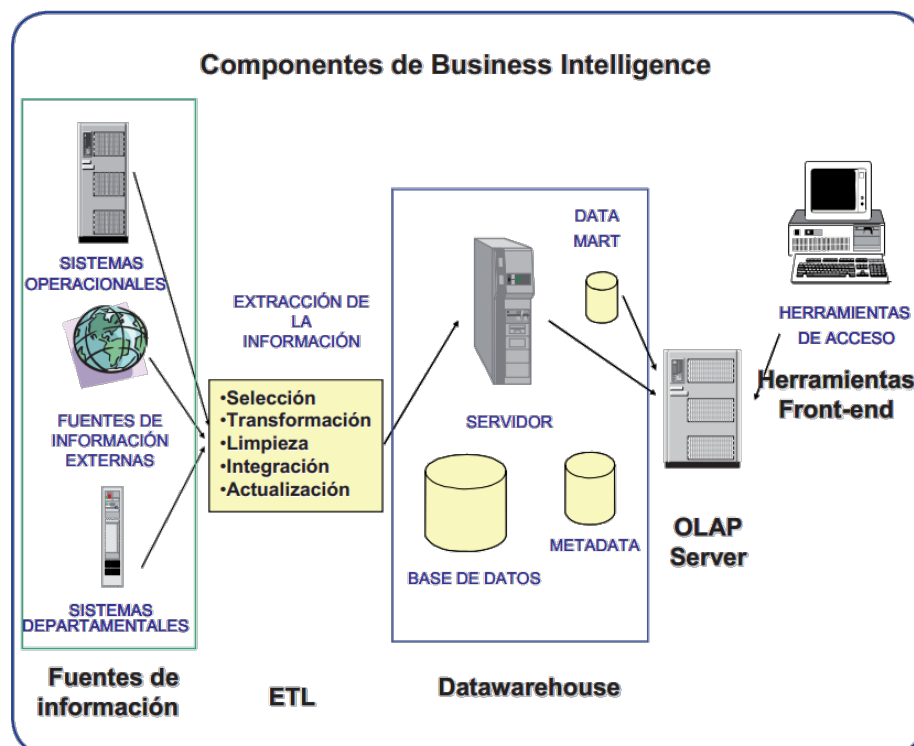


Figura 3.1: Componentes de Business Intelligence [Can07]

A continuación, una breve descripción de los componentes de Business Intelligence:

3.3.1. Fuentes de Información

Las fuentes de información representan el origen de los datos con las cuales se alimenta de información al datawarehouse. Estas pueden provenir de diferentes sistemas como:

- Sistemas transaccionales, aplicaciones a medida, ERP, CRM, SCM, etc.
- Sistemas de información departamentales: previsiones, presupuestos, hojas de cálculo, etc.
- Fuentes de información externa, en algunos casos comprada a terceros. Las fuentes de información externas podrían ser importantes para enriquecer la información acerca de los clientes. Ej.: Base de datos de contribuyentes con RUC, base de datos del registro civil de las personas, etc.

También las fuentes de información son usualmente heterogéneas, pueden contener los siguientes tipos de datos [CC10]:

- Estructurados: almacenados en las bases de datos.
- Semi estructurados: son formatos entendibles por los computadores como HTML tabulado, Excel, CSV u otros que pueden ser obtenidos mediante técnicas estándar de extracción de datos.

- No estructurados: son formatos no legibles para computadoras como Word, HTML no tabulado, PDF, etc. que pueden obtenerse mediante técnicas avanzadas de extracción de datos.

3.3.2. Extracción, Transformación y Carga

La extracción, transformación y carga, comúnmente abreviado por las siglas ETL (*del inglés “Extract, Transform and Load”*) es un tipo de integración de datos que consiste en todo el proceso que se realiza entre las fuentes de información y el área de presentación de los datos [Kim92]. Es utilizado para extraer los datos del origen, transformarlos de acuerdo a la necesidad o requerimientos y cargar los datos al datawarehouse.

El proceso ETL se divide en 5 subprocesos [Can07]:

Extracción

La extracción es el primer paso en el proceso de obtención de los datos, recupera los datos físicamente de las distintas fuentes de información. En este punto se dispone de los datos en bruto. El objetivo primordial es obtener los datos de las fuentes transaccionales necesarias y ajustarlos para los demás subprocesos de ETL. Para ello se deben determinar las mejores fuentes de información, las de mejor calidad. Para tal finalidad, se debe analizar las fuentes disponibles y escoger aquellas que sean mejores.

Limpieza

Este proceso obtiene los datos originales y verifica la calidad de los datos, elimina datos repetidos, corrige los datos incorrectos y completa los valores vacíos, es decir se transforman los datos -siempre que sea posible- para reducir los errores de carga. En este momento se cuentan con datos sin errores y confiables.

Las fuentes transaccionales poseen datos que deben ser depurados y limpiados. Ciertos motivos que ocasionan que los datos estén “sucios” son:

- Valores por defecto.
- Campos vacíos.
- Campos con distintos usos.
- Uso inapropiado de los campos.
- Re utilización de claves primarias.
- Selección del primer valor de una lista.
- Problemas de carga de antiguos sistemas o de integración entre sistemas.

La limpieza de datos se divide en distintas etapas:

- **Depurar los valores** (*parsing*): busca e identifica los datos individuales de información de las fuentes origen. Ej.: separa el nombre completo en: nombre, primer apellido, segundo apellido; o la dirección en: calle, número, etc.
- **Corregir** (*correcting*): modifica los valores de los campos utilizando algoritmos de corrección y orígenes externos de datos. Ej.: verifica que el código postal se corresponda con la dirección.
- **Estandarizar** (*standardizing*): aplica estrategias para la transformación de los valores en formatos consistentes. Ej.: cambiar los términos Sr. o Sra. por sus correspondientes nombres completos.
- **Relacionar** (*matching*): busca eliminar duplicados relacionando los valores de registros, mediante la corrección y la estandarización. Ej.: identificación de nombres, apellidos o direcciones semejantes.

Transformación

La transformación inicia una vez que los datos se encuentren limpios, los datos se transforman según las necesidades y las reglas del negocio. Como resultado se obtienen datos limpios, sólidos, fiables, resumidos y útiles.

La transformación incluye:

- Cambios de formato.
- Sustitución de códigos.
- Valores derivados y agregados.

Los agregados como sumas de las ventas generalmente se realizan mediante un cálculo previo y luego son almacenados para conseguir rendimientos más óptimos. En este proceso también se ajusta el nivel de granularidad o detalle, por ejemplo: se puede tener detalles a nivel de líneas de factura en los datos extraídos, pero en el datawarehouse lo que se almacena son las ventas semanales o mensuales. La diferencia del nivel de detalle en el análisis es lo que se denomina granularidad.

Integración

La integración valida que los datos cargados en el datawarehouse sean consistentes con las definiciones y formatos especificados; los integra en los distintos modelos de las distintas áreas de negocio. Estos procesos pueden ser complejos.

Actualización

Este proceso es el que permite agregar los nuevos datos al datawarehouse, determina la frecuencia con el que se realizarán nuevas cargas de datos al datawarehouse.

3.3.3. Datawarehouse

El datawarehouse abreviado como DW o “almacén de datos” son conjuntos de datos acerca de las operaciones de una organización, brindan una visión total, común e integrada de los datos. Hoy en día se piensa que son las principales tecnologías que apoyan el entorno heterogéneo de toma de decisiones, el datawarehouse tiene las siguientes propiedades: no volátil, coherente, fiable y con información histórica [Inm92, CC10].

El profesor Hugh J. Watson [Wat06] lo define como:

“Un datawarehouse es una colección de información creada para soportar las aplicaciones de toma de decisiones. Datawarehousing es el proceso completo de extraer información, transformarla y cargarla en un datawarehouse y el acceso a esta información por los usuarios finales y las aplicaciones.”

Bill Inmon [Inm92] definió las características que debe cumplir un datawarehouse:

- Orientado a un área: cada parte del datawarehouse está construida para resolver un problema de negocio. Ej.: entender los hábitos de compra de clientes, analizar la calidad de los productos, analizar la productividad de una línea de fabricación.
- Integrado: la información debe ser transformada en medidas, códigos y formatos comunes para que sea de utilidad. Ej.: la moneda en que están expresadas los importes es común.
- Indexado en el tiempo: se mantiene la información histórica. Ej.: analizar la evolución de las ventas en los periodos deseados.
- No volátil: los usuarios no la mantienen como lo harían en los entornos transaccionales. No se ve actualizado continuamente, sino periódicamente de forma preestablecida. La información se almacena para la toma de decisiones

Ralph Kimbal [Kim92] define los objetivos que debería cumplir un datawarehouse:

- El alcance de un datawarehouse puede ser a nivel de departamento o corporativo.
- El datawarehouse no es sólo información sino también las herramientas de consulta, análisis y presentación de la información.
- La información del datawarehouse es consistente.
- La calidad de información en el datawarehouse es el motor de business reengineering.

Se debe recalcar que existen otros varios elementos relacionados a un datawarehouse [CC10]:

- **Datawarehousing:** consiste en extraer y filtrar los datos de las operaciones transaccionales provenientes de las distintas fuentes de información o sistemas externos para transformar, integrar y almacenar en el almacén de datos, con el objetivo de acceder a ellos para brindar soporte en la toma de decisiones.
- **Data Mart:** representa un conjunto parcial de datos del datawarehouse que permite el análisis de un determinado proceso.
- **Operational Data Store (ODS):** es un repositorio de datos temporal que contiene solamente los últimos valores de los datos extraídos.
- **Staging Área:** es un sistema intermedio existente entre las fuentes de origen de los datos y el datawarehouse con el objetivo de:
 - Agilizar la extracción de los datos de las fuentes de origen heterogéneas y complejas.
 - Mejorar la calidad de los datos.
 - Ser utilizado como caché de datos con el que luego se realiza el proceso de datawarehousing.
 - Ser utilizada para el acceso a información más detallada no contenida en el datawarehouse.
- **Procesos ETL:** es la herramienta de integración de datos que consolida los datos de las distintas fuentes de origen, para luego almacenarlos en el datawarehouse, data mart, staging área y ODS.
- **Metadatos:** es la representación estructurada de los datos que detallan características de las instancias aportando informaciones para identificar y administrar dichas instancias.

Elementos de un datawarehouse

Una base de datos con una estructura relacional se rige por las “formas normales” [KH03]. Un datawarehouse debe dejar de lado dicho patrón de diseño. La idea principal es que los datos almacenados en el datawarehouse estén desnormalizadas, con el fin de agilizar las consultas. Para ello se debe identificar los procesos, las vistas y las medidas cuantificables del negocio que están relacionados. Según [CC10] los elementos de un datawarehouse son:

- **Tabla de hecho:** es la representación de un proceso de negocio en el datawarehouse. En el diseño se representa como una tabla que permite guardar dos tipos de atributos diferenciados:
 - Medidas del proceso de trabajo que se pretende modelar.
 - Claves foráneas hacia registros de una tabla de dimensión.
- **Dimensión:** se corresponde con una vista del proceso de negocio en el datawarehouse.

- **Métrica:** son los indicadores de un proceso de negocio que permiten medir o cuantificar un proceso.

Tipos de esquemas para estructurar los datos en un datawarehouse

- **Esquema en estrella:** consiste en que la información sea estructurada en procesos, vistas y métricas semejante a una estrella. En el diseño, se tiene una tabla de hechos que se encuentra en el centro y es el objeto de análisis, y rodeada por una o varias tablas dimensionales que representan cada una de las vistas que interactúan en la descripción del hecho.
- **Esquema en copo de nieve:** es un esquema derivado del esquema estrella, en el que las tablas de dimensión son normalizadas en múltiples tablas. Por este motivo la tabla de hechos ya no es la única tabla en relacionarse con otras tablas.

3.3.4. OLAP

Existen múltiples tecnologías que permiten analizar la información almacenada en un datawarehouse, uno de los más importantes es OLAP (*Online Analytical Processing*), esta tecnología permite realizar un análisis multidimensional de un hecho, desde distintas perspectivas o dimensiones mediante consultas complejas que van desde pocas hasta docenas de operaciones de unión, filtrado, agrupación y agregación. El principal objetivo es agilizar la consulta de grandes cantidades de datos [Wre06][Can07].

Una definición formal de OLAP sería [CC10]:

“Se entiende por OLAP o proceso analítico en línea, al método ágil y flexible para organizar datos, especialmente metadatos, sobre un objeto o jerarquía de objetos como en un sistema u organización multidimensional, y cuyo objetivo es recuperar y manipular datos y combinaciones de los mismos a través de consultas o incluso informes”

En la figura 3.2 se observa una representación gráfica del OLAP también conocida como cubo. En el ejemplo el cubo tiene 3 dimensiones (Tiempo, Producto y Región) sobre las cuales se pueden realizar consultas, ej.: Monto total de venta del producto “A” en el año “B” en la región “C”. Los cubos OLAP también permiten representar jerarquías, en el caso de la dimensión Tiempo la jerarquía podría estar compuesta por año, semestre, mes, semana y día.



Figura 3.2: Cubo OLAP

Existen distintos tipos de OLAP, las cuales difieren principalmente en la forma de guardar los datos:

- MOLAP (Multidimensional OLAP): es la forma tradicional del OLAP, accede directamente sobre una base de datos multidimensional, que utiliza estructuras de datos optimizadas para la recuperación de los mismos, es eficaz en los tiempos de respuestas de las consultas.
- ROLAP (Relational OLAP): accede directamente a las bases de datos relacionales que almacenan los datos base y las tablas dimensionales como tablas relacionadas.
- HOLAP (Hybrid OLAP): es una combinación de las dos anteriores, permite almacenar parte de los datos en una base de datos multidimensional y otra parte en una relacional. En la base de datos relacional se guardan grandes cantidades de información detallada, mientras que en la multidimensional se almacenan datos menos detallados o agregados.

3.3.5. Herramientas de BI

Las principales herramientas de Business Intelligence son [EH05]:

- *Generadores de informes*: La utilizan generalmente los desarrolladores expertos para la creación de informes estándar para departamentos o grupos de una organización.
- *Herramientas de usuario final de consultas e informes*: Utilizada por los propios usuarios finales que crean sus propios reportes y donde no se requiere experiencia en programación.
- *Herramientas OLAP*: Permiten un análisis multidimensional de la información desde distintas perspectivas y periodos de tiempo.

- *Herramientas de Dashboard y Scorecard*: Permite visualizar la información crítica sobre el rendimiento con un vistazo general mediante gráficos y con opción de visualizar información detallada para un mejor análisis.
- *Herramientas de planificación, modelización y consolidación*: Posibilita a los usuarios generar planes y simulaciones con la información de BI. Proveen a los dashboards y scorecards las métricas con los umbrales y objetivos.
- *Herramientas datamining*: Utilizado por los analistas para crear modelos estadísticos de las actividades. Consiste en un proceso para descubrir patrones desconocidos contenidos en la información con los cuales es posible resolver problemas del negocio.

3.4. Indicadores Clave de Rendimiento

Los KPI o Indicadores Clave de Rendimiento se tratan de indicadores que son decisivos para analizar de forma rápida la situación del negocio y facilitar la toma de decisiones. Una característica es que todos los KPI son indicadores, pero no todos los indicadores son KPI. Otra característica es que cada organización debe definir sus propios KPI según la actividad realizada, el tipo de producto o la estrategia de negocios, no pueden copiarse ya que cada organización es diferente y requiere de una reflexión estratégica de donde saldrán los KPI. Otro elemento importante de los KPI es que pueden no ser exclusivamente del tipo financiero, pueden medir otros valores como por ejemplo: calidad del servicio. Por último los KPI son un elemento importante en la estrategia de negocios por su capacidad de comunicar resultados a las personas que forman parte del proyecto. [Alv13]

Un cuadro de gestión o de mando no debe excederse en la cantidad de KPI, porque puede darse el problema de “la parálisis por el análisis”, que ocurre cuando se pasa de no tener ninguna información a contar con decenas de indicadores, y una de las características del entorno competitivo actual es que se deben tomar decisiones de forma rápida y antes de que lo hagan los demás competidores [Alv13].

En el capítulo 5 haremos uso de los conceptos, herramientas y tecnologías que nos provee Business Intelligence para obtener los datos que ayudarán a desarrollar una solución al problema de estudio sobre pronóstico de la demanda, iniciaremos con una breve descripción de la fuente de información, el proceso ETL (Extracción, Transformación y Carga), la especificación del datawarehouse, la definición de los indicadores clave de rendimiento y finalmente el etiquetado a cada tupla de indicadores que se transformarán en datos de entrada para el proceso de aprendizaje automático.

3.5. Aplicaciones de BI

Los sistemas de Business Intelligence abarcan un grupo cada vez mayor de usuarios como especialistas en tareas de control, información financiera, personal de ventas, directivos y gerentes

[RGS02]. En el cuadro 3.1 se puede ver algunas de las áreas en las que se pueden implementar sistemas de Business Intelligence, en la primera columna se citan las áreas que utilizan con mayor frecuencia como: compañías comerciales, compañías de seguros, entidades financieras, telecomunicaciones y empresas de manufactura; en la segunda columna se exponen algunos casos de aplicación como: análisis de ventas, planificación de pedidos, análisis de clientes, gestión de créditos, etc.

Tabla 3.1: Áreas de aplicación de BI

<i>Áreas de aplicación de BI</i>	<i>Casos de uso</i>
Empresas Retail	<ul style="list-style-type: none"> ■ Proporcionar un análisis de las transacciones de los clientes. Determinar los productos más vendidos, promociones, hábitos de compras. ■ Pronóstico. Uso de datos históricos para pronosticar la demanda y basarse en el pronóstico para definir los requisitos de inventario con mayor precisión.
Inventario	<ul style="list-style-type: none"> ■ Planificación de Inventarios. Ayudar a identificar el nivel de inventario necesario para garantizar la demanda de los clientes.
Gestión de Pedidos	<ul style="list-style-type: none"> ■ Pedido y reposición. Uso de la información para tomar decisiones más rápidas sobre los artículos a pedir y determinar cantidades óptimas.
Contabilidad	<ul style="list-style-type: none"> ■ El uso de datos contables permite una mejor oportunidad de administración del riesgo, agilizar las operaciones, identificar ahorros de costos y oportunidades estratégicas de ingresos.

Bancos, Financieras y Valores	<ul style="list-style-type: none"> ■ Análisis de rentabilidad del cliente: Analizar la rentabilidad global del cliente, a corto y largo plazo, proporcionar la base para las ventas de alta rentabilidad y la banca de relación, maximizar las ventas a clientes de alto valor, reducir los costos para los clientes de bajo valor, maximizar la rentabilidad de nuevos productos y servicios. ■ Gestión de créditos: Establecer patrones de progresión de problemas de crédito por clase y tipo de clientes, alertar a los clientes para evitar problemas de crédito, gestionar límites de crédito, evaluar la cartera crediticia del banco, reducir pérdidas crediticias. ■ Atención en sucursales: Mejorar el servicio y la atención al cliente, aperturas de nuevas cuentas, fortalecer la lealtad del cliente.
Telecomunicaciones	<ul style="list-style-type: none"> ■ Perfil y segmentación de clientes. Determinar perfiles de productos de alto beneficio y segmentos de clientes, proporcionar perfiles de clientes detallados e integrados, desarrollar programas individualizados de llamadas frecuentes, determinar futuras necesidades de los clientes. ■ Previsión de la demanda del cliente. Prever las necesidades futuras del producto o actividad de servicio, proporcionar una base para el análisis y control de la rotación para mejorar la retención de clientes.
Transporte	<ul style="list-style-type: none"> ■ Aerolíneas. Analizar conductas, tarifas abonadas, respuestas a promociones, canje de millas, categorías de los pasajeros frecuentes de la empresa.
Educación	<ul style="list-style-type: none"> ■ Universidades y Colegios. Análisis del proceso de admisión de alumnos, registros de cursos y asistencia estudiantil, monitoreo del progreso de los alumnos.

Salud	<ul style="list-style-type: none"> ■ Analizar los resultados, identificar tendencias, detectar patrones y predecir los resultados para mejorar el desempeño clínico y operacional. ■ Monitorear iniciativas de calidad y programas de atención. ■ Seguimiento y monitoreo de ingresos, márgenes y rendimiento operacional.
-------	---

3.6. Resumen

En el capítulo se introdujeron los conceptos más relevantes de Business Intelligence. El término acuñado en 1989 por Howard Dressner consultor de Gartner Group lo describe como un conjunto de conceptos y métodos para transformar los datos empresariales en información significativa y útil para mejorar las decisiones. Se realiza una breve introducción de los principales componentes de Business Intelligence citadas a continuación:

- Fuentes de información.
- Extracción, transformación y carga.
- Datawarehouse.
- Análisis OLAP, y
- Herramientas de visualización.

