



UNIVERSIDAD NACIONAL DE ASUNCIÓN
Facultad Politécnica

Técnica de pronóstico de la demanda basada en Business Intelligence y Machine Learning

Alberto Garcete
Raúl Benítez
Ingeniería en Informática

Asesores:
Ph.D. Diego Pinto
Ing. Aditardo Vázquez

CLEI/JAIIO 2017 – Córdoba – Argentina



Objetivo propuesto

Diseñar una nueva técnica de pronóstico de la demanda para la toma de decisión en la reposición de stock utilizando herramientas de Business Intelligence y Machine Learning.

El problema del pronóstico

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.

Técnicas de pronóstico de la demanda

1. Métodos de pronósticos cualitativos

- 1.1 Opinión del Gerente
- 1.2 Junta de opinión ejecutiva
- 1.3 Consulta a la fuerza de ventas
- 1.4 Encuesta en el mercado de consumo
- 1.5 Método Delphi

2. Métodos de pronósticos cuantitativos

2.1 Métodos de series de tiempo

2.1.1 El método de pronóstico del último valor: $\text{Pronóstico} = \text{último_valor}$

2.1.2 El método de pronóstico por promedios: $\text{Pronóstico} = \text{promedio_de_todos_los_valores_hasta_la_fecha}$

2.1.3 El método de pronóstico de promedio móvil: $\text{Pronóstico} = \text{promedio_de_los_últimos_n_valores}$

2.1.4 El método de pronóstico por suavizamiento exponencial:

$$\text{Pronóstico} = \alpha (\text{último_valor}) + (1 - \alpha) (\text{último_pronóstico})$$

2.1.5 El método de suavizamiento exponencial con tendencia:

$$\text{Pronóstico} = \alpha (\text{último_valor}) + (1 - \alpha) (\text{último_pronóstico}) + \text{tendencia_estimada}$$

$$\text{tendencia_estimada} = \beta (\text{última_tendencia}) + (1 - \beta) (\text{estimación_anterior})$$

$$\text{última_tendencia} = \alpha (\text{último_valor} - \text{penúltimo_valor}) + (1 - \alpha) (\text{último_pronóstico} - \text{penúltimo_pronóstico})$$

2.2 Pronósticos causales

2.2.1 Regresión lineal: $y = a + bx$

$$y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

Business Intelligence

"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', consultas, informes, análisis y las herramientas para mostrar información (herramientas de Business Intelligence) y los procesos en el 'front end'".

Componentes de BI

- ✓ Fuentes de información
- ✓ ETL
- ✓ El Datawarehouse
- ✓ El motor OLAP
- ✓ Las herramientas de visualización

Modelado del problema