Los Indicadores Claves de Rendimiento, como lo dice el nombre se tratan de indicadores que son decisivos y brindan un análisis rápido de la situación actual del negocio para facilitar la toma de decisiones. Son un elemento importante en la estrategia de negocios por su capacidad de comunicar resultados a las personas que forman parte del proyecto.

Por último, en el cuadro 3.1 se resumen algunas de las áreas de mayor aplicación de Business Intelligence, entre las que cabe destacar: salud, educación, transporte, telecomunicaciones, bancos, financieras, inventario, empresas retail.

Capítulo 4

Machine Learning

El objetivo de este capítulo es visualizar qué aspectos de Machine Learning fueron tomados como componentes de solución al problema de estudio.

En lo que va de la Edad Contemporánea, el hombre siente fuertemente la necesidad de encontrar respuestas a ciertos aspectos de sus propias capacidades, respuestas de cómo funciona el cerebro humano, cómo se van originando e hilando los pensamientos, cómo se va adquiriendo el conocimiento, cómo la racionalidad está presente en las decisiones humanas, cuál es el mecanismo de toma de decisiones de la mente humana, cómo utiliza su inteligencia para resolver problemas e ideas abstractas, cuál es el proceso que sigue la mente para poder aprender, interrogantes sobre la memoria del cerebro y cómo los sentidos van alimentando de percepciones para tener una visión única del universo. La ciencia de la computación ha abierto la puerta a dichas respuestas.

En 1950 Alan Turing dio uno de los saltos más importantes al proponer el enfoque de su “Prueba de Turing”, donde un computador es considerado como agente inteligente si un evaluador humano, sin interactuar con el computador, realiza preguntas y no es capaz de distinguir si las respuestas vienen de una persona o del computador. En su artículo “Computing Machinery and Intelligence” [Tur50] expuso este y otros conceptos y como tal es considerado padre de la Inteligencia Artificial (IA).

Aunque hoy en día aún es válida la Prueba de Turing, el desafío no está en construir un agente inteligente total, sino mas bien buscar que un computador tenga capacidades como [RN04]:

1. Procesamiento del lenguaje natural
2. Representación del conocimiento
3. Razonamiento automático
4. Aprendizaje automático (Machine Learning)
5. Visión computacional
6. Robótica

Entre los precursores de IA están John McCarthy, Marvin Minsky, Allen Newell y Herbert Simon, todos ganadores del premio “ACM A.M. Turing Award” [AWA] por sus notables aportes en los inicios de IA. McCarthy acuñó el término Artificial Intelligence en el taller “Dartmouth Summer Research Project on Artificial Intelligence” que organizó y se desarrolló en junio del año 1956.

En palabras transcritas, el taller perseguía el siguiente objetivo: *“El estudio debe proceder sobre la base de la conjetura de que cada aspecto del aprendizaje o cualquier otra característica de la inteligencia puede, en principio, describirse tan precisamente que se puede hacer que una máquina lo simule. Se intentará encontrar cómo hacer que las máquinas usen el lenguaje, formen abstracciones y conceptos, resuelvan tipos de problemas ahora reservados para los humanos y se mejoren a sí mismos. Creemos que se puede lograr un avance significativo en uno o más de estos problemas si un grupo de científicos cuidadosamente seleccionados trabajan juntos durante un verano”*.

Entre los aportes más destacados de McCarthy se pueden mencionar la definición del lenguaje de alto nivel Lisp, la creación del tiempo compartido y el artículo “Programs with Common Sense” [McC59] donde define su *Generador de Consejos*, todos estos aportes desarrollados en 1958. Minsky diseñó SNARC (Stochastic Neural Analog Reinforcement Calculator) la primera máquina a partir de una red neuronal en el año 1951, junto a Seymour Papert publicó el libro “Perceptrons: an introduction to computational geometry” [MP87] en el año 1969 y también creó un modelo de redes semánticas denominadas “marcos” publicado en “A framework for representing knowledge” en el año 1974 [Min74].

En setiembre de 1956 en el “MIT Symposium on Information Theory” se mostraron trabajos muy importantes como el de George Miller que presentó “The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information”, Noam Chomsky presentó “Three Models for the Description of Language”, y Allen Newell y Herbert Simon presentaron “The Logic Theory Machine”: programa de computador capaz de hacer la demostración de un teorema y considerado el primer programa de IA.

El campo de la ciencia cognitiva echó sus raíces en este simposio evidenciando que los modelos informáticos se pueden utilizar para modelar la psicología de la memoria, el lenguaje y el pensamiento lógico. En 1959 presentaron el programa “General Problem Solver” (GPS), pretendía funcionar como una máquina universal para resolver problemas y fue el primer programa de computador que separó su conocimiento de los problemas (reglas representadas como datos de entrada) de su estrategia de cómo resolver problemas (un motor de resolución genérico).

Machine Learning como subcampo de IA está presente en muchas aplicaciones de la vida real, especialmente en aquellas donde se requiere el procesamiento de grandes cantidades de datos. Por esta razón la tecnología de la información la toma como aliado esencial. Muchos avances tecnológicos de última generación utilizan algoritmos de Machine Learning para realizar un análisis inteligente de los datos [SV08]. Entre algunas aplicaciones conocidas se destacan:

- Reconocimiento facial de Facebook: técnicas de Computer Vision and Pattern Recogni-

tion. <https://research.fb.com/learning-to-segment/>

- Kinect para Xbox 360: técnicas de reconocimiento de voz y reconocimiento facial para la identificación automática de los usuarios.
- Voice recognition.
- Campo relacionado del reconocimiento de caracteres manuscritos.
- Motores de búsqueda.
- Recommender system en plataformas como Amazon, Netflix, Facebook.
- Reconocimiento automático de ciertas áreas en el mundo realizado por satélites.

Para dimensionar el gran crecimiento de los datos digitales en el mundo, en un estudio publicado por la International Data Corporation (IDC) y patrocinado por Dell EMC que se denomina: “The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East” [GR12] se analizó qué tan rápido crecerían los datos cada año. La medida incluyeron todos los datos digitales creados, replicados o consumidos en un solo año. En la Figura 4.1 se muestra una proyección del tamaño hasta el 2020.

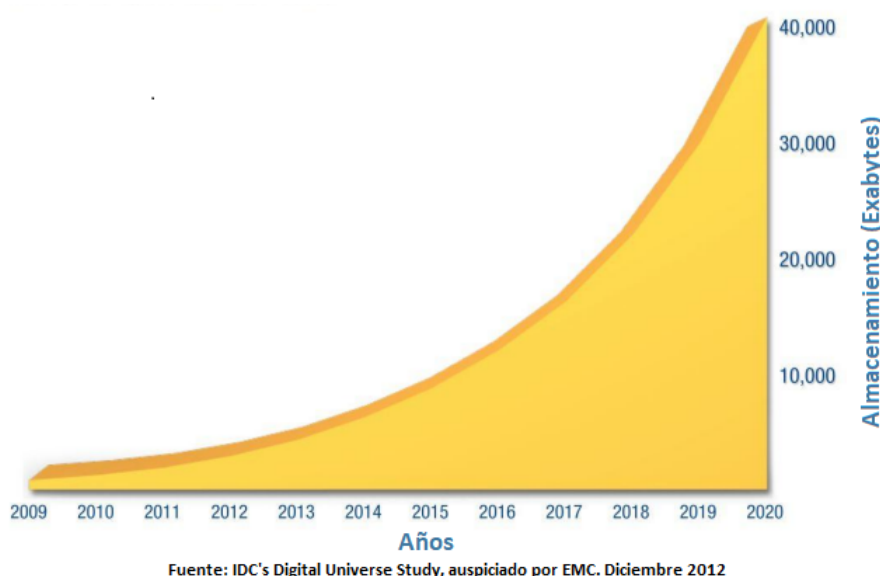


Figura 4.1: Proyección de crecimiento de datos del 2005 al 2020.

Los datos incluían: videos de YouTube, películas, datos bancarios, imágenes de cámaras de seguridad en aeropuertos, datos de las colisiones subatómicas del Gran Colisionador de Hadrones en el CERN, llamadas telefónicas, mensajes de texto, etc. Los resultados indicaron que en el 2005 existían 130 exabytes en datos digitales, en el 2010 llegó a 1200 exabytes, en el 2015 a unos 7900 exabytes y para el 2020 se pronostica que llegará a los 40000 exabytes. IDC estima que para el 2020 hasta el 33 % de los datos digitales contendrán información valiosa.

4.1. Definición

El término Machine Learning fue propuesto por Arthur Samuel en el año 1959 mientras trabajaba para IBM. En el IBM Journal publicó “Some Studies in Machine Learning Using the Game of Checker” en el cual escribió: *“Programming computers to learn from experience should eventually eliminate the need for much of this detailed programming effort”* [Sam59] por lo que de su artículo se interpreta que *“Machine Learning es un campo de estudio que da a las computadoras la capacidad de aprender a resolver problemas sin ser explícitamente programadas”*.

Samuel desarrolló en el año 1952 un programa que aprendió a jugar Damas, hasta llegar a una categoría amateur. Este pionero ya demostraba que los programas a través del aprendizaje pueden efectuar tareas de toma de decisiones sin ser programadas explícitamente dichas decisiones. Cabe destacar que Arthur Samuel fue uno de los asistentes del “Dartmouth Summer Research Project on Artificial Intelligence”.

Otro investigador, Tom Mitchell propuso en 1998 la siguiente definición: *“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E ”*. Donde se nos indica que el aprendizaje en las máquinas deberá ser parecido al aprendizaje en los humanos, por ejemplo cuando una criatura comienza a hablar a través de la experiencia de pronunciar las palabras y de su interacción con otras personas, entonces sucede que su capacidad de hablar se va perfeccionando o mejorando [Mit97].

Otra definición es *“The purpose of machine learning is to learn from training data in order to make as good as possible predictions on new, unseen, data”*[Pug16]. La dificultad radica en que debemos construir modelos que nos acerquen a una buena predicción sobre datos aún no conocidos o imprevistos.

En el contexto de IA y según el enfoque propuesto en el libro “Inteligencia Artificial Un Enfoque Moderno” [RN04], se conciben los sistemas inteligentes como agentes racionales dotados de capacidades específicas. Entre las capacidades principales están el poder percibir el entorno de trabajo (“problemas para los cuales fueron hechos”) con la ayuda de sus sensores, actuar en ese entorno mediante sus actuadores y contar con una medida de rendimiento para medir el éxito.

Estructuralmente un agente = arquitectura + programa. Los sensores y actuadores forman parte de la arquitectura, la arquitectura es el medio físico que puede ser una PC o un robot con sensores, etc. El objetivo principal de la IA es construir el programa del agente, que es donde se implementa la función del agente. Para que el agente sea considerado un agente que aprende, el programa debe contener el “Elemento de aprendizaje”. En la Figura 4.2 se grafica el concepto, un modelo de agente inteligente que aprende.

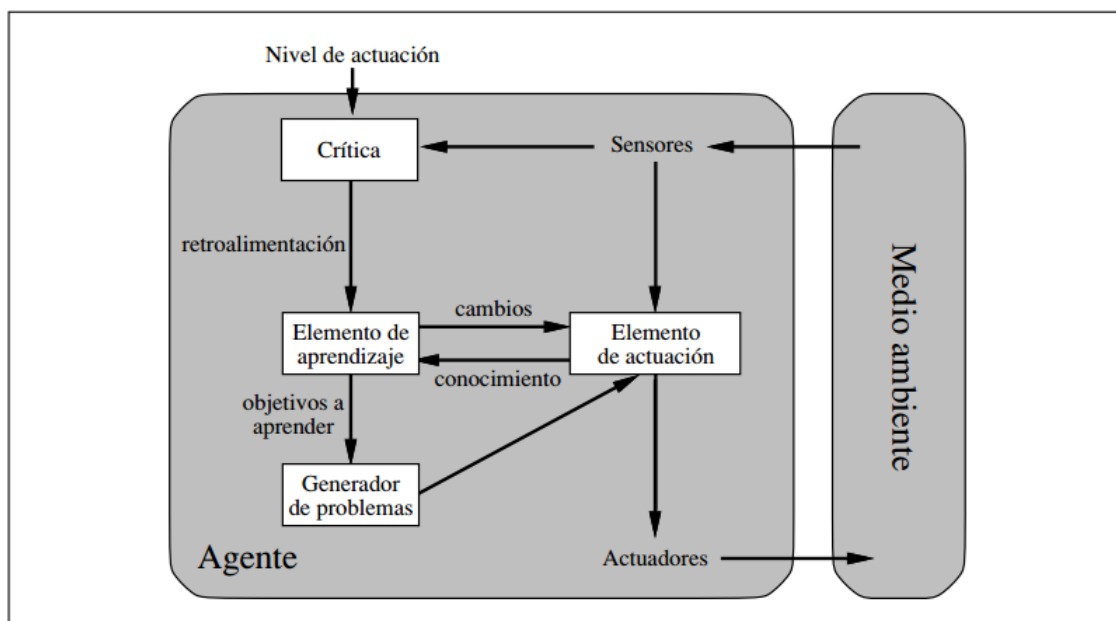


Figura 4.2: Modelo general para agentes que aprenden [RN04].

4.2. Formas de Aprendizaje

Los algoritmos de aprendizaje automático se pueden agrupar según la forma en que se realiza el aprendizaje, de acuerdo a la información que poseen o que pueden llegar a poseer. Para entender la naturaleza del problema es vital conocer el tipo de retroalimentación que se dispone para el aprendizaje. Hay tres tipos distintos de aprendizaje: supervisado, no supervisado y por refuerzo [RN04].

4.2.1. Aprendizaje supervisado

Los agentes inteligentes que implementan algoritmos de aprendizaje supervisado (*Supervised Learning*), tienen como objetivo aprender una función de hipótesis h que se aproxime a la función verdadera f y que es solución al problema que se intenta resolver. Lo que se conoce de f son los llamados “ejemplos”, que son puntos concretos dentro de la desconocida línea que representa la función verdadera f . En otras palabras, la definición matemática de la función f no se conoce, solamente se conoce un conjunto de ejemplos definidos por valores correctos de entradas y salidas de la función f .

Otra forma de verlo es de la siguiente manera, se dispone de variables de entrada (X) y una variable de salida (Y), mediante un algoritmo y a partir de un conjunto de datos de entrenamiento se busca aprender una función $Y = f(X)$ que mapee la salida desde la entrada. Los datos de entrenamiento constituyen las respuestas correctas conocidas o “ejemplos”. El objetivo esencial es aproximar la función de mapeo lo mejor que se pueda, de manera que cuando se tengan nuevos datos de entrada (X) se pueda predecir la variable de salida (Y) [CCCC08].

Problemas del tipo clasificación o regresión se resuelven típicamente con aprendizaje supervisado. Un problema de clasificación es cuando la variable de salida es una categoría y un problema de regresión es cuando la variable de salida es un valor real [JWHT14]. Como ejemplo de aplicaciones que implementan esta técnica están los vehículos autoconducidos, que deben aprender a diferenciar una calle de la que no es (salida booleana: es calle o no es calle?), también debe aprender a frenar (salida booleana: frenar o no frenar?), etc. Entre los algoritmos que implementan aprendizaje supervisado están: Linear Regression, Decision Trees, Support Vector Machines (SVM), Bayesian Networks, etc [Kot07]. El problema de estudio de la tesis utiliza algoritmos de aprendizaje supervisado.

4.2.2. Aprendizaje no supervisado

En los problemas de aprendizaje no supervisado (*Unsupervised Learning*) los algoritmos reciben como entrada datos de entrenamiento (ejemplos) que no tienen respuestas correctas conocidas. Solo se dispone de variables de entrada (X) pero no hay correspondiente variable de salida (Y). Los algoritmos buscan estructuras presentes y como resultado pueden extraer reglas generales, o reducir sistemáticamente la redundancia, u organizar los datos por similitud. Los problemas de agrupación (clustering) y asociación (association) se resuelven normalmente con algoritmos de aprendizaje no supervisados.

En empresas retail buscan descubrir grupos de clientes que tienen el mismo comportamiento en las compras (los que toman cerveza por ejemplo), lo que representa una situación de agrupación inherente en los datos. Cuando buscan descubrir reglas que se deduzcan del conjunto de datos (por ejemplo, los que toman vino compran agua tónica o mineral), se tiene una situación de asociación inherente en los datos. Entre las aplicaciones que implementan esta técnica está el reconocimiento de dígitos escritos a mano. Entre los algoritmos de aprendizaje no supervisado están: k-Means, Principal Component Analysis (PCA), reducción de la dimensionalidad, etc [Kot07].

4.2.3. Aprendizaje por refuerzo

El aprendizaje por refuerzo (*Reinforcement Learning*) es el más general entre las tres categorías. En vez de que un instructor indique al agente qué hacer, el agente inteligente debe aprender cómo se comporta el entorno mediante recompensas (*refuerzos*) o castigos, derivados del éxito o del fracaso respectivamente. El objetivo principal es aprender la función de valor que le ayude al agente inteligente a maximizar la señal de recompensa y así optimizar sus políticas de modo a comprender el comportamiento del entorno y a tomar buenas decisiones para el logro de sus objetivos formales.

Los principales algoritmos de aprendizaje por refuerzo se desarrollan dentro de los métodos de resolución de problemas de decisión finitos de Markov, que incorporan las ecuaciones de Bellman y las funciones de valor. Los tres métodos principales son: la Programación Dinámica

(*Dynamic Programming o DP*), los métodos de Monte Carlo y el aprendizaje de Diferencias Temporales (*Temporal-Difference Learning o TD*) [SB98b].

Entre las implementaciones desarrolladas está AlphaGo, un programa de IA desarrollado por Google DeepMind para jugar el juego de mesa Go. En marzo de 2016 AlphaGo le ganó una partida al jugador profesional Lee Se-Dol que tiene la categoría noveno dan y 18 títulos mundiales. Entre los algoritmos que utiliza se encuentra el árbol de búsqueda Monte Carlo, también utiliza aprendizaje profundo con redes neuronales. Puede ver lo ocurrido en el documental de Netflix “AlphaGo” <https://www.netflix.com/title/80190844>.

4.3. Algoritmos de aprendizaje supervisado y no supervisado

El Dr. Jason Brownlee es un especialista en aprendizaje automático, desarrollador, escritor y empresario. Ha trabajado en sistemas de aprendizaje automático para la defensa, startups y pronósticos meteorológicos. Tiene una comunidad en <https://machinelearningmastery.com/>, la cual empezó porque le apasiona ayudar a los desarrolladores profesionales a comenzar y aplicar con confianza Machine Learning que les permita resolver problemas complejos. Los algoritmos de aprendizaje automático se pueden agrupar según la similaridad en términos de su forma o función, como por ejemplo los métodos basados en árboles y los métodos inspirados en redes neuronales. Se muestra en la Figura 4.3 lo propuesto por Brownlee [REF].

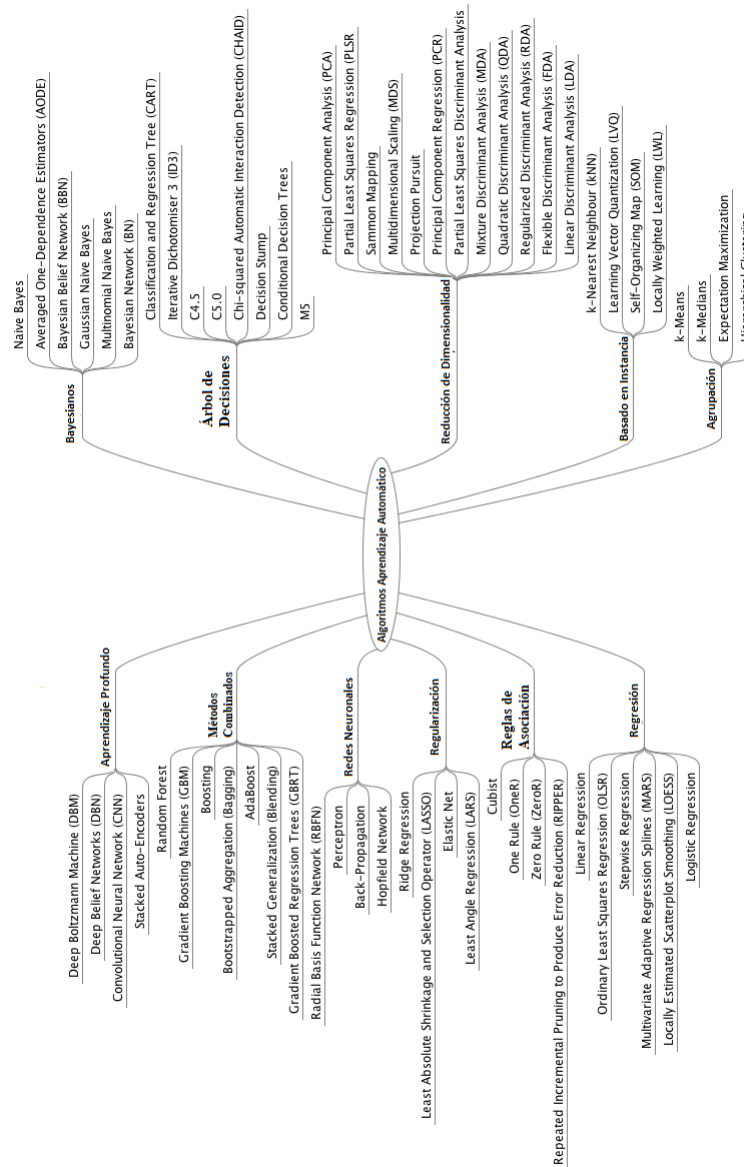


Figura 4.3: Agrupación de algoritmos según Jason Brownlee [\[REF\]](#).

No hay un consenso general de cómo agrupar los algoritmos de Machine Learning en términos de su función o de cómo trabajan. La Figura 4.3 muestra un método útil de agrupación, que no es perfecto y ni exhaustivo en los grupos y algoritmos. Hay algoritmos que pueden encajar en varias categorías como Learning Vector Quantization que es a la vez un método inspirado en una red neuronal y un método basado en instancia. También hay categorías que tienen el mismo nombre que describen el problema y la clase de algoritmo como Regression y Clustering. Se podría manejar estos casos listando los algoritmos dos veces o insertando en el grupo al que subjetivamente se ajusta mejor. Se utiliza este último enfoque de no duplicar algoritmos.

4.3.1. Algoritmos de regresión

Los Algoritmos de Regresión (*Regression Algorithms*) reciben como datos de entrada un conjunto de datos llamados instancias o ejemplos. Cada instancia puede verse como un registro

de base de datos, donde uno de los campos es la variable dependiente y las demás variables independientes. Estos algoritmos buscan la función que defina lo más adecuadamente la relación existente entre la variable dependiente y las variables independientes. Si la variable dependiente es un valor discreto, entonces se trata de un problema de clasificación, si es un valor continuo se trata de un problema de regresión [Iko12]. Estos algoritmos se implementan en problemas de aprendizaje supervisado.

4.3.2. Algoritmos basados en instancia

El modelo de aprendizaje Basados en Instancia (*Instance-Based Learning*) se basa en un conjunto de instancias que pertenecen a la descripción del concepto y que son almacenadas y actualizadas constantemente. A la descripción de concepto se suman dos funciones principales: la función de similitud y la función de clasificación. Por cada instancia de entrenamiento la función de similitud calcula la similitud que tiene con las instancias de la descripción del concepto. La función de clasificación realiza la predicción basada en la función de similitud y en ciertas propiedades de la descripción del concepto [AKA91]. Estos algoritmos se implementan en problemas de aprendizaje supervisado.

4.3.3. Algoritmos de regularización

Los Algoritmos de Regularización (*Regularization*) ayudan a evitar modelos sobreajustados agregando penalizaciones a los parámetros. Entre los métodos más destacados se encuentran: regularización L1 (regularización LASSO), regularización L2 (regularización de Tíjonov) y una versión combinada de ambos [DKKAK11]. Estos algoritmos se implementan en problemas de aprendizaje supervisado y no supervisado.

4.3.4. Algoritmos de árbol de decisiones

Los métodos de Árbol de Decisiones (*Decision Tree*) permiten “Aproximar funciones objetivo de valores discretos, en el que la función aprendida se representa mediante un árbol de decisiones” [Mit97]. Estos algoritmos se implementan en problemas de aprendizaje supervisado.

4.3.5. Algoritmos bayesianos

Los métodos Bayesianos (*Bayesian*) modelan el aprendizaje en base al teorema de Bayes. Se considera una técnica indicada para problemas donde es obligatorio tener en cuenta la incertidumbre [Bar12]. Estos algoritmos se implementan en problemas de aprendizaje supervisado.

4.3.6. Algoritmos de agrupación

Los métodos de Agrupación (*Clustering*) son técnicas que persiguen el objetivo de formar subconjuntos homogéneos a partir de un conjunto principal de datos y en base a determina-

das características [JMF99]. Estos algoritmos se implementan en problemas de aprendizaje no supervisado.

4.3.7. Algoritmos de reglas de asociación

Los algoritmos basados en Reglas de Asociación (*Association Rules*) construyen un conjunto de reglas del tipo *si \Rightarrow entonces*. Una regla $\{chorizo, costilla\} \Rightarrow \{carbón\}$ podría indicar que los clientes que compren chorizos y costilla de vaca eventualmente compren carbón para hacer asado. [BR18]. Estos algoritmos se implementan en problemas de aprendizaje no supervisado.

4.3.8. Algoritmos de redes neurales artificiales

Las Redes Neuronales artificiales (*Neural Networks*) son modelos basados en la abstracción de una red de neuronas del cerebro humano. “*El cerebro es una computadora altamente compleja, no lineal y paralela (sistema de procesamiento de información)*” [Hay11]. Estos algoritmos se implementan en problemas de aprendizaje supervisado, no supervisado y por refuerzo.

4.3.9. Algoritmos de aprendizaje profundo

Los algoritmos de Aprendizaje Profundo (*Deep Learning*) se construyen sobre Redes Neuronales con muchas capas ocultas [GBC16]. Para Quoc V. Le [Le15] son algoritmos que utilizan redes neuronales como arquitectura y que además son capaces de aprender automáticamente las características. Estos algoritmos se implementan en problemas de aprendizaje supervisado y no supervisado.

4.3.10. Algoritmos de reducción de dimensionalidad

La Reducción de la Dimensionalidad (*Dimensionality Reduction*) es aplicado a problemas donde se tiene una enorme cantidad de características en el modelo, de los cuales no todos son relevantes. Estos métodos proveen una manera eficiente de encontrar un grupo reducido de dimensiones determinantes [SVPM14]. Estos algoritmos se implementan en problemas de aprendizaje no supervisado.

4.3.11. Algoritmos combinados

Los métodos combinados (*Ensemble*) se basan en la idea de que varios clasificadores participan en el proceso de predicción sobre nuevos datos [Sew07]. Estos algoritmos se implementan en problemas de aprendizaje supervisado y no supervisado.

4.3.12. Algoritmos máquinas de soporte vectorial

Las máquinas de vectores de soporte (Support Vector Machines o SVM) son una serie de potentes algoritmos que resuelven problemas de regresión y clasificación, se basan en la construcción de hiperplanos n-dimensionales [Wan05]. Estos algoritmos se implementan en problemas de aprendizaje supervisado.

4.4. Algoritmos de aprendizaje por refuerzo

Se puede complementar el listado propuesto por el Dr. Jason con algunos de los algoritmos de aprendizaje por refuerzo.

4.4.1. Programación dinámica

Estos son los principales algoritmos [BBSE10]:

- Programación dinámica con Policy Iteration.
- Programación dinámica con Value Iteration.
- Programación dinámica con Policy search

4.4.2. Método de Monte Carlo

Tiene como principales algoritmos [HS16]:

- On-Policy Monte Carlo Control.
- Off-Policy Monte Carlo Control.

4.4.3. Aprendizaje por Diferencias Temporales

Sus principales algoritmos son [SB98b]:

- Sarsa: On-Policy TD Control.
- Q-learning: Off-Policy TD Control.
- Actor-Critic Methods.
- R-Learning for Undiscounted Continual Tasks.

4.5. Problemas de clasificación y regresión

4.5.1. Clasificación

4.5.1.1. Clasificación binaria

En su forma más simple se reduce a la siguiente cuestión: dado un patrón x extraído de un dominio X , estimar qué valor asumirá una variable aleatoria binaria asociada $y \in \{\pm 1\}$.

La clasificación binaria es probablemente el problema más estudiado en el aprendizaje automático y ha dado lugar a una gran cantidad de desarrollos algorítmicos y teóricos importantes durante el siglo pasado [SV08].

Por ejemplo, si se muestran imágenes de manzanas y naranjas, podemos indicar si el objeto en cuestión es una manzana o una naranja. Igualmente, si se quiere predecir si un propietario de vivienda podría incumplir su préstamo dado sus datos de ingresos y su historial de crédito, o si un correo electrónico determinado es spam o no.

4.5.1.2. Clasificación multiclase

Es la extensión lógica de la clasificación binaria. La principal diferencia es que ahora $y \in \{1, 2, 3, \dots, N\}$ puede asumir un rango de valores diferentes. El problema de estudio en cuestión utiliza clasificación multiclase, donde $y \in \{Nada, Poco, Medio, Mucho\}$. Por ejemplo, es posible que se desee clasificar un documento de acuerdo con el idioma en el que fue escrito (inglés, francés, alemán, español, hindú, japonés, chino, ...) [SV08].

La principal diferencia es que el costo del error puede depender en gran medida del tipo de error que se comete. Por ejemplo, en el problema de evaluar el riesgo de cáncer, hace una diferencia significativa si clasificamos erróneamente una etapa temprana del cáncer como saludable (en cuyo caso es probable que el paciente muera) o una etapa avanzada de cáncer (en cuyo caso es probable que el paciente sufra molestias por un tratamiento excesivamente agresivo).

Para desarrollar un modelo o esquema de Machine Learning para resolver problemas de clasificación multiclase, es necesario conocer los componentes esenciales que la forman [FHW11].

Ejemplos o instancias

La entrada de un esquema de aprendizaje automático es un conjunto de instancias. Estas instancias son las cosas que deben ser clasificadas, asociadas o agrupadas. En el escenario estándar, cada instancia es un ejemplo individual e independiente del concepto que se debe aprender.

Características o atributos

Las instancias son caracterizadas mediante los valores de un conjunto predeterminado de atributos. Cada instancia proporciona una entrada al aprendizaje automático y es caracterizado por los valores de un conjunto fijo y predefinido de características o atributos.

Etiquetas

Las cantidades nominales tienen valores que son símbolos distintos. Los valores mismos sirven como etiquetas o nombres, de ahí el término nominal, que viene de la palabra latina nombre. Los atributos nominales a veces se llaman categorizados, enumerados o discretos.

Conjunto de entrenamiento

El grupo de ejemplos utilizados en el proceso de entrenamiento de los algoritmos de aprendizaje automático constituyen el conjunto de entrenamiento.

Algoritmos de clasificación multiclase

Constituye el conjunto de algoritmos de machine learning que soportan problemas de clasificación multiclase. Cada algoritmo se construye con su Función Objetivo (f), Variables de entrada (X), Variable de salida (Y), donde $Y = f(X)$ [REF].

Conjunto de prueba

Para medir el rendimiento de un clasificador sobre nuevos datos, necesitamos evaluar su tasa de error en un conjunto de datos que no desempeñó ningún papel en la formación del clasificador. Este conjunto de datos independiente se denomina conjunto de prueba.

4.5.2. Regresión

Es otra aplicación prototípica. Aquí el objetivo es estimar una variable de valor real $y \in \mathbb{R}$ dado un patrón x . Por ejemplo, se podría querer estimar el valor de un stock al día siguiente, el rendimiento de un fabuloso semiconductor dado el proceso actual, el contenido de hierro de las mediciones de espectroscopia de masas dadas por el mineral o la frecuencia cardíaca de un atleta, dada la información del acelerómetro. Una de las cuestiones clave en las que los problemas de regresión difieren entre sí es la elección de una pérdida. Por ejemplo, al estimar los valores de stock, nuestra pérdida para una opción de venta será decididamente unilateral. Por otro lado, a un deportista aficionado solo le importaría que nuestra estimación de la frecuencia cardíaca coincidiera con la media real [REF].

4.6. Algoritmos de clasificación en Weka

Weka es una colección de algoritmos de aprendizaje automático para tareas de minería de datos. Los algoritmos pueden ser aplicados directamente a un conjunto de datos o llamados desde código Java. Weka contiene herramientas para pre-procesamiento de datos, clasificación, regresión, clustering, reglas de asociación y visualización. También es adecuado para desarrollar nuevos esquemas de aprendizaje automático [atUoW]. En el problema de estudio se utiliza el

conjunto de algoritmos de clasificación de Weka [FHWP16]. Los algoritmos de clasificación de Weka que se utilizarán son los siguientes [htt]:

Tabla 4.1: Clasificadores Weka

<i>Categoría del clasificador</i>	<i>Nombre del clasificador</i>	<i>Modelo, técnica o algoritmo que implementa</i>
Clasificadores bayesianos	BayesNet	Bayes Network (Red Bayesiana) [BFH ⁺ 16]
Clasificadores bayesianos	NaiveBayes	Naive Bayes [JL95]
Clasificadores bayesianos	NaiveBayesMultinomial	Naive Bayes multinomial [MN98]
Clasificadores bayesianos	NaiveBayesMultinomialUpdateable	Naive Bayes multinomial actualizable [MN98]
Clasificadores bayesianos	NaiveBayesUpdateable	Naive Bayes actualizable [JL95]
Basado en funciones	Logistic	Regresión Logística [lCvH92]
Basado en funciones	MultilayerPerceptron	Red Neuronal con <i>back propagation</i>
Basado en funciones	SimpleLogistic	Regresión Logística lineal con LogitBoost [LHF05] [SFH05]
Basado en funciones	SMO	Sequential Minimal Optimization con Support Vector [Pla98] [KSBM01] [HT98]
Clasificadores perezosos	IBk	K-nearest neighbours (K vecinos más cercanos) [AKA91]
Clasificadores perezosos	KStar	K* con función de distancia basada en entropía [CT95]
Clasificadores perezosos	LWL	Locally Weighted Learning (Aprendizaje Ponderado Localmente) [FHP03] [AMS96]
Meta algoritmos	AdaBoostM1	Adaboost M1 [FS96]
Meta algoritmos	AttributeSelectedClassifier	Selección de atributos
Meta algoritmos	Bagging	Bagging [Bre96]
Meta algoritmos	ClassificationViaRegression	Métodos de regresión [FWI ⁺ 98]

Meta algoritmos	CVParameterSelection	Selección de parámetros [Koh95b]
Meta algoritmos	FilteredClassifier	Filtro arbitrario
Meta algoritmos	IterativeClassifierOptimizer	Optimización del número de iteraciones
Meta algoritmos	LogitBoost	Regresión Logística aditiva [FHT98]
Meta algoritmos	MultiClassClassifier	Metaclasificador
Meta algoritmos	MultiClassClassifierUpdateable	Metaclasificador actualizable
Meta algoritmos	MultiScheme	Selección del clasificador
Meta algoritmos	RandomCommittee	Conjunto aleatorizado de clasificadores base
Meta algoritmos	RandomizableFilteredClassifier	Clasificador arbitrario con filtro arbitrario
Meta algoritmos	RandomSubSpace	Árbol de decisión [Ho98]
Meta algoritmos	Stacking	Combinación de clasificadores utilizando apilamiento [Wol92]
Meta algoritmos	Vote	Combinación de clasificadores [Kun04] [KHDM98]
Meta algoritmos	WeightedInstancesHandlerWrapper	Soporte de instancias ponderadas
Sistema de reglas	DecisionTable	Tabla de decisión simple [Koh95a]
Sistema de reglas	JRip	“Repeated Incremental Pruning to Produce Error Reduction” (RIPPER) [Coh95]
Sistema de reglas	OneR	Clasificador 1R [Hol93]
Sistema de reglas	PART	Divide y vencerás para construir un árbol de decisión C4.5 parcial [FW98]
Sistema de reglas	ZeroR	Clasificador 0-R
Árboles de decisión	DecisionStump	Decision stump in conjunction with a boosting algorithm
Árboles de decisión	HoeffdingTree	Algoritmo de inducción incremental del árbol de decisión [HSD01]

Árboles de decisión	J48	Árbol de decisión C4.5 podado o no podado [Qui93]
Árboles de decisión	LMT	“Árboles de Modelos Logísticos” o “Logistic Model Trees” (LMT) [LHF05] [SFH05]
Árboles de decisión	RandomForest	“Bosque de Árboles Aleatorios” o “Forest of Random Trees” [Bre01]
Árboles de decisión	RandomTree	Considera K atributos elegidos al azar en cada nodo. No realiza poda.
Árboles de decisión	REPTree	Construye un árbol de decisión/regresión utilizando la información de ganancia/varianza

4.7. Técnicas de evaluación para aprendizaje supervisado

La evaluación es la clave para lograr avances reales en el aprendizaje automático. Entre las técnicas de evaluación para problemas de aprendizaje supervisado se destacan [\[REF\]](#):

- Hold-out validation,
- K-fold cross-validation

4.7.1. Método de retención

La técnica de Retención (*Holdout method*) divide el conjunto de datos en dos partes, de modo que el modelo pueda ser entrenado y probado sobre diferentes datos. Presenta una mejor estimación del rendimiento fuera de la muestra, pero sigue siendo una estimación de “alta varianza”. Útil debido a su velocidad, simplicidad y flexibilidad.

Divide el conjunto de datos en dos conjuntos: un conjunto de entrenamiento (generalmente el 70 % del conjunto de datos completo) del que aprende el modelo y un conjunto de prueba (el otro 30 %). Debido a que el conjunto de prueba se retiene del modelo durante el entrenamiento, puede contribuir a una evaluación imparcial de qué tan bien se desempeña un modelo en datos nunca antes vistos. Esto protege contra el sobreajuste y permite evaluar cómo el modelo funcionaría con los nuevos datos a medida que surgen [\[REF\]](#).

4.7.2. Validación cruzada de k iteraciones

La técnica de Validación Cruzada de k iteraciones (k-fold Cross-Validation) consiste en dividir los datos en un número de particiones iguales. En la figura 4.4 se muestra un ejemplo de una validación cruzada de 10 iteraciones donde se divide el conjunto de datos en diez particiones de igual tamaño. Luego se entrena el modelo en nueve de esas diez particiones y se prueba el modelo en la partición restante. Luego, se repite el proceso, seleccionando una partición diferente para que sea el grupo de prueba y entrenando un nuevo modelo en el conjunto restante de nueve particiones. El proceso se repite ocho veces más, para un total de diez rondas de validación cruzada, una por cada pasada.

Luego se tienen diez modelos diferentes, cada uno ha sido entrenado y probado en un subconjunto diferente de datos y cada uno tiene su propio peso y exactitud de predicción. Al final, se combinan estos modelos al promediar sus pesos para estimar un modelo predictivo final. La validación cruzada es otro antídoto contra el sobreajuste [REF].



Figura 4.4: Esquema general de iteraciones para 10-fold Cross-Validation [REF]

La técnica estándar para predecir la tasa de error es Stratified k-fold Cross-Validation, donde la estratificación se refiere al proceso de reorganizar los datos de tal manera a asegurar que cada partición sea una buena representación del conjunto. Comúnmente se acepta que 10 es el número de particiones con el que se obtiene la mejor estimación de error, idea basada en diversas pruebas sobre conjuntos de datos diferentes y para distintas técnicas de aprendizaje [FHW11].

4.8. Métricas de desempeño para problemas de clasificación

Las métricas de clasificación son las medidas contra las cuales se evalúan los modelos de clasificación. Para los problemas de clasificación, es natural medir el rendimiento de un clasifi-

cador en términos de la tasa de error (*error rate*) o tasa de acierto (*sucess rate*). El clasificador predice la clase de cada instancia: si es correcta se cuenta como un éxito, sino se cuenta como un error. La tasa de error es sólo la proporción de errores cometidos sobre un conjunto de instancias, y mide el rendimiento general del clasificador. Por supuesto, lo que interesa es el probable desempeño futuro en nuevos datos, no el rendimiento sobre datos pasados.

Para predecir el rendimiento de un clasificador sobre nuevos datos, necesitamos evaluar su tasa de error en un conjunto de datos que no desempeñó ningún papel en la formación del clasificador. Este conjunto de datos independiente se denomina conjunto de prueba. En tales situaciones se suele hablar de tres conjuntos de datos, los datos de entrenamiento, los datos de validación y los datos de prueba [FHW11]. Los datos de entrenamiento son utilizados por uno o más esquemas de aprendizaje para conocer clasificadores.

Los datos de validación se utilizan para optimizar los parámetros de los clasificadores, o para seleccionar uno determinado. A continuación, los datos de prueba se utilizan para calcular la tasa de error del método final optimizado. Cada uno de los tres conjuntos debe ser independiente: El conjunto de validación debe ser diferente del conjunto de entrenamiento para obtener un buen desempeño en la etapa de optimización o selección y el conjunto de prueba debe ser diferente de ambos para obtener una estimación confiable de la tasa de error real.

Para fines prácticos generalmente se utiliza la metodología de dividir los datos en dos conjuntos, el conjunto de entrenamiento y el conjunto de prueba, como es el caso del trabajo actual donde se utilizan valores por defecto en los parámetros de los clasificadores.

4.8.1. Porcentaje de acierto

Es el porcentaje de instancias correctamente clasificadas o de predicciones correctas. La métrica más simple y más común es el porcentaje de acierto. Se calcula dividiendo el número de predicciones correctas sobre el número total de predicciones realizadas, multiplicado por 100. Por ejemplo, en un modelo de clasificación supervisada para transacciones bancarias, si probamos el modelo en cien transacciones y se predice correctamente la etiqueta (fraude/no fraude) para noventa de ellas, entonces el porcentaje de acierto del modelo es del 90 %.

$$\text{Porcentaje de acierto} = \frac{\text{Número de predicciones correctas}}{\text{Número total de predicciones realizadas}} \times 100 \quad (4.1)$$

4.8.2. Matriz de confusión

El porcentaje de acierto es la métrica de clasificación más fácil, simple y entendible que se puede utilizar. Pero no dice sobre la distribución subyacente de los valores de respuesta, ni dice qué "tipos" de errores está cometiendo el clasificador. No distingue entre falsos positivos, transacciones incorrectamente clasificadas como fraude y falsos negativos, transacciones incorrectamente clasificadas como no fraudulentas. Es por eso que se necesita la matriz de confusión.

Una matriz de confusión para una clasificación binaria es una tabla de 2 x 2 como se muestra en la figura 4.5 que clasifica las predicciones en una de cuatro clasificaciones: verdadero positivo (VP), verdadero negativo (VN), falso positivo (FP) y falso negativo (FN). En problemas de clasificación para aplicaciones reales normalmente los errores cuestan diferentes cantidades.

Por ejemplo en bancos y financieras el costo de prestar a una persona que no paga sus deudas es mayor que el costo de rechazar un préstamo a una persona que es pagadora. Los Verdaderos Positivos y Verdaderos Negativos son clasificaciones correctas. Un Falso Positivo es cuando el resultado se predice incorrectamente como positivo cuando es realmente negativo. Un Falso Negativo es cuando el resultado se predice incorrectamente como negativo cuando es realmente positivo.

		Clase Predicha	
		P	N
Clase Actual	P	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	N	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Figura 4.5: Matriz de confusión para clasificación binaria.

En la figura 4.6 se muestra una generalización de la matriz de confusión para clasificación multiclase.

		Clase Predicha			
		Clase 1	Clase 2	Clase ...	Clase N
Clase Real	Clase 1	f_{11}	f_{12}	...	f_{1N}
	Clase 2	f_{21}	f_{22}	...	f_{2N}
	Clase
	Clase N	f_{N1}	f_{N2}	...	f_{NN}

Figura 4.6: Matriz de confusión para clasificación multiclase.

La matriz de confusión brinda una imagen más completa de cómo está funcionando el clasificador. También permite calcular varias métricas de clasificación y estas métricas pueden guiar la selección del modelo. La elección de la métrica depende del objetivo de su negocio. Es importante identificar si FP o FN es más relevante reducir, para luego elegir la métrica con la variable relevante (FP o FN en la ecuación). Entre las métricas calculadas a partir de una matriz de confusión se encuentran [\[REF\]](#):

4.8.2.1. Precisión (accuracy)

En general, ¿con qué frecuencia es correcto el clasificador?.

$$Precisión = \frac{VP + VN}{VP + VN + FP + FN} \quad (4.2)$$

La generalización de la fórmula para clasificación multiclase es la siguiente, donde n es el número total de instancias evaluadas y N el número total de clases.

$$Precisión = \frac{1}{n} \sum_{i=1}^N f_{ii} \quad (4.3)$$

4.8.2.2. Exactitud (precision)

Cuando se predice un valor positivo, ¿con qué frecuencia es correcta la predicción?. ¿Cuán “exacto” es el clasificador al predecir instancias positivas?.

$$Exactitud = \frac{VP}{VP + FP} \quad (4.4)$$

La generalización de la fórmula para clasificación multiclase es la siguiente, donde n es el número total de instancias evaluadas y N el número total de clases.

$$Exactitud_i = \frac{f_{ii}}{\sum_{j=1}^N f_{ji}}$$

4.8.2.3. Sensibilidad (sensitivity)

Cuando el valor real es positivo, ¿con qué frecuencia es correcta la predicción?. Algo que queremos maximizar. ¿Qué tan “sensible” es el clasificador para detectar instancias positivas?. También conocido como True Positive Rate o Recall.

$$Sensibilidad = \frac{VP}{VP + FN} \quad (4.5)$$

La generalización de la fórmula para clasificación multiclase es la siguiente, donde n es el número total de instancias evaluadas y N el número total de clases.

$$Sensibilidad_i = \frac{f_{ii}}{\sum_{j=1}^N f_{ij}}$$

4.8.2.4. Especificidad (specificity)

Cuando el valor real es negativo, ¿con qué frecuencia es correcta la predicción?. Algo que queremos maximizar. ¿Qué tan “específico” (o “selectivo”) es el clasificador en predecir las

instancias negativas?

$$Especificidad = \frac{VN}{VN + FP} \quad (4.6)$$

4.8.2.5. Medida F (f-measure)

La Medida F es una medida de la exactitud de una prueba. La Medida F puede interpretarse como un promedio ponderado de la precisión y sensibilidad, donde alcanza su mejor valor en 1 y el peor en 0.

$$Medida F_1 = \frac{2 \times Sensibilidad \times Precision}{(Sensibilidad + Precision)} \quad (4.7)$$

La generalización de la fórmula para clasificación multiclase es la siguiente:

$$Medida F_1(i) = \frac{2 \times Sensibilidad_i \times Precision_i}{(Sensibilidad_i + Precision_i)} \quad (4.8)$$

4.8.2.6. Estadística kappa (kappa Statistic)

En la predicción multiclase, cada elemento de la matriz de confusión muestra el número de ejemplos de prueba para los que la clase real es la fila y la clase prevista es la columna. Son buenos resultados los grandes números en la diagonal principal e idealmente cero fuera de la diagonal principal. El coeficiente kappa de Cohen (κ) [Coh60] es una estadística que mide el acuerdo entre evaluadores para los ítems cualitativos (categóricos). Si los examinadores están completamente de acuerdo Kappa es igual a 1. Si no hay total acuerdo entre los observadores (aparte de lo que se esperaría por casualidad), entonces Kappa tiene un valor < 1 .

$$\kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \quad (4.9)$$

Donde: $Pr(a)$ es el acuerdo relativo observado entre examinadores o el número de instancias que fueron clasificadas correctamente sobre el total de instancias, $Pr(e)$ es la probabilidad de que el acuerdo entre examinadores se deba al azar y está relacionada con el número de instancias de cada clase y el número de instancias que el clasificador concuerda con el valor verdadero de la clase. El resultado del numerador da el grado de acuerdo realmente logrado por encima del azar y el del denominador da el grado de acuerdo que es alcanzable por encima del azar.

La generalización de la fórmula para clasificación multiclase es la siguiente, donde n es el número total de instancias evaluadas, N el número total de clases.

$$Pr(a) = \frac{1}{n} \sum_{i=1}^N f_{ii} \quad (4.10)$$

$$Pr(e) = \sum_{k=1}^N \left[\frac{1}{n^2} \left(\sum_{i=1}^N f_{ik} \times \sum_{j=1}^N f_{kj} \right) \right] \quad (4.11)$$

Otra forma de interpretar los resultado es el siguiente:

- < 0 Acuerdo deficiente
- 0.01 - 0.20 Acuerdo leve
- 0.21 - 0.40 Acuerdo justo
- 0.41 - 0.60 Acuerdo moderado
- 0.61 - 0.80 Acuerdo sustancial
- 0.81 - 1.00 Acuerdo casi perfecto

4.9. Componentes de los errores de predicción: Ruido, Sesgo y Varianza

Es importante evaluar el rendimiento de un modelo de aprendizaje de la mejor manera posible. Una de las razones es para determinar si usar o no el modelo para la predicción de valores futuros, otra de las razones es que la evaluación es una componente integral de muchos métodos de aprendizaje. Los errores de predicción en los modelos de aprendizaje automático tienen tres componentes: Ruido (*Noise*), Sesgo (*Bias*) y Varianza (*Variance*) [\[REF\]](#).

4.9.1. Ruido

El ruido es un error aleatorio de varianza de la variable que está siendo medida [HPK11], representa el componente “irreducible” de error de generalización en una función, este término está fuera del control incluso aún conociendo la verdadera función de predicción. El ruido es una distorsión de los datos originales, está presente en casi cualquier problema del mundo real, sin embargo no siempre son conocidas. Como ejemplo de ruido en los datos para un sistema de monitoreo ambiental puede darse por mal funcionamientos, calibraciones defectuosas en sensores de medición, problemas de la red que transporta los datos [KACN03].

4.9.2. Sesgo

La precisión observada de la hipótesis aprendida a través de los ejemplos de entrenamiento es a menudo una mala estimación de su exactitud sobre futuros ejemplos. Debido a que la hipótesis aprendida se derivó de estos ejemplos, típicamente proporcionarán una estimación optimista sesgada de la precisión de la hipótesis sobre ejemplos futuros, esto se conoce como sesgo en la estimación [Mit97]. El sesgo conocido también como “subajuste”, ocurre cuando la función de aprendizaje mapea pobremente los datos de entrenamiento y esto da un error de generalización con datos futuros. En estadística se entiende por sesgo como la diferencia al cuadrado entre el verdadero valor y el valor esperado de la estimación [FHT01].

El sesgo define la capacidad de ajuste del modelo de la función de aprendizaje, en la figura 4.7 se observa que a menor complejidad, mayor sesgo, esto indica que el error de predicción tanto en el entrenamiento como en el testeo es mayor cuando el modelo seleccionado es mas simple [DBK95].

4.9.3. Varianza

Si la precisión de la hipótesis se mide en un conjunto imparcial de ejemplos de prueba independientes de los ejemplos de entrenamiento, la precisión medida puede variar de la precisión real, dependiendo de la configuración del conjunto particular de ejemplos de prueba. Cuanto menor es el conjunto de ejemplos de prueba, mayor es la varianza esperada [Mit97]. La varianza llamada también “sobreajuste” se da cuando la función seleccionada aprende casi perfectamente la tendencia de los datos del ejemplo de entrenamiento, pero falla en la generalización con datos nuevos del futuro.

La varianza es una medida de generalización, tanto el sesgo como la varianza están relacionadas con la complejidad del modelo de la función, la figura 4.7 muestra que a medida que aumenta la complejidad del modelo, la varianza tiende a aumentar y el sesgo tiende a disminuir [FHT01]. En Machine Learning es esencial encontrar modelos con un balance entre sesgo y varianza, o visto de otra forma, un equilibrio entre error y complejidad.

4.10. Complejidad del modelo y curvas de aprendizaje

La capacidad de generalización está determinada principalmente por dos factores [REF]:

- La complejidad del modelo (model complexity).
- El tamaño del conjunto de entrenamiento (learning curves).

La figura 4.7 muestra la relación existente entre la complejidad del modelo (eje x) y el error de predicción (eje y). Indica cómo va variando el error de predicción en los conjuntos de entrenamiento y de prueba a medida que aumenta la complejidad del modelo. Los modelos muy simples dan una situación no deseada de sesgo (subajuste) y los modelos muy complejos igualmente una situación no deseada de varianza (sobreajuste). Para obtener una buena capacidad de generalización se debe encontrar un compromiso entre sesgo y varianza.

Este compromiso entre sesgo y varianza se puede encontrar probando distintas configuraciones de los parámetros de ajuste (tuning parameters) de los algoritmos de aprendizaje y luego eligiendo el que arroje mejores resultados. Normalmente se recomienda probar las distintas configuraciones de los parámetros sobre un conjunto de datos llamado de validación, que no se debe confundir con los conjuntos de entrenamiento y de prueba.

Algunos ejemplos de familias de algoritmos y sus parámetros de ajuste son:

- Algoritmos de regresión lineal: la cantidad de atributos del modelo (que determinará la cantidad de variables independientes).

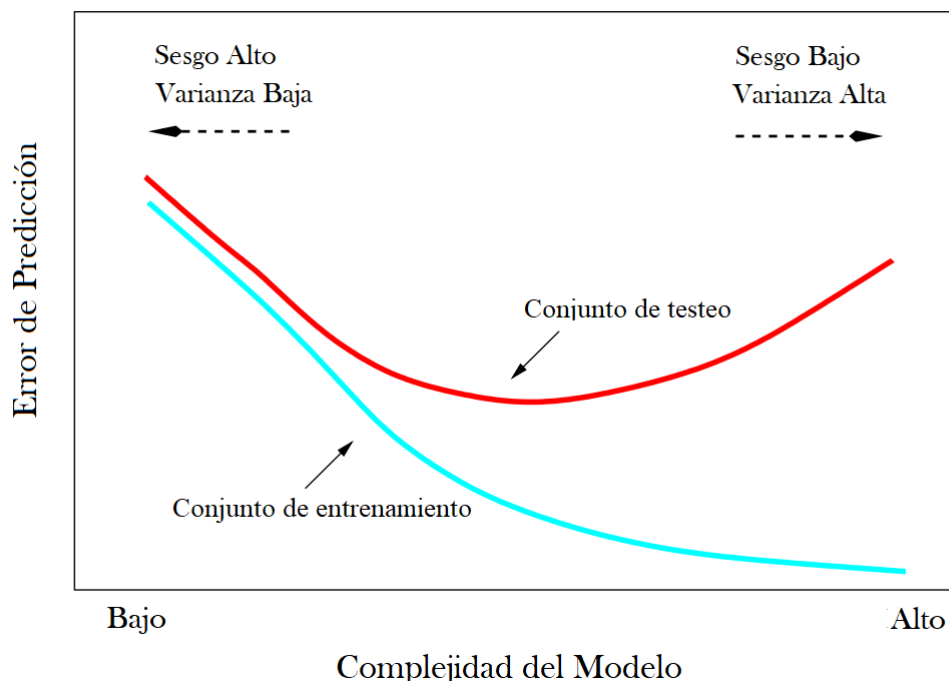


Figura 4.7: Complejidad del modelo [FHT01]

- Modelos bayesianos: en las redes bayesianas por ejemplo se debe elegir el algoritmo estimador (estimación de las tablas de probabilidad condicional de las redes bayesianas) y el método para encontrar estructuras de red.
- Árboles de decisión: el factor de confianza de la poda, mínimo número de instancias, número de pasadas para reducir el error de poda [MHC⁺16].
- Redes neuronales artificiales: la topología de la red donde se debe elegir la cantidad de neuronas para la capa de entrada, salida y la oculta. También se deben probar distintas tasas de aprendizaje (learning rate) y funciones de activación.
- Support vector machines: los parámetros gamma (γ), costo (C), y epsilon (ϵ). γ es una constante que reduce el espacio modelo y controla la complejidad de la solución, C es una constante positiva que es un parámetro de control de capacidad, mientras que ϵ es la función de pérdida que describe el vector de regresión sin todos los datos de entrada [MA16].

La figura 4.8 muestra la relación existente entre el tamaño del conjunto de entrenamiento (eje x) y el error de predicción (eje y). Indica cómo va variando el error de predicción en los conjuntos de entrenamiento y de prueba a medida que aumenta el tamaño del conjunto de entrenamiento. Los modelos muy simples dan problema de alto sesgo (figura 4.8 (a)) y los modelos muy complejos dan problema de alta varianza (figura 4.8 (b)).

Cuando el modelo está sesgado (figura 4.8 (a)) por mas que se aumenten los datos del conjunto de entrenamiento el nivel de error del conjunto de entrenamiento y el de prueba se

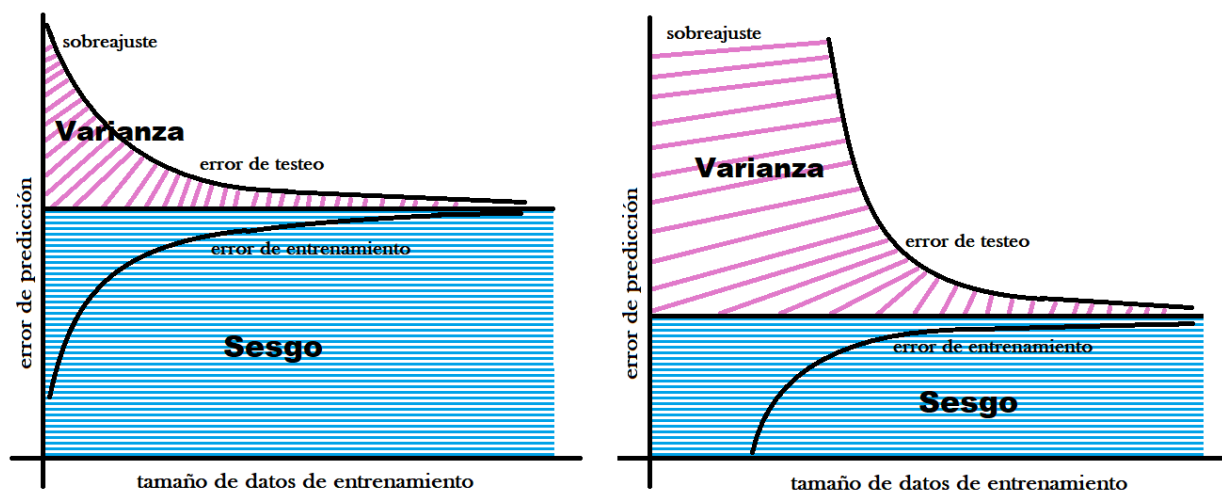


Figura 4.8: Curvas de aprendizaje

mantienen alto, por lo que para modelos sesgados no es una solución aumentar significativamente la cantidad de los datos del conjunto de entrenamiento. La solución se debería buscar aumentando la complejidad del modelo.

Cuando el modelo tiene alta varianza (figura 4.8 (b)) el error en el conjunto de entrenamiento se mantiene bajo aún aumentando su cantidad de datos. El conjunto de prueba en cambio inicia con alto porcentaje de error con pocos datos, pero a medida que aumentan los datos en el conjunto de entrenamiento el error del conjunto de prueba disminuye significativamente. Por tanto, ante alta varianza es una posible solución aumentar el tamaño de los datos. También se puede probar disminuyendo la complejidad del modelo.

4.11. Trabajos relacionados a Pronóstico - Business Intelligence - Machine Learning

En esta sección se pretende dar a conocer un análisis de los trabajos relacionados con los tres puntos más importantes del este trabajo. En los textos revisados cabe destacar que no se ha encontrado un trabajo o estudio que aborda el problema del Pronóstico de la Demanda con Business Intelligence y Machine Learning en la misma solución. Sin embargo, publicaciones que comprenden los últimos años presentan trabajos muy interesantes en el ámbito de Machine Learning asociado al Pronóstico de la Demanda (*Demand Forecasting*), como así también Pronóstico de la demanda en conjunto con Business Intelligence. Se citan a continuación algunos de los problemas afrontados en publicaciones, que dan una idea del estado del arte en este tema:

En el año 2016 [Cas16] se afronta el problema de mejorar el tiempo y la calidad en la toma de decisiones gerenciales en la empresa “Figueri SRL”, el tiempo y el esfuerzo requerido por el área de sistemas para la generación de informes a la gerencia afectan en la toma de decisiones. El enfoque adoptado para la solución consiste en la implementación de un sistema de Business

Intelligence basado en un modelo de pronóstico de ventas.

En el desarrollo de Business Intelligence se construye un Data mart bajo la metodología de Kimball para guardar los datos históricos, se realiza el proceso ETL para obtener y cargar los datos al dataware. El desarrollo del pronóstico se basa en el modelo de series de tiempo autorregresivo ARMA de la demanda de productos. Como resultado de la solución propuesta la empresa obtuvo una disminución en el porcentaje de devolución y el porcentaje de rotación de productos, los pronósticos ayudaron a la toma de decisiones, generando mayores ventas y nuevas estrategias de marketing.

En el trabajo presentado en el año 2016 [MA16] los autores abordan cómo pronosticar la demanda de agua urbana en la ciudad de Montreal (Canadá) en plazos de 1 y 3 días. Para llegar a la solución propuesta se compararon cuatro modelos de pronóstico: Artificial Neural Network (ANN), Support Vector Regression (SVR) y Extreme Learning Machine (ELM) y el tradicional Multiple Linear Regression (MLR).

Los modelos se basaron en combinaciones de variables de entrada como la temperatura diaria máxima, precipitación diaria total y la demanda diaria de agua con datos disponibles desde 1999 a 2010. Los dos índices de rendimiento utilizados para evaluar los modelos MLR, ANN, SVR y ELM fueron el Coeficiente de Determinación (R^2) y el Error Cuadrático Medio (RMSE). ELM alcanzó una mayor precisión en ambos períodos de pronóstico 1 y 3 días, en comparación con MLR, ANN y SVR que alcanzaron buena precisión para el plazo de 1 día. En general, ELM resulta ser un método de aprendizaje eficiente cuando se trata de pronosticar a corto plazo.

En el año 2015 se presentó [FLSL15] un caso de estudio basado en la empresa de retailer online *Rue La La* (<http://www.ruelala.com>). Los autores buscan estrategias para estimar la demanda de estilos nunca antes vendidos y también buscan un algoritmo que optimice combinaciones de precios para que sirva como herramienta de apoyo diario en la toma de decisiones de precios y que maximice los ingresos de dichos primeros estilos de exposición. Los datos de transacciones de ventas abarcaron desde principios de 2011 hasta mediados de 2013, donde cada registro representa una venta con su sello de tiempo, cantidad vendida del artículo, precio, fecha/hora de inicio del evento, duración del evento y el inventario inicial del artículo.

Además, se disponen de datos relacionados al producto como la marca, el tamaño, el color, el precio sugerido por el fabricante (MSRP o Manufacturer's Suggested Retail Price) y la clasificación jerárquica del producto. Los modelos probados fueron: Least Squares Regression, Principal Components Regression, Partial Least Squares Regression, Multiplicative (power) Regression, Semilogarithmic Regression y Regression Trees. Para comparar estos modelos se dividió aleatoriamente los datos en conjuntos de entrenamiento y prueba, y se utilizó los datos de entrenamiento para construir los modelos de regresión.

Para los modelos que requerían parámetros de ajuste se utilizó validación cruzada de cinco pasadas sobre los datos de entrenamiento. Se encontró que Regression Trees con Bagging superó a los demás métodos de regresión. Luego de poner en producción la herramienta las ventas no disminuyen debido a la implementación de aumentos de precios recomendados por el

algoritmo de optimización de precios. Se logró un aumento en los ingresos del grupo de prueba en aproximadamente 9.7 %.

El trabajo presentado en el año 2014 [SLH14], plantea como objetivo pronosticar la demanda del petróleo crudo importado (Imported Crude Oil - ICO) en Taiwán. Para el estudio se utilizaron datos reales, desde los años 1993 a 2010, que fueron registrado por la Oficina de Energía de Taiwán. Como solución se propone un modelo híbrido de dos etapas. Para evidenciar la efectividad de los modelos híbridos se lo comparan con modelos de una sola etapa.

Los modelos de pronóstico de una sola etapa incluyen: Multiple Linear Regression (MLR), Support Vector Regression (SVR), Artificial Neural Networks (ANN), y Extreme Learning Machine (ELM), y una selección de 23 variables influyentes como posibles datos de entrada. Para ANN y SVR se utilizaron las 23 variables, para ELM entre 1 y 15 variables, y para MLR se utilizaron 2 variables significativas.

Para los modelos de pronóstico híbridos se utilizaron las 2 variables seleccionadas para MLR como variables de entrada, dando así 3 modelos híbridos: MLR(sel)-ANN, MLR(sel)-SVR y MLR(sel)-ELM. Como medidas de rendimiento de los pronósticos se utilizaron: Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), y el Mean Absolute Difference (MAD) fueron utilizados. Los resultados mostraron que los enfoques híbridos propuestos son más precisos que los de una sola etapa, por lo tanto son capaces de predecir con mayor precisión la demanda de petróleo crudo en Taiwán.

Capítulo 5

Propuesta de solución

5.1. Esquema general de solución

En la figura 5.1 se ilustra el esquema general de la solución.

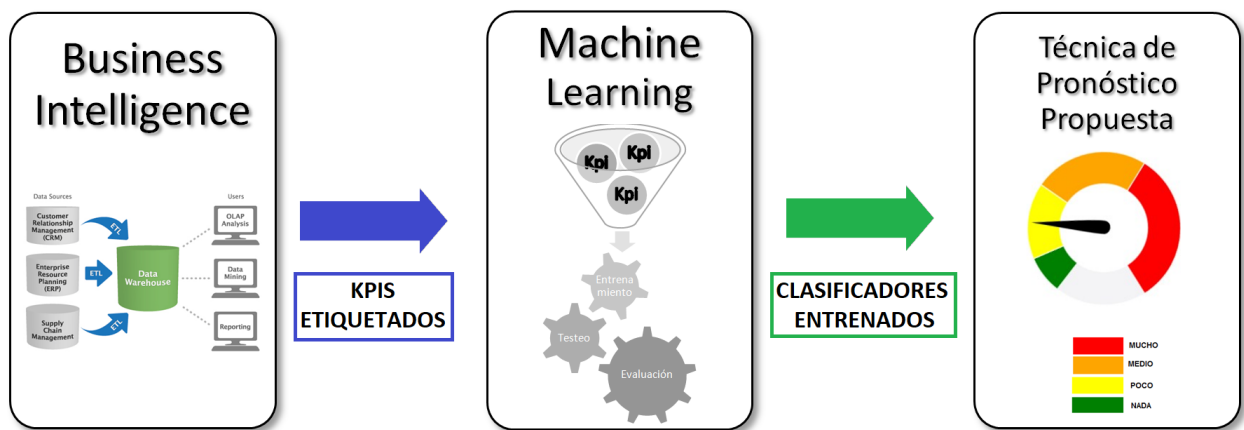


Figura 5.1: Esquema general de solución

En este capítulo se desarrolla el primer componente de la solución propuesta que es Business Intelligence. Para ello, se obtienen los datos de la fuentes de información, se sigue el proceso de ETL y se construye el datawarehouse. Luego se definen y obtienen los valores de los KPI, a continuación se define la estrategia de etiquetado y se asignan las etiquetas a cada tupla de KPI.

En el capítulo siguiente se desarrollan el segundo y tercer componente de la solución propuesta que son: Machine Learning y la Técnica de solución propuesta.

5.2. Componente de solución “Business Intelligence”

Esta sección se centra en los tres primeros componentes de Business Intelligence siguiendo el proceso de modelado dimensional[Kim92]. Uno de los principales problemas con los que se enfrentan las empresas retail, trata acerca de pronosticar la demanda, es decir, determinar la

cantidad de productos que se deben disponer para satisfacer la demanda de los clientes en el siguiente periodo. Utilizando los conceptos y herramientas de Business Intelligence, se parte con el análisis de una fuente de información auténtica obtenida de una empresa retail; se diseña el datawarehouse que será poblado con información de las transacciones operacionales diarias del negocio.

Posteriormente se definen los indicadores claves de rendimiento como métricas, se calculan los valores de los indicadores con los datos históricos almacenados en el datawarehouse para cada periodo de tiempo establecido y se asigna una etiqueta a cada tupla de valores KPI que finalmente servirán como parámetros de entrada de las herramientas de aprendizaje automático para diseñar una solución que pronostique la demanda.

5.2.1. Fuente de Información

Para el presente trabajo, se cuenta con una base de datos relacional Oracle 10g con las operaciones transaccionales de una empresa retail dedicada a la venta de productos alimenticios y artículos de limpieza, algunas de las líneas de productos con que cuenta la empresa son: aceites corporales, acondicionadores, aromatizantes, cuidado corporal, desodorantes, limpiadores, salud e higiene, salud y belleza, aguas, gaseosas, cervezas, vinos, chocolates, galletitas, enlatados, lácteos, yerbas y varias líneas de productos más. La base de datos almacena datos de las operaciones comprendidas entre noviembre de 2013 y octubre de 2016. A continuación una reseña de las principales tablas tenidas en cuenta para el diseño del datawarehouse.

- **Tabla Productos:** almacena datos como descripción, unidad de medida, categoría, tipo de impuesto, línea del producto, marca, código de barras, proveedor, costo, precio de venta, fecha de registro, etc., de los productos disponibles para la venta, la tabla cuenta con 13.200 artículos registrados.
- **Tabla Proveedores:** almacena datos como denominación, dirección, teléfono, ruc, email, página web, ciudad, país, propietario, etc., de los proveedores de la empresa, la tabla cuenta con 1.623 proveedores registrados.
- **Tabla de Ventas Cabecera:** es una de las tablas principales donde se registran los movimientos de ventas de la empresa. Contiene datos como número de factura, moneda, tipo de comprobante, caja, usuario, fecha, cliente, monto total, monto gravado, monto impuesto, monto exenta, etc., la tabla cuenta con 301.316 registros de ventas correspondientes al periodo mencionado previamente.
- **Tabla de Ventas Detalle:** contiene los registros de los productos que fueron comercializados, cada detalle está relacionado a un registro de la tabla venta cabecera. Contiene datos de la fecha, el producto, precio de costo unitario, precio de venta unitario, cantidad, importe grabado, importe del impuesto entre otros datos, la tabla cuenta con 981.402 detalles de ventas registrados.

- **Tabla de Movimientos de Stock:** contiene los registros de movimientos de stock de ventas y compras detallado. Contiene datos de fecha, producto, cantidad, tipo de movimiento, costo unitario, precio unitario, entre otro datos, la tabla cuenta con 1.062.440 movimientos de compra y ventas registradas.

5.2.2. Proceso ETL.

En el proceso ETL posterior a la extracción y previa a la carga de los datos al datawarehouse fueron realizadas algunas transformaciones y limpieza sobre los datos de origen.

- **Tabla de Productos:** se detectaron registros de artículos con las siguientes inconsistencias:
 - Datos del proveedor con valores nulos, los cuales eran completados con un proveedor por defecto de la tabla dimensional de proveedores.
 - Artículos con valores de costo nulo, en tales casos los valores eran asignados con un costo promedio tomados de la tabla de Ventas Detalle.
 - Artículos con valores donde el costo eran mayor al precio de venta unitario, en dichos casos los datos fueron completados con el costo promedio tomados de la tabla de Ventas Detalle.
 - Artículos cuyo precio de venta unitario era nulo, en los cuales los datos eran completados con el precio de venta mas reciente tomado de la tabla de Ventas Detalle.
- **Tabla de Ventas Cabecera:** se encontraron registros donde los datos del cliente eran nulos, en dichos casos fueron asignados un cliente por defecto tomados de la tabla dimensional de clientes.
- **Tabla de Ventas Detalle:** se detectaron registros con las siguientes falencias:
 - Registros de detalles donde los valores de costo eran iguales a cero, los cuales eran modificados por el costo promedio de la tabla dimensional de productos.
 - Registros de detalle donde el costo unitario eran mayores al precio de venta unitario, las cuales fueron corregidas con el costo promedio, tomados de la tabla dimensional de productos.

5.2.3. Datawarehouse

El datawarehouse se diseña partiendo de la definición de las tablas de hechos y dimensiones utilizando el modelo en esquema estrella [Kim92].

Tablas de hechos

De las tablas transaccionales se definen 3 tablas de hechos a partir de las cuales se obtienen los valores de los KPI que se utilizan en la solución propuesta.

- **Tabla de hechos Cabecera:** almacena los datos históricos de las ventas, cada registro de la tabla de hechos guarda datos de: fecha, cliente, caja, número de factura y monto total, monto exento, monto gravado IVA. Las métricas definidas para la tabla de hechos son: monto total, monto exento, monto gravado IVA.
- **Tabla de hechos Detalles:** almacena los datos históricos en detalle por cada producto vendido, cada registro guarda información del número de comprobante, fecha, producto, proveedor, cliente, cantidad, costo unitario, precio unitario, impuesto e importe total. Las métricas asociadas a la tabla de hechos son: cantidad, precio unitario, impuesto, costo unitario y el importe total.
- **Tabla de hechos Stock:** almacena los datos históricos de cada movimiento ya sea compra o de venta realizada, cada registro contiene información como fecha, producto, tipo de movimiento, cantidad, precio unitario y costo unitario. Las métricas definidas para la tabla de hechos son: cantidad, precio unitario y costo unitario.

Dimensiones

Las tablas de dimensiones diseñadas para el modelado del datawarehouse y que se encuentran relacionadas a las tablas de hechos son:

- **Dimensión Fecha:** La tabla de dimensión fecha está ligada a todas las tablas de hechos, sirve para limitar o agrupar los datos de las tablas de hechos al momento de realizar consultas sobre estas en el tiempo. Con la dimensión fecha se pueden establecer niveles jerárquicos en días, semanas, meses, trimestres, semestres y años.
- **Dimensión Productos:** La tabla de dimensión producto está relacionada a las tablas de hechos Detalles y Stock. Contiene los atributos o campos (ej.: nombre del producto) por las cuales se pueden filtrar o agrupar datos.
- **Dimensión Proveedores:** La tabla de dimensión proveedores está relacionada a la tabla de hechos Detalles. Contiene los atributos o campos (ej.: nombre del proveedor) por las cuales se pueden realizar filtros o agrupaciones de datos.
- **Dimensión Clientes:** La tabla de dimensión Clientes está relacionada a las tablas de hechos Cabecera y Detalles. Contiene los atributos o campos (ej.: nombre del cliente) por las cuales se pueden hacer consultas de datos.

- **Dimensión Cajas:** La tabla de dimensión Cajas está relacionada a la tabla de hechos Cabecera. Contiene los atributos o campos (ej.: número de caja) por las cuales se pueden filtrar o agrupar datos.

En figura 5.2 observamos un ejemplo del modelo en esquema estrella de la tabla de hechos Detalles.

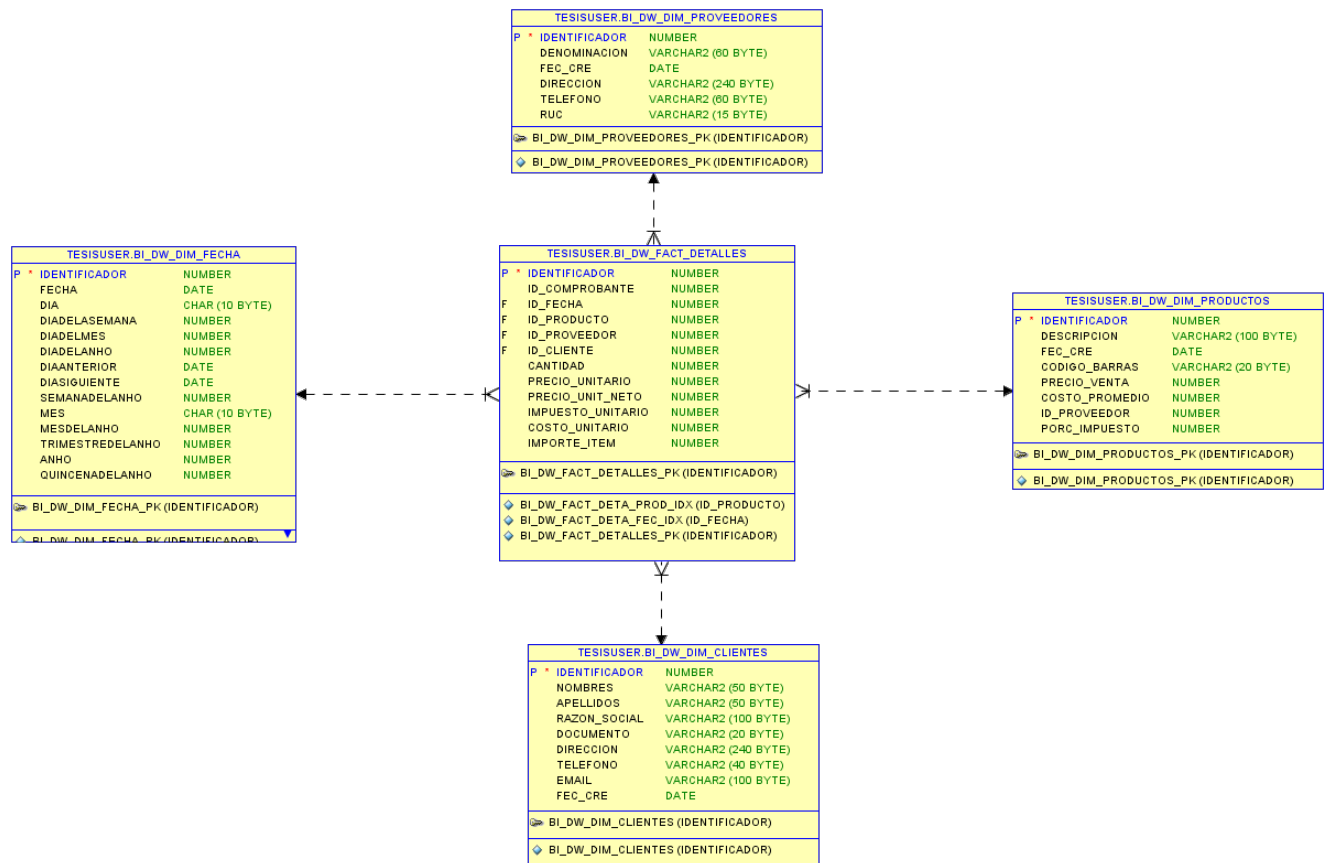


Figura 5.2: Tabla de hechos Detalles

5.3. Definición de los Indicadores Claves de Rendimiento

En el marco de este trabajo, en esta sección se definen los KPIs y se obtienen los valores que serán utilizados como datos de entrada en las herramientas de aprendizaje automático para el desarrollo de una técnica para pronosticar la demanda y que servirá de apoyo en la toma de decisiones para la reposición de stock del siguiente periodo[Alv13] (Ej.: cantidad a comprar para satisfacer la demanda de la siguiente semana, quincena, o mes).

Los KPIs definidos fueron adaptados a la solución planteada debido a que en los textos consultados los KPI engloban a toda la organización en áreas como compras, ventas, marketing, recursos humanos y otros. Cada KPI mide un valor obtenido de los datos históricos almacenados en el datawarehouse. El cálculo de cada KPI se realiza por cada producto y periodo de tiempo (semanal, quincenal o mensual). A continuación se definen los KPIs

Ticket Medio. Es la cantidad media por cada transacción de venta que se realiza de un determinado producto. El indicador viene determinado por dos variables: a) la cantidad total vendida del producto, b) el total de tickets en las que fue vendido el producto. Aplicando la siguiente fórmula obtenemos el valor de la cantidad media de venta para cada producto.

$$TM = \frac{Cantidad\ Vendida\ Periodo}{Total\ Tickets\ Periodo} \quad (5.1)$$

Cifra de Ventas La cifra de ventas es un KPI que sirve para explicar el importe total de ventas que se ha obtenido para un producto. Se obtiene de la siguiente fórmula.

$$CV = Precio \times Cantidad\ Vendida\ Periodo \quad (5.2)$$

Margen Comercial Es la razón entre el precio de venta y precio de costo del producto, es un indicador que permite conocer el porcentaje de rentabilidad del producto. Se obtiene de la siguiente fórmula.

$$MC = \frac{(Precio - Costo) \times Cantidad\ Vendida\ Periodo}{Precio \times Cantidad\ Vendida\ Periodo} \times 100 \quad (5.3)$$

Rotación de Stock Este indicador mide la cantidad de veces que el stock del producto se renueva durante un determinado ciclo comercial, es decir, la cantidad de veces que se recupera la inversión. Se obtiene de la siguiente fórmula.

$$RS = \frac{Cantidad\ Vendida\ Periodo}{\left(\frac{Stock\ Inicial\ Periodo + Stock\ Final\ Periodo}{2}\right)} \quad (5.4)$$

Coeficiente de Rentabilidad El indicador mide la rentabilidad obtenida por la empresa basada en el margen y la rotación, el objetivo de toda empresa retail es aumentar los niveles de rotación. El coeficiente se obtiene de la siguiente fórmula.

$$CR = ((Precio - Costo) \times Cantidad\ Vendida\ Periodo) \times RS \quad (5.5)$$

Cobertura de Stock Este indicador muestra el periodo de tiempo (habitualmente se expresa en días o semanas) que el negocio puede continuar vendiendo con el stock de que dispone en el momento, sin incorporar nuevas cantidades de ese producto.

$$CS = \frac{Stock\ Inicio\ Periodo}{Promedio\ Cantidad\ Vendida\ Ultimos\ N\ Periodos} \quad (5.6)$$

5.4. Obtención de los valores para los Indicadores Clave de Rendimiento.

Se obtiene los valores de KPI por cada producto y periodo, mediante codificación de sentencias SQL que operan sobre los datos almacenados en el datawarehouse, y todos los resultados obtenidos de la ejecución se almacenan en una tabla base. En dicha tabla base además de los valores de KPI, en cada registro se guarda adicionalmente información de cantidad, fecha, año, mes, quincena, semana y periodo. Durante el cálculo de los valores de los KPI se establecieron ciertas restricciones, como por ejemplo si un producto no fue vendido durante un número consecutivo de periodos entonces era descartado, ya que los resultados para algunos KPI daban como resultado un valor cero, el cual no tiene relevancia.

PERIODOS DE ANÁLISIS

De la tabla base donde se almacenan los valores obtenidos de los KPI, por la columna *periodo* se pueden agrupar los registros en los tres periodos de análisis: semanal, quincenal y mensual.

SEMANAL En la figura 5.3 se puede observar un extracto de los valores de la tabla base agrupados por periodo semanal para un determinado producto. Por ejemplo, el primer registro corresponde al producto con Id 133, periodo 'S' (semanal), semana 49, mes 12 y año 2013, observamos que para el KPI Rotación de Stock según la fórmula 5.4 se obtuvo un valor 0,769230.

KPI_TIKET_MEDIO	KPI_CIFRA_VENTAS	KPI_MARGEN_COMERCIAL	KPI_ROTACION_STOCK	KPI_COEF_RENTABILIDAD	KPI_COBERTURA_STOCK	CANTIDAD	FECHA	ANHO	MES	QUINCENA	SEMANA	ID_PRODUCTO	PERIODO
5000	25000	7380,95238095238	0,769230769230769	5677,65567765568	1,2	5	01/12/2013 ▾	2013	12		49	133 S	
5000	20000	5904,7619047619	0,571428571428571	3374,14865986395	2,07692307692308	4	08/12/2013 ▾	2013	12		50	133 S	
5000	15000	4428,57142857143	0,272727272727273	1207,79220779221	0,9375	3	15/12/2013 ▾	2013	12		51	133 S	
5000	10000	2952,38095238095	0,125	369,047619047619	4,25	2	22/12/2013 ▾	2013	12		52	133 S	
5000	10000	2952,38095238095	0,142857142857143	421,768707482993	5	2	29/12/2013 ▾	2013	12		53	133 S	
5000	5000	1476,19047619048	0,08	118,095238095238	5,57142857142857	1	01/01/2014 ▾	2014	1		1	133 S	
5000	15000	4428,57142857143	0,285714285714286	1265,30612244898	7,2	3	05/01/2014 ▾	2014	1		2	133 S	
5000	20000	5904,7619047619	0,571428571428571	3374,14865986395	4,5	4	12/01/2014 ▾	2014	1		3	133 S	
5000	20000	5904,7619047619	1,33333333333333	7873,01587301587	1,875	4	19/01/2014 ▾	2014	1		4	133 S	
5000	10000	2952,38095238095	0,2	590,47619047619	0,272727272727273	2	26/01/2014 ▾	2014	1		5	133 S	

Figura 5.3: Ejemplo de valores KPI para un periodo semanal

QUINCENAL En la figura 5.4 se puede observar un extracto de los valores de la tabla base agrupados por periodo quincenal para un determinado producto. Por ejemplo, el primer registro corresponde al producto con Id 133, periodo 'Q' (quincenal), quincena 23, mes 12 y año 2013, observamos que para el KPI Cobertura de Stock según la fórmula 5.6 se obtuvo un valor de 1,2.

KPI_TIKET_MEDIO	KPI_CIFRA_VENTAS	KPI_MARGEN_COMERCIAL	KPI_ROTACION_STOCK	KPI_COEF_RENTABILIDAD	KPI_COBERTURA_STOCK	CANTIDAD	FECHA	AÑO	MES	QUINCENA	SEMANA	ID_PRODUCTO	PERIODO
1	5000	13285,7142857143	2	26571,4285714286	1,2	1,2	9/01/12/2013	2013	12	23		133 Q	
2	5000	10333,3333333333	0,77777777777778	8037,0303030304	0,789473684210526	0,789473684210526	7/16/12/2013	2013	12	24		133 Q	
3	5000	10333,3333333333	0,736842105263158	7614,030877193	1,625	1,625	7/01/01/2014	2014	1	1		133 Q	
4	5000	10333,3333333333	0,56	5786,66666666667	0,782608695652174	0,782608695652174	7/16/01/2014	2014	1	2		133 Q	
5	5000	10333,3333333333	0,451612903225806	4656,66666666667	2,71428571428571	2,71428571428571	7/01/02/2014	2014	2	3		133 Q	
6	5000	15000	0,285714285714286	1265,30612244898	1,71428571428571	1,71428571428571	3/16/02/2014	2014	2	4		133 Q	
7	5000	1476,19047619048	0,117647058823529	173,669467787115	1,58823529411765	1,58823529411765	1/01/03/2014	2014	3	5		133 Q	
8	0	0	0	0	2,18181818181818	2,18181818181818	0/16/03/2014	2014	3	6		133 Q	
9	0	0	0	0	6	6	0/01/04/2014	2014	4	7		133 Q	
10	2500	1476,19047619048	0,133333333333333	196,825396825397	24	24	1/16/04/2014	2014	4	8		133 Q	
11	5000	2952,38095238095	0,333333333333333	984,126984126984	21	21	2/01/05/2014	2014	5	9		133 Q	
12	5000	2952,38095238095	0,5	1476,19047619048	5	5	2/16/05/2014	2014	5	10		133 Q	
13	5000	2952,38095238095	1	2952,38095238095	1,8	1,8	2/01/06/2014	2014	6	11		133 Q	

Figura 5.4: Ejemplo de valores KPI para un periodo quincenal

MENSUAL En la figura 5.5 se puede observar un extracto de los valores de la tabla base agrupados por periodo mensual para un determinado producto. Por ejemplo, el primer registro corresponde al producto con Id 133, periodo 'M' (mensual), mes 12 y año 2013, observamos que para el KPI Margen Comercial según la fórmula 5.3 se obtuvo un valor 23619,047619.

KPI_TIKET_MEDIO	KPI_CIFRA_VENTAS	KPI_MARGEN_COMERCIAL	KPI_ROTACION_STOCK	KPI_COEF_RENTABILIDAD	KPI_COBERTURA_STOCK	CANTIDAD	FECHA	ANHO	MESES	QUINCENA	SEMANA	ID_PRODUCTO	PERIODO
5000	80000	23619,0476190476	1,88235294117647	44459,383755014	1,2	16	01/12/2013	2013	12			133	M
5000	70000	20666,6666666667	0,875	18083,3333333333	1,5	14	01/01/2014	2014	1			133	M
5000	50000	14761,9047619048	0,714285714285714	10544,2176870748	1,425	10	01/02/2014	2014	2			133	M
5000	5000	14761,9047619048	0,117647058823529	173,669467787115	0,675	1	01/03/2014	2014	3			133	M
2500	5000	14761,9047619048	0,133333333333333	196,825398825397	0,96	1	01/04/2014	2014	4			133	M
5000	20000	5904,7619047619	0,8	4723,8095230952	1,75	4	01/05/2014	2014	5			133	M
5000	15000	4428,57142857143	3	13285,7142857143	1,5	3	01/06/2014	2014	6			133	M
0	0	0	0	0	-0,375	0	01/07/2014	2014	7			133	M
0	0	0	0	0	-0,428571428571428	0	01/08/2014	2014	8			133	M
5000	30000	8571,42857142857	0,363636363636364	3116,8831688312	19	6	01/09/2014	2014	9			133	M
5000	70000	20000	2	40000	7	14	01/10/2014	2014	10			133	M

Figura 5.5: Ejemplo de valores KPI para un periodo mensual

5.5. Asignación de etiquetas

A cada tupla de KPI se le debe asignar una sola etiqueta de entre las siguientes “*Nada*”, “*Poco*”, “*Medio*” o “*Mucho*”. El etiquetado es uno de los puntos focales para el aprendizaje automático.

Para una mayor fiabilidad esta asignación de etiquetas debe ser realizada y revisada por el experto del área de compras (que podría ser el gerente de administración de compras u otra persona a cargo de la reposición de stock), es decir, el experto debería analizar los KPI por cada tupla y tomar la decisión en base a sus conocimientos y experiencia el volumen de compra apropiado para la reposición de stock para el siguiente periodo. Por ejemplo si el experto escoge la etiqueta *Nada* para una tupla dada, significa que el considera no realizar ninguna compra para el siguiente periodo.

Para el presente trabajo el etiquetado fue realizado en forma empírica, sin la intervención de un experto por la dificultad de contar con una persona especializada en el área. La estrategia para el etiquetado se basa en los siguientes pasos:

Paso 1:

Para cada KPI se define una tabla con un rango de valores, luego de acuerdo al valor calculado para dicho KPI se le asigna una determinada letra. A continuación mostramos la tabla de rangos para cada KPI.

Tabla 5.1: Rango KPI Ticket Medio

(=) igual a 0	a
> (mayor) a 0 y < (menor) a 1	b
>= (mayor o igual) a 1 y <= (menor o igual) a 3	c
> (mayor) a 3	d

Tabla 5.2: Rango KPI Cifra Ventas (%)

>= (mayor o igual) a 0 y <= (menor o igual) a 20	e
> (mayor) a 20 y <= (menor o igual) a 50	f
> (mayor) a 50 y <= (menor o igual) a 80	g
> (mayor) a 80 y <= (menor o igual) a 100	h

Tabla 5.3: Rango KPI Margen Comercial (%)

>= (mayor o igual) a 0 y <= (menor o igual) a 20	i
> (mayor) a 20 y <= (menor o igual) a 50	j
> (mayor) a 50 y <= (menor o igual) a 80	k
> (mayor) a 80 y <= (menor o igual) a 100	l

Tabla 5.4: Rango KPI Rotación Stock

(=) igual a 0	m
> (mayor) a 0 y < (menor) a 1	n
>= (mayor o igual) a 1 y <= (menor o igual) a 3	o
> (mayor) a 3	p

Tabla 5.5: Rango KPI Cobertura Stock

(=) igual a 0	q
> (mayor) a 0 y < (menor) a 1	r
>= (mayor o igual) a 1 y <= (menor o igual) a 3	s
> (mayor) a 3 y <= (menor o igual) a 10	t
> (mayor) a 10	u

Paso 2:

Posteriormente cada tupla queda definida por un conjunto de 5 letras, luego a cada combinación de letras se asignó de forma empírica una de las posibles etiquetas “Nada”, “Poco”, “Medio” o “Mucho” denotado en el cuadro 5.6.

Tabla 5.6: Tabla de etiquetado por el experto

aeimq	Nada	bejnq	Poco	bejoq	Poco
aeimr	Nada	bejnr	Poco	bejpq	Medio
aeims	Nada	bejns	Nada	beknq	Poco
aeimt	Nada	bejnt	Nada	beknr	Poco
aeimu	Nada	bejnu	Nada	bekns	Nada
...

En la figura 5.6 se muestra un extracto del etiquetado para un periodo semanal. La última columna contiene las etiquetas como resultado del proceso.

5.6. Resumen

Sintetizando el capítulo se puede resumir que el modelado del problema del componente BI se basa en tres grandes procesos: a) proceso de Business Intelligence y sus etapas b) proceso de obtención de KPI y c) proceso de etiquetado.

El proceso de Business Intelligence incluyó tres de sus componentes, 1) el acceso a las fuentes de información, 2) el proceso de extracción, transformación y carga de los datos obtenidos de la fuente de origen y 3) el diseño y construcción del datawarehouse. El acceso a la fuente de información de los datos no representó mayores inconvenientes, el proceso consistió en restaurar en un servidor local el backup de los datos de la empresa que proporcionó la información.

Durante el ETL se tuvo que identificar las columnas de las tablas relevantes para la *extracción* que posteriormente generan las métricas de las tablas de hechos, los problemas encontrados en esta etapa se dieron por ciertas inconsistencias de datos, valores vacíos y datos nulos los cuales fueron corregidos en la *transformación* y posterior *carga* de los datos al datawarehouse. Para construir el datawarehouse el trabajo clave consistió en definir las tablas de hechos que sirven de base para el cálculo de los KPI junto con sus principales dimensiones utilizando el esquema estrella.

En el proceso de obtención de KPI, el principal trabajo fue la selección de los KPI relevantes en las áreas de negocios ventas y productos. Debido a que en la literatura los KPI miden los procesos de negocios a un nivel de generalización donde se ven a los productos como un todo, las fórmulas fueron adaptadas de modo que las mediciones fueron bajadas a un nivel de productos individuales.

En el etiquetado, lo esencial es contar con la participación de un experto en administración en compras con conocimientos y capacidad de interpretación de KPI. En este proceso se simula al experto mediante un conjunto de reglas para indicar la etiqueta a cada tupla de KPI.

Capítulo 6

Experimentación

6.1. Componente de solución “Machine Learning”

Se describirá cómo es la implementación del proceso de aprendizaje automático para este caso de estudio. Se mostrará primeramente cómo está constituida la salida del proceso de Business Intelligence, que en esencia proveen las instancias necesarias para la entrada del proceso de aprendizaje automático. También se verá qué clasificadores fueron utilizados, cómo se realizó el proceso de entrenamiento y de evaluación, y cuáles son las métricas de evaluación consideradas para medir el rendimiento de los clasificadores.

6.2. Datos de entrada de Machine Learning

La salida de Business Intelligence provee tres conjuntos de datos independientes que se corresponden con los períodos de análisis: Mensuales, Quincenales y Semanales.

- Períodos Mensuales: Se analizaron 309 productos diferentes, y por cada producto se tiene un máximo de 34 instancias. Cada instancia tiene los siguientes atributos: Ticket Medio, Cifra de Ventas, Margen Comercial, Rotación de Stock, Coeficiente de Rentabilidad, Cobertura de Stock, Cantidad, Año, Mes. La clase de cada instancia está definida por $y \in \{Nada, Poco, Medio, Mucho\}$.
- Períodos Quincenales: Se analizaron 228 productos diferentes, y por cada producto se tiene un máximo de 68 instancias. Cada instancia tiene los siguientes atributos: Ticket Medio, Cifra de Ventas, Margen Comercial, Rotación de Stock, Coeficiente de Rentabilidad, Cobertura de Stock, Cantidad, Año, Quincena. La clase de cada instancia está definida por $y \in \{Nada, Poco, Medio, Mucho\}$.
- Períodos Semanales: Se analizaron 127 productos diferentes, y por cada producto se tiene un máximo de 151 instancias. Cada instancia tiene los siguientes atributos: Ticket Medio, Cifra de Ventas, Margen Comercial, Rotación de Stock, Coeficiente de Rentabilidad, Co-

bertura de Stock, Cantidad, Año, Semana. La clase de cada instancia está definida por $y \in \{Nada, Poco, Medio, Mucho\}$.

6.3. Entrenamiento, testeo y evaluación

En la figura 6.1 se muestra la estrategia general utilizada para el entrenamiento, testeo y evaluación en los periodos semanales, quincenales y mensuales. El conjunto de entrenamiento se basa en el 70 % de las instancias y el conjunto de prueba corresponde al 30 % restante. Esta estrategia es la utilizada en el curso The Machine Learning Masterclass [Sci16], un curso moderno de Machine Learning para proyectos de análisis predictivo.

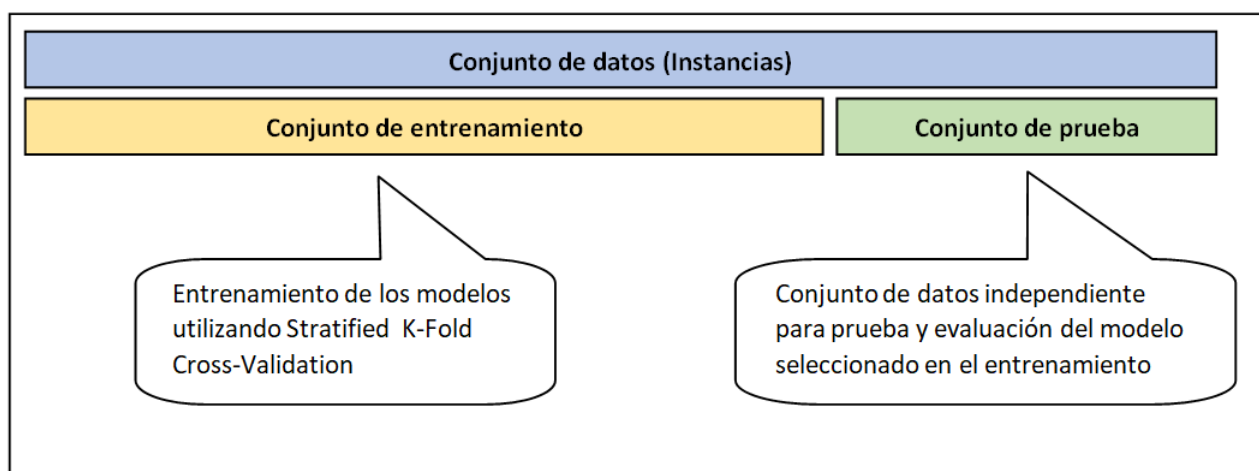


Figura 6.1: Esquema de entrenamiento y prueba.

En el algoritmo 6.1 se presenta el pseudocódigo de aprendizaje y selección de los clasificadores.

Algoritmo 6.1 Pseudocódigo para el proceso de clasificación.

```

for cada período de análisis {mensual, quincenal, semanal}:
    for cada producto con sus instancias:
        establecer conjunto de entrenamiento;
        establecer conjunto de testeo;
        for cada algoritmo de clasificación:
            construir clasificador (conjunto de entrenamiento);
            validar clasificador (stratified k-folds cross-validation);
            obtener métricas de validación;
        endfor;
        criterios de línea de base (ZeroR, criterios del experto u otro);
        seleccionar mejor clasificador (max(Kappa));
        evaluar mejor clasificador (conjunto de testeo);
        guardar clasificador;
    endfor;
endfor;
  
```

Los algoritmos utilizados para construir los clasificadores son los proveídos por la herramienta WEKA según tabla 4.1 y los parámetros de configuración establecidos por defecto. Otra forma de categorizar los clasificadores incluidos en WEKA es como sigue:

- Bayesianos: BayesNet, NaiveBayes, NaiveBayesUpdateable.
- Basados en funciones: Logistic, MultilayerPerceptron, SimpleLogistic, SMO.
- Basados en reglas: OneR, DecisionTable, JRip, PART, ZeroR.
- Basados en árboles: DecisionStump, J48, LMT, RandomForest, RandomTree, REPTree.

La validación se hace por el método Stratified k-fold Cross-Validation para un valor de k igual a 10 y las métricas de desempeño consideradas son el *Porcentaje de Acierto* y la *Estadística Kappa*. El criterio de línea de base utilizado es el clasificador ZeroR, el cual es uno de los criterios de línea de base más representativos para problemas de clasificación. Se considera que también puede resultar conveniente que el experto en compras establezca su propio criterio de línea de base, como puede ser un umbral mínimo de porcentaje de acierto aceptado.

6.4. Análisis del desempeño

6.4.1. Comparación con línea base ZeroR

En este apartado se contrastan los resultados obtenidos mediante una comparación entre el porcentaje de acierto de la línea base ZeroR, el porcentaje de acierto del mejor algoritmo alcanzado en la validación y el porcentaje de acierto en el testeo del mismo algoritmo. Como se mencionó la línea base es el clasificador ZeroR, el cual es el criterio mínimo de aceptación de un clasificador. Los puntos de la línea de entrenamiento son los porcentajes de acierto de los clasificadores seleccionados durante la validación, según el algoritmo 6.1. La línea de test contiene los puntos que representan los porcentajes de aciertos durante la evaluación sobre el conjunto de prueba.

Los productos incluidos en el gráfico constituyen un subconjunto representativo, esto debido a la extensa cantidad de productos y su dificultad para graficarlos por completo.

En las figuras 6.2 y 6.3 los gráficos contienen las líneas comparativas para periodos semanales y quincenales respectivamente. De acuerdo a las experimentaciones el porcentaje de acierto en el entrenamiento es siempre mayor o igual a la línea base, en concordancia con el algoritmo 6.1. En el testeo por tratarse de un conjunto de datos separado del conjunto de entrenamiento los porcentajes de aciertos varían.

Se dieron dos casos atípicos donde el porcentaje de acierto fue menor al de la línea base, estos casos aparecen cuando hay una etiqueta dominante, en periodos semanales fue el producto 222 con el 88 % de etiquetas “Nada” y en periodos quincenales fue el producto 400 con el 92 % de etiquetas “Nada”. De los resultados se puede deducir que hay aprendizaje, ya que los porcentajes

de acierto alcanzados en el testeo se encuentran por encima de la línea base en la mayoría de los casos.

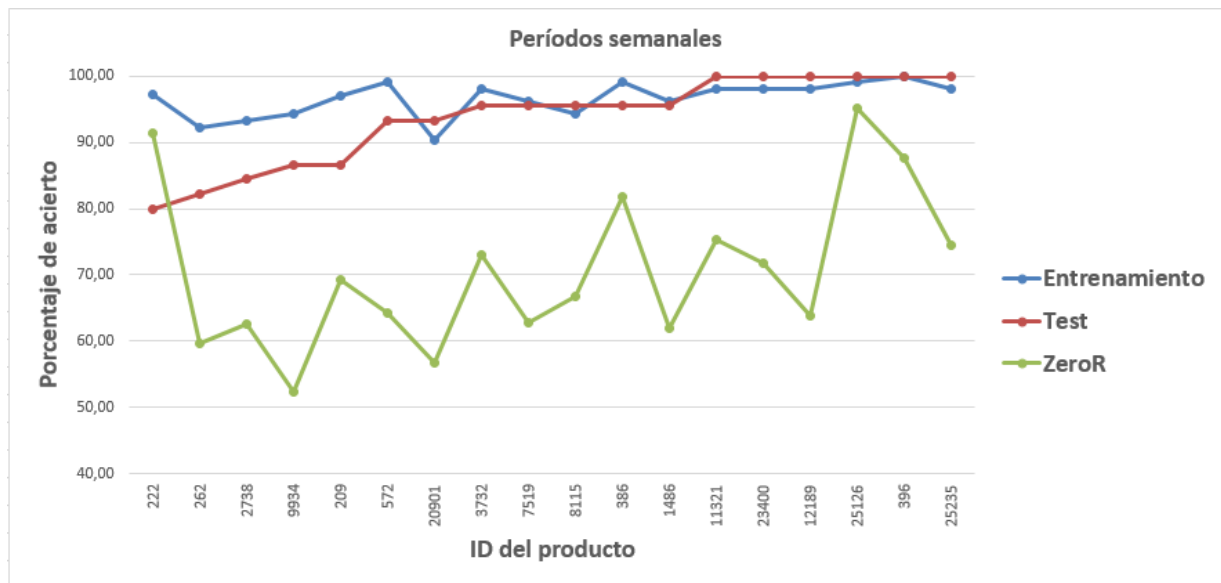


Figura 6.2: Entrenamiento vs Test vs ZeroR para períodos semanales.

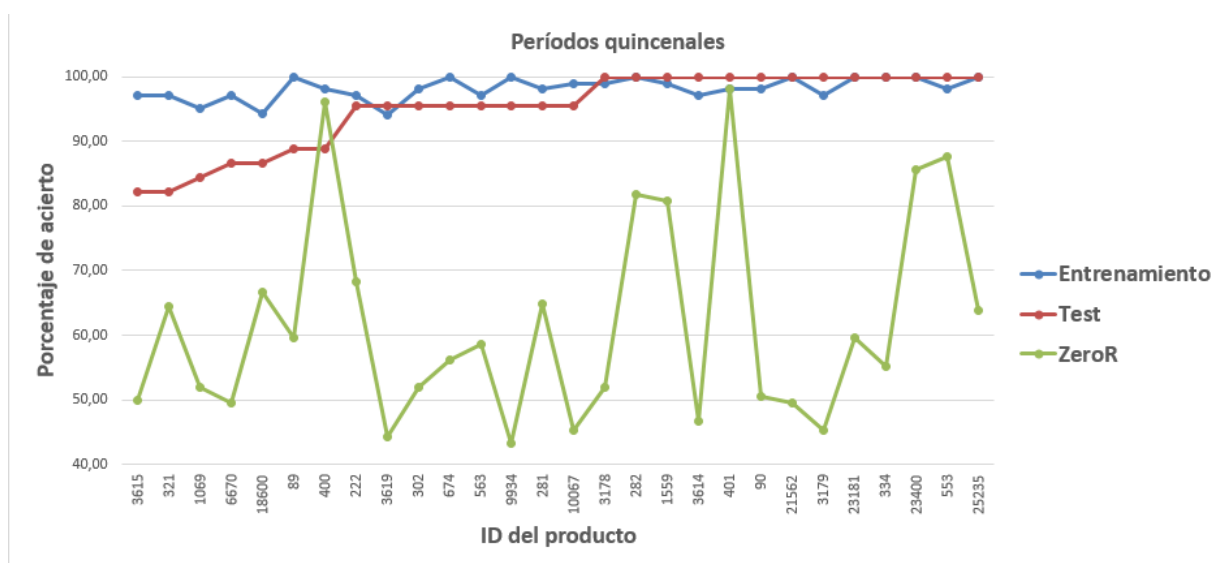


Figura 6.3: Entrenamiento vs Test vs ZeroR para períodos quincenales.

La figura 6.4 muestra las comparativas para períodos mensuales. Al igual que los períodos semanales y quincenales el porcentaje de acierto en el entrenamiento es siempre mayor o igual a la línea base. En el testeo por tratarse de un conjunto de datos separado del conjunto de entrenamiento los porcentajes de aciertos varían. Se tiene un caso atípico donde el porcentaje de acierto fue menor al de la línea base, este caso aparece cuando hay una etiqueta dominante el cual se dio en el producto 309 con el 93 % de etiquetas “Medio”.

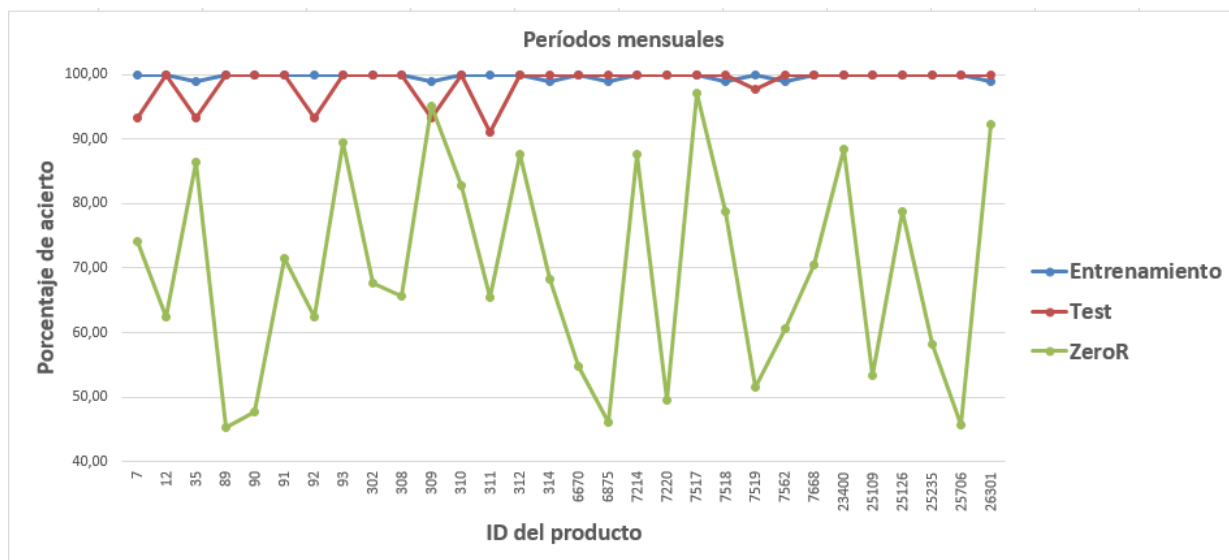


Figura 6.4: Entrenamiento vs Test vs ZeroR para períodos mensuales.

6.4.2. Recuentos por tipo de clasificador

Por cada producto y período de análisis se elige como clasificador aquel que haya alcanzado el mayor valor de *Kappa*. En la figura 6.5 se muestra como quedó la distribución de clasificadores para períodos mensuales. Por ejemplo, el clasificador *IterativeClassifierOptimizer* resultó una mejor solución en 23 productos para periodos semanales.

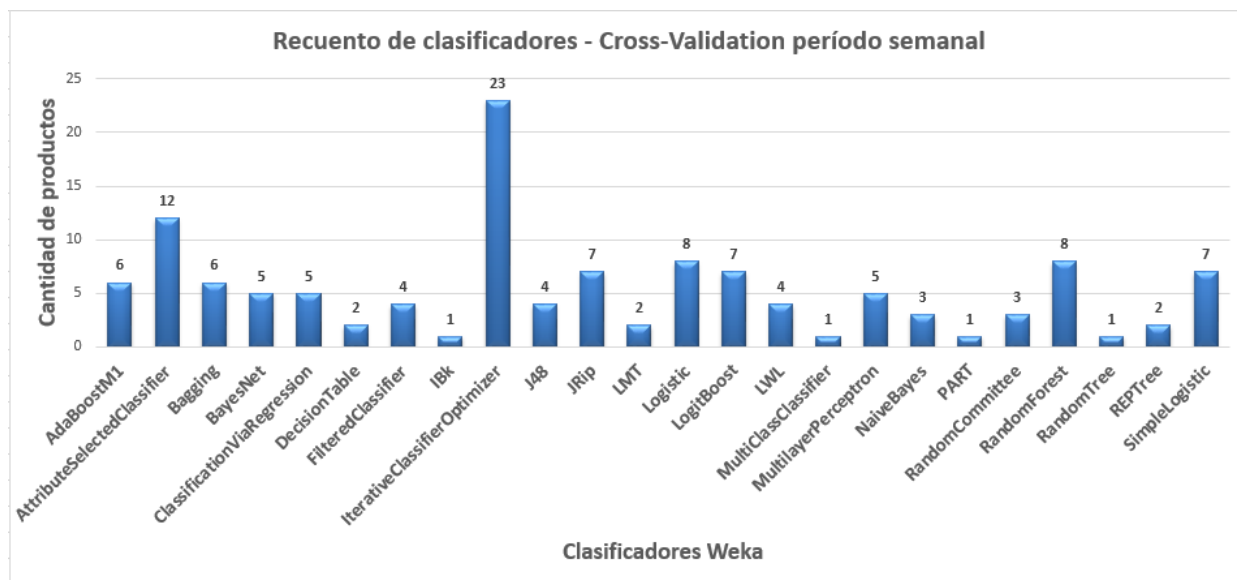


Figura 6.5: Conteo de clasificadores para períodos semanales.

En la figura 6.6 se muestra como quedó la distribución de clasificadores para períodos quincenales. Por ejemplo, el clasificador *IterativeClassifierOptimizer* resultó una mejor solución en 60 productos para periodos quincenales.

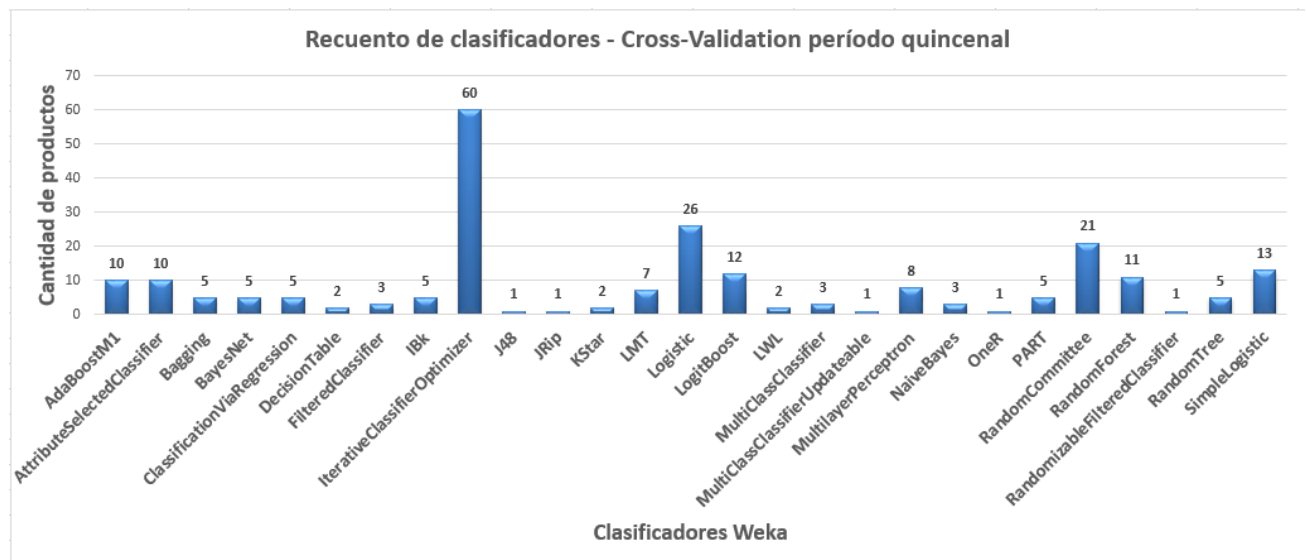


Figura 6.6: Recuento de clasificadores para períodos quincenales.

En la figura 6.7 se muestra como quedó la distribución de clasificadores para períodos mensuales. Por ejemplo, el clasificador *Logistic* resultó una mejor solución en 92 productos para periodos mensuales.

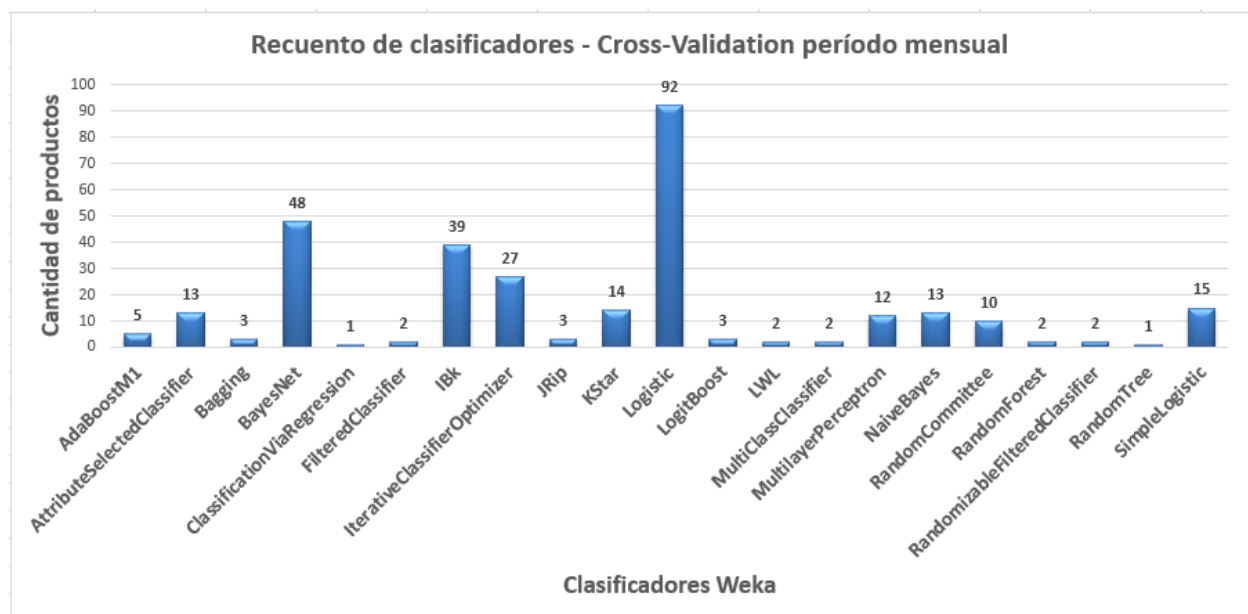


Figura 6.7: Recuento de clasificadores para períodos mensuales.

6.4.3. Porcentajes de acierto por período

Se puede observar en la figura 6.8 que la técnica propuesta alcanza altos porcentajes de aciertos en promedio, tanto para periodos mensuales, quincenales como semanales.

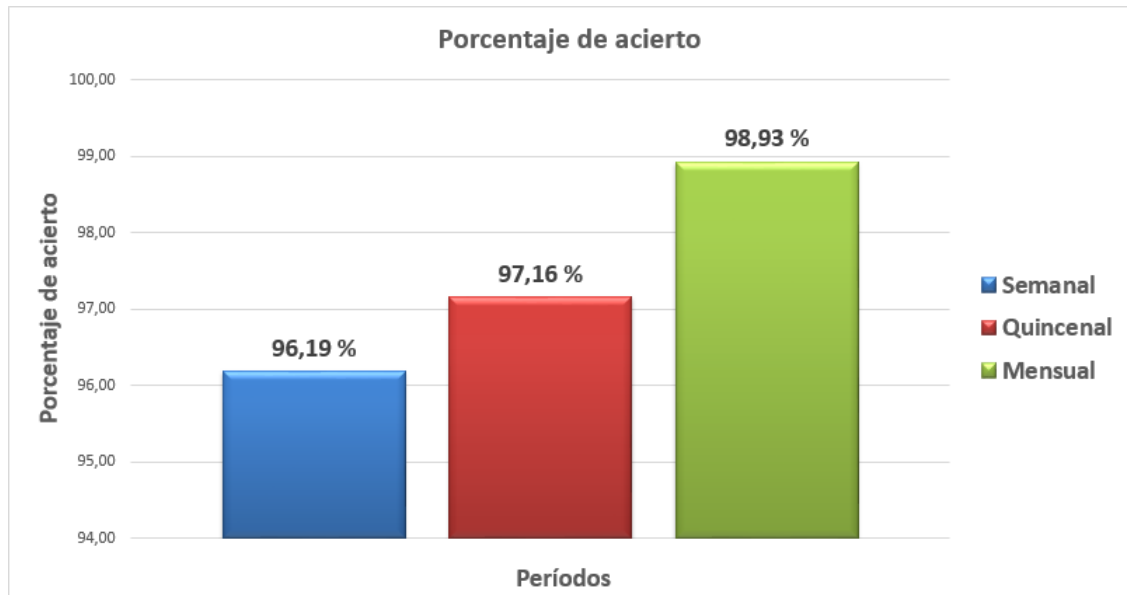


Figura 6.8: Promedio de porcentaje de aciertos para los tres períodos de análisis.

Como se trata de una prueba exhaustiva, por cada producto se intenta con todos los algoritmos de clasificación posible y se evalúa con el método Stratified 10-fold Cross Validation. Estas métricas de desempeño preliminares dan indicio de que la técnica propuesta en este trabajo puede alcanzar altos grados de confiabilidad. Obtener buenos resultados depende en gran medida de que la obtención de valores de *KPI* y la asignación de etiquetas hayan sido realizadas correctamente.

6.5. Componente de solución “Técnica de solución propuesta”

En el algoritmo 6.2 se muestra el mecanismo para obtener el pronóstico de la demanda en un ambiente de producción.

Algoritmo 6.2 Pseudocódigo para el proceso de pronóstico de la demanda.

```

for cada próximo período a pronosticar {mensual,quincenal,semanal}:
  for cada producto:
    obtener KPIs del período actual finalizado;
    ejecutar su mejor clasificador (KPIs);
    obtener etiqueta {nada,poco,medio,mucho}
    extrapolar a valores continuos(criterio experto);
  endfor;
endfor;

```

Como se mencionó anteriormente, la técnica propuesta arroja como resultado un valor discreto $y \in \{Nada, Poco, Medio, Mucho\}$. Luego en función de la etiqueta resultante, del tipo de

producto y del período seleccionado, el experto extrapola a un valor continuo que representa la cantidad en la orden de compra. El significado de las etiquetas varía en función del tipo de producto.

6.6. Discusión

6.6.1. Impacto del período de análisis

Una de las decisiones que se debe tomar es acerca del tiempo asignado al período de análisis. En este trabajo se analizaron tres períodos distintos: mensuales, quincenales y semanales con propósitos experimentales y por ser los más comunes en el ámbito comercial. En la práctica, la elección del período es una decisión estratégica a nivel gerencial que depende en gran medida del sector y tamaño de la empresa, tipos de productos y otros criterios.

En el presente trabajo, por tratarse de períodos de tiempo muy cercanos (1, 2 y 4 semanas) no se observan diferencias significativas en los porcentajes de aciertos. Otro factor a tener en cuenta es que para períodos de tiempo muy extensos (6, 12 meses) existe mayor incertidumbre en el pronóstico.

6.6.2. Impacto del etiquetado

La técnica propuesta se trata de un sistema parametrizado, donde las variables principales son el período comercial y las etiquetas seleccionadas para la clasificación. Por cuestiones de practicidad y generalidad se eligió para este trabajo un enfoque de problema de clasificación. El etiquetado proporciona mayor flexibilidad al sistema y un entorno más controlable, en comparación a un sistema de asignación de valores continuos. La flexibilidad del sistema permitió emular la opinión del experto en compras y encontrar una cantidad eficiente de etiquetas.

Capítulo 7

Conclusiones

7.1. Conclusiones Generales

Este trabajo se enfocó en proponer una nueva técnica de pronóstico de la demanda de productos, para reposición de stock en empresas retail. En el capítulo 2 se mencionó que la gestión de compras es uno de los ejes centrales en la actividad empresarial y la decisión del volumen de compras para cada producto es un desafío que enfrentan las empresas al momento de reponer el stock. Partiendo de esta premisa y analizando las técnicas de pronóstico de la demanda empleadas en la actualidad y el creciente incremento del uso de tecnologías de Business Intelligence en las organizaciones, se encontró la oportunidad de desarrollar una nueva técnica de pronóstico.

En esta nueva técnica se utilizan los Indicadores Claves de Rendimiento y apoyados en la experiencia de un experto en compras (gerente o encargado de compras) se realiza la experimentación utilizando algoritmos de clasificación de Machine Learning.

De acuerdo a los resultados experimentales se obtuvieron altas tasas de aciertos, haciendo pruebas exhaustivas con varios algoritmos de clasificación y validando con un método ampliamente aceptado. Es importante considerar que para una implementación real es primordial tener en cuenta los siguientes aspectos: el etiquetado multiclase puede ser reemplazado por valores continuos; el experto en compras debe realizar o guiar todo el proceso de etiquetado ya sea con valores continuos o discretos.

Para periodos de análisis cortos se obtuvieron buenos resultados, pero también podrían considerarse periodos de tiempo mas extensos; el conjunto de KPI seleccionados fueron en base a las limitaciones de la base de datos, es importante incluir otros KPI que consideren más aspectos como costos de inventario, marketing, clientes y otros.

Como se mencionó en la introducción, para el sector retail la mayoría de las empresas puede pronosticar la demanda total de todos los productos como un grupo con errores menores al 5 %, no obstante el pronóstico de la demanda de un producto puede generar errores considerablemente mayores con las técnicas vigentes. Con la técnica propuesta se logran errores menores al 5 % por cada producto. Se pretende que esta técnica se convierta en una herramienta de apoyo en la toma de decisiones del gerente de compras en el proceso de reposición de stock.

7.2. Aporte de este Trabajo

Este trabajo de grado brinda los siguientes resultados:

- Definición de KPI adaptados para la medición del rendimiento de productos individuales.
- Esquema básico y automatizado para el etiquetado de las instancias.
- Integración de las tecnologías de Business Intelligence centrado en los Indicadores Claves de Rendimiento y Machine Learning basado en aprendizaje supervisado.
- Herramienta de pronóstico de la demanda de productos para reposición de stock.

7.3. Trabajos Futuros

Con el propósito de futuras mejoras del presente trabajo, a continuación se citan una serie de propuestas:

- Incorporar mas KPI referentes a otros procesos de negocios que inciden en la venta de un producto.
- Asignación por parte del experto en compras de valores continuos a la clase de las instancias.
- Etiquetado basado en clustering (aprendizaje no supervisado) con la aprobación del experto en compras.
- Optimizar los algoritmos de aprendizaje mediante el ajuste de los parámetros (*Tuning parameters*).
- Incluir costos asociados a un producto (costos de almacenamiento, seguro, mantenimiento, etc.).
- Desarrollar un software SaaS (Software as a Service) que provea un servicio de pronóstico de la demanda.

Bibliografía

- [AAS14] Lakshmi Anusha, Swati Alok, and Ashiff Shaik. Demand forecasting for the indian pharmaceutical retail: A case study. *Journal of Supply Chain Management Systems*, 3(2):1, 2014. 15
- [AE91] Everett Adam and Ronald Ebert. *Administración de la producción y las operaciones: conceptos, modelos y funcionamiento*. Prentice Hall, 1991. 5
- [AKA91] David Aha, Denis Kibler, and Marc Albert. Instance-based learning algorithms. *Machine Learning*, 6:37–66, 1991. 38, 43
- [Alv13] Marcos Alvarez. *Cuadro de Mando Retail*. Profit, 2013. 1, 26, 61
- [AMS96] Christopher Granger Atkeson, Andrew Moore, and Stefan Schaal. Locally weighted learning. *AI Review*, 1996. 43
- [ASW⁺11] David Anderson, Dennis Sweeney, Thomas Williams, Jeffrey Camm, and Kipp Martin. *Métodos cuantitativos para los negocios*. © D.R. 2011 por Cengage Learning Editores, S.A. de C.V., una compañía de Cengage Learning, Inc, 11 edition, 2011. ix, ix, 1, 2, 7, 8, 9, 10, 16
- [atUoW] Machine Learning Group at the University of Waikato. Weka 3: Data mining software in java. 42
- [AWA] A.M. TURING AWARD. Alphabetical listing of a.m. turing award winners. <https://amturing.acm.org/alphabetical.cfm>. 31
- [AY16] Mustafa Akpınar and Nejat Yumusak. Year ahead demand forecast of city natural gas using seasonal time series methods. *Energies (19961073)*, 9(9):1 – 17, 2016. 15
- [Bar12] David Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012. 38
- [BBSE10] Lucian Busoniu, Robert Babuska, Bart De Schutter, and Damien Ernst. *Reinforcement Learning and Dynamic Programming Using Function Approximators*. CRC Press, Inc., Boca Raton, FL, USA, 1st edition, 2010. 40

- [BFH⁺16] Remco Bouckaert, Eibe Frank, Mark Hall, Richard Kirkby, Peter Reutemann, Alex Seewald, and David Scuse. *WEKA Manual for Version 3-8-0*, April 2016. 43
- [Big12] Milos Bigovic. Demand forecasting within montenegrin tourism using box-jenkins methodology for seasonal arima models. *Tourism and Hospitality Management*, 18(1):1 – 18, 2012. 15
- [BJ73] George Box and Gwilym Jenkins. Some comments on a paper by chatfield and prothero and on a review by kendall. *Journal of the Royal Statistical Society. Series A (General)*, 136(3):337–352, 1973. 13
- [BR18] Petr Berka and Jan Rauch. Machine learning and association rules. 09 2018. 39
- [Bre96] Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996. 43
- [Bre01] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. 45
- [Can07] Josep Lluís Cano. *Busines Intelligence: Competir con información*. ESADE, Banesto, Banesto Pyme, 2007. ix, 17, 18, 19, 20, 24
- [Cas16] Juan Castro. Aplicacion de business intelligence basado en un modelo de pronostico de series de tiempo para mejorar el volumen de ventas de la empresa fsrl de lima — castro — revista de investigacion business intelligence. 2016. 54
- [CC10] Josep Curto and Jordi Conesa. *Introducción al Business Intelligence*. Editorial UOC, 2010. 1, 17, 19, 22, 23, 24
- [CCCC08] Matthieu Cord, Pdraig Cunningham, Matthieu Cord, and Pdraig Cunningham. *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval (Cognitive Technologies)*. Springer-Verlag TELOS, Santa Clara, CA, USA, 1 edition, 2008. 34
- [Cha06] Stephen Chapman. *Planificación y control de la producción*. Pearson educación. Pearson Educación, 2006. 5
- [Coh60] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960. 50
- [Coh95] William Cohen. Fast effective rule induction. In *Twelfth International Conference on Machine Learning*, pages 115–123. Morgan Kaufmann, 1995. 44
- [CT95] John Cleary and Leonard Trigg. K*: An instance-based learner using an entropic distance measure. In *12th International Conference on Machine Learning*, pages 108–114, 1995. 43

- [DBK95] Thomas Dietterich and Eun Bae Kong. Machine learning bias, statistical bias, and statistical variance of decision tree algorithms. 255, 03 1995. 52
- [DDM15] Amelia. De Diego Morillo. *Gestión de pedidos y stock*. Paraninfo, 2015. 6
- [DKKAK11] Ozgur Demir-Kavuk, Mayumi Kamada, Tatsuya Akutsu, and Ernst-Walter Knapp. Prediction using step-wise l1, l2 regularization and feature selection for small data sets with large number of features, 2011. 38
- [EH05] Wayne Eckerson and Cindi Howson. Enterprise business intelligence: Strategies and technologies for deploying bi on an enterprise scale tdwi report series. 2005. 18, 25
- [FHB05] Bent Flyvbjerg, Mette Skamris Holm, and Søren Buhl. How (in)accurate are demand forecasts in public works projects?: The case of transportation. *Journal of the American Planning Association*, 71(2):131–146, 2005. 2
- [FHP03] Eibe Frank, Mark Hall, and Bernhard Pfahringer. Locally weighted naive bayes. In *19th Conference in Uncertainty in Artificial Intelligence*, pages 249–256. Morgan Kaufmann, 2003. 43
- [FHT98] Jerone Friedman, Trevor Hastie, and Robert Tibshirani. Additive logistic regression: a statistical view of boosting. Technical report, Stanford University, 1998. 44
- [FHT01] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001. ix, 51, 52, 53
- [FHW11] Eibe Frank, Mark Hall, and Ian Witten. *Data Mining - Practical Machine Learning Tools and Techniques - Third Edition*. Copyright © 2017 Elsevier Inc. All rights reserved, tercera edition, 2011. 41, 46, 47
- [FHWP16] Eibe Frank, Mark Hall, Ian Witten, and Christopher Pal. *Data Mining - Practical Machine Learning Tools and Techniques - Fourth Edition*. 4ta edition, 2016. 43
- [FLSL15] Kris Johnson Ferreira, Bin Hong Alex Lee, and David Simchi-Levi. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 18(1):69–88, 2015. 55
- [FS96] Yoav Freund and Robert Schapire. Experiments with a new boosting algorithm. In *Thirteenth International Conference on Machine Learning*, pages 148–156, San Francisco, 1996. Morgan Kaufmann. 43

- [FW98] Eibe Frank and Ian Witten. Generating accurate rule sets without global optimization. In J. Shavlik, editor, *Fifteenth International Conference on Machine Learning*, pages 144–151. Morgan Kaufmann, 1998. 44
- [FWI⁺98] Eibe Frank, Yong Wang, Stuart Inglis, Geoffrey Holmes, and Ian Witten. Using model trees for classification. *Machine Learning*, 32(1):63–76, 1998. 43
- [Gar06] Gartner. Glosario de gartner, www.gartner.com, enero 2006. gartner es una consultora internacional especializada en tecnologías de información y comunicación, January 2006. 17
- [GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>. 39
- [Gio17] Marcello Giordano. Pronostico de la demanda - meetlogistics. <https://meetlogistics.com/demand-planning/pronostico-de-la-demanda/>, 2017. (Accessed on 01/16/2018). 1, 2
- [GR12] John Gantz and David Reinsel. The digital universe in 2020: Big data , bigger digital shadows, and biggest grow th in the far east. *IDC iView*, 2012. 32
- [Hay11] Simon Haykin. *Neural Networks and Learning Machines*. Pearson Education, 2011. 39
- [HGS17] Jakob Huber, Alexander Gossmann, and Heiner Stuckenschmidt. Cluster-based hierarchical demand forecasting for perishable goods. *Expert Systems with Applications*, 76:140 – 151, 2017. 14
- [HH08] Frederick Hillier and Mark Hillier. *Métodos cuantitativos para administración*. Tercera edition, 2008. 7, 8, 9, 11, 13, 16
- [Ho98] Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998. 44
- [Hol93] Robert Holte. Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, 11:63–91, 1993. 44
- [HPK11] Jiawei Han, Jian Pei, and Micheline. Kamber. *Data Mining: Concepts and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science, 2011. 51
- [HS16] Matthew Hausknecht and Peter Stone. On-policy vs. off-policy updates for deep reinforcement learning. In *Deep Reinforcement Learning: Frontiers and Challenges, IJCAI Workshop*, July 2016. 40

- [HSD01] Geoff Hulten, Laurie Spencer, and Pedro Domingos. Mining time-changing data streams. In *ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, pages 97–106. ACM Press, 2001. 44
- [HT98] Trevor Hastie and Robert Tibshirani. Classification by pairwise coupling. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Advances in Neural Information Processing Systems*, volume 10. MIT Press, 1998. 43
- [htt] <http://weka.sourceforge.net/doc.dev/weka/classifiers/Classifier.html>. Interface classifier. 43
- [Iko12] Elena Ikonomovska. *Algorithms for Learning Regression Trees and Ensembles on Evolving Data Streams: Doctoral Dissertation*. E. Ikonomovska, 2012. 38
- [Inm92] William H. Inmon. *Building the datawarehouse*. QED Press, 3ra. edition, 1992. 22
- [JL95] George John and Pat Langley. Estimating continuous distributions in bayesian classifiers. In *Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 338–345, San Mateo, 1995. Morgan Kaufmann. 43
- [JLF12] Fraser Johnson, Michiel Leenders, and Anna Flynn. *Administración de compras y abastecimientos*. McGraw-Hill / Interamericana Editores, S.A., 14 edition, 2012. 1, 5, 7, 9, 16
- [JMF99] Anil Jain, Narasimha Murty, and Patrick Flynn. Data clustering: A review. *ACM Comput. Surv.*, 31(3):264–323, September 1999. 39
- [JWHT14] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated, 2014. 35
- [KACN03] Elias Kalapanidas, Nikolaos Avouris, Marian Craciun, and Daniel Neagu. Machine learning algorithms: a study on noise sensitivity. In *Proc. 1st Balcan Conference in Informatics*, pages 356–365, 2003. 51
- [KH03] David. Kroenke and Ana Hernández. *Procesamiento de bases de datos: fundamentos, diseño e implementación*. Pearson Educación, 2003. 23
- [KHDM98] Josef Kittler, Mohamad Hatef, Robert PW Duin, and Jiri Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998. 44
- [Kim92] Ralph Kimball. *The datawarehouse Toolkit*. John Wiley & Sons, Inc, 1992. 20, 22, 57, 59

- [Koh95a] Ron Kohavi. The power of decision tables. In *8th European Conference on Machine Learning*, pages 174–189. Springer, 1995. 44
- [Koh95b] Ron Kohavi. *Wrappers for Performance Enhancement and Oblivious Decision Graphs*. PhD thesis, Stanford University, Department of Computer Science, Stanford University, 1995. 44
- [Kot07] Sotiris Kotsiantis. Supervised machine learning: A review of classification techniques. In *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*, pages 3–24, Amsterdam, The Netherlands, The Netherlands, 2007. IOS Press. 35
- [KSBM01] Sathiya Keerthi, Shirish Shevade, Chiranjib Bhattacharyya, and KRK Murthy. Improvements to platt’s smo algorithm for svm classifier design. *Neural Computation*, 13(3):637–649, 2001. 43
- [Kun04] Ludmila Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley and Sons, Inc., 2004. 44
- [lCvH92] Saskia le Cessie and Hans van Houwelingen. Ridge estimators in logistic regression. *Applied Statistics*, 41(1):191–201, 1992. 43
- [Le15] Quoc Le. A tutorial on deep learning. 2015. 39
- [LHF05] Niels Landwehr, Mark Hall, and Eibe Frank. Logistic model trees. 95(1-2):161–205, 2005. 43, 45
- [MA03] Larissa Moss and Shaku Atre. *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-support Applications*. Addison-Wesley information technology series. Addison-Wesley, 2003. 17
- [MA16] Soukayna Mouatadid and Jan Adamowski. Using extreme learning machines for short-term urban water demand forecasting. *Urban Water Journal*, 14(6):630–638, 2016. 53, 55
- [McC59] John McCarthy. Programs with common sense. *Computer Science Department - Stanford University*, 1959. 31
- [MHC⁺16] Rafael Gomes Mantovani, Tomas Horvath, Ricardo Cerri, Joaquin Vanschoren, and André Carlos Ponce de Leon Ferreira de Carvalho. Hyper-parameter tuning of a decision tree induction algorithm. In *2016 5th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 37–42, Oct 2016. 53
- [Min74] Marvin Minsky. A framework for representing knowledge. *MIT- AI Laboratory Memo 306*, 1974. 31

- [Mit97] Tom Mitchell. *Machine Learning*. McGraw-Hill International Editions. McGraw-Hill, 1997. 33, 38, 51, 52
- [MN98] Andrew McCallum and Kamal Nigam. A comparison of event models for naive bayes text classification. In *AAAI-98 Workshop on 'Learning for Text Categorization'*, 1998. 43
- [MP87] Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry, Expanded Edition*. The MIT Press, 1987. 31
- [PB10] Mónica Míguez Pérez and Ana Isabel Bastos Boubeta. *Introducción a la gestión de stocks: El proceso de control, valoración y gestión de stocks*. Ideaspropias Editorial, 2da. edition, 2010. 6
- [PBP15] Michal Patak, Lenka Branska, and Zuzana Pecinova. Demand forecasting in retail grocery stores in the czech republic. *International Multidisciplinary Scientific Conference on Social Sciences & Arts SGEM*, page 693, 2015. 15
- [Pla98] John Platt. Fast training of support vector machines using sequential minimal optimization. In B. Schoelkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*. MIT Press, 1998. 43
- [PMAR07] Arley Pérez, Alberto Medina, Pavel Alonso, and Nguyen Ramírez. Métodos y técnicas para la previsión de la demanda. *Universidad de Matanzas Camilo Cienfuegos - Facultad Industrial-Economía*, 2007. 6, 9
- [Pug16] Jean Francois Puget. What is machine learning?, May 2016. 33
- [Qui93] Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 1993. 45
- [RGS02] Nils Rasmussen, Paul Goldy, and Per Solli. *Financial Business Intelligence: Trends, Technology, Software Selection, and Implementation*. Wiley, 2002. 27
- [RN04] Stuart Russell and Peter Norvig. *Inteligencia Artificial. Un Enfoque Moderno. Segunda Edición*. PEARSON EDUCACIÓN, S.A., 2004. ix, 30, 33, 34
- [Sam59] Arthur Samuel. Some studies in machine learning using the game of checker. *IBM Journal* 3, 211-229, 1959. 33
- [SB98a] Daniel Sipper and Roberto Bulfin. *Planeación y control de la producción*. McGraw-Hill, 1998. 2, 5
- [SB98b] Richard Sutton and Andrew Barto. *Reinforcement Learning: An Introduction*. A Bradford Book. MIT Press, Cambridge, Massachusetts., 1998. 36, 40

- [Sch08] Swain Scheps. *Business Intelligence For Dummies*. For Dummies, 2008. 17
- [Sci16] Elite Data Science.com. *The Machine Learning Masterclass*. 2016. 72
- [Sew07] Martin Sewell. Ensemble methods. 2007. 39
- [SFH05] Marc Sumner, Eibe Frank, and Mark Hall. Speeding up logistic model tree induction. In *9th European Conference on Principles and Practice of Knowledge Discovery in Databases*, pages 675–683. Springer, 2005. 43, 45
- [SLH14] Yuehjen Shao, , Chi-Jie Lu, and Chia-Ding Hou. Hybrid soft computing schemes for the prediction of import demand of crude oil in taiwan. *Mathematical Problems in Engineering*, 2014. 56
- [SNN18] Sima Siami-Namini and Akbar Siami Namin. Forecasting economics and financial time series: Arima vs. lstm. 2018. 12
- [SS17] Ergun Sevgi and Sahin Suleyman. Literature survey about demand forecasting in industry. *Ulakbilge: Sosyal Bilimler Dergisi, Vol 5, Iss 10, Pp 469-487 (2017)*, (10):469, 2017. 14
- [SV08] Alex Smola and SVN Vishwanathan. *Introduction to Machine Learning*. The Press Syndicate of The University of Cambridge, 2008. 1, 31, 41
- [SVPM14] Carlos Sorzano, Javier Vargas, and Alberto Pascual-Montano. A survey of dimensionality reduction techniques. *CoRR*, abs/1403.2877, 2014. 39
- [Tur50] Alan Turing. Computing machinery and intelligence. *Mind* 49: 433-460, 1950. 30
- [VAH11] Naim Caba Villalobos, Oswaldo Chamorro Altahona, and Tomás José Fontalvo Herrera. *Gestión de la Producción y Operaciones*. 2011. 7, 16
- [Wan05] Lipo Wang. *Support Vector Machines: Theory and Applications (Studies in Fuzziness and Soft Computing)*. Springer-Verlag, Berlin, Heidelberg, 2005. 40
- [Wat06] Hugh James Watson. Recent developments in datawarehousing: A tutorial. 2006. 22
- [Wol92] David Wolpert. Stacked generalization. *Neural Networks*, 5:241–259, 1992. 44
- [Wre06] Robert Wrembel. *Data Warehouses and OLAP: Concepts, Architectures and Solutions: Concepts, Architectures and Solutions*. Gale virtual reference library. IRM Press, 2006. 24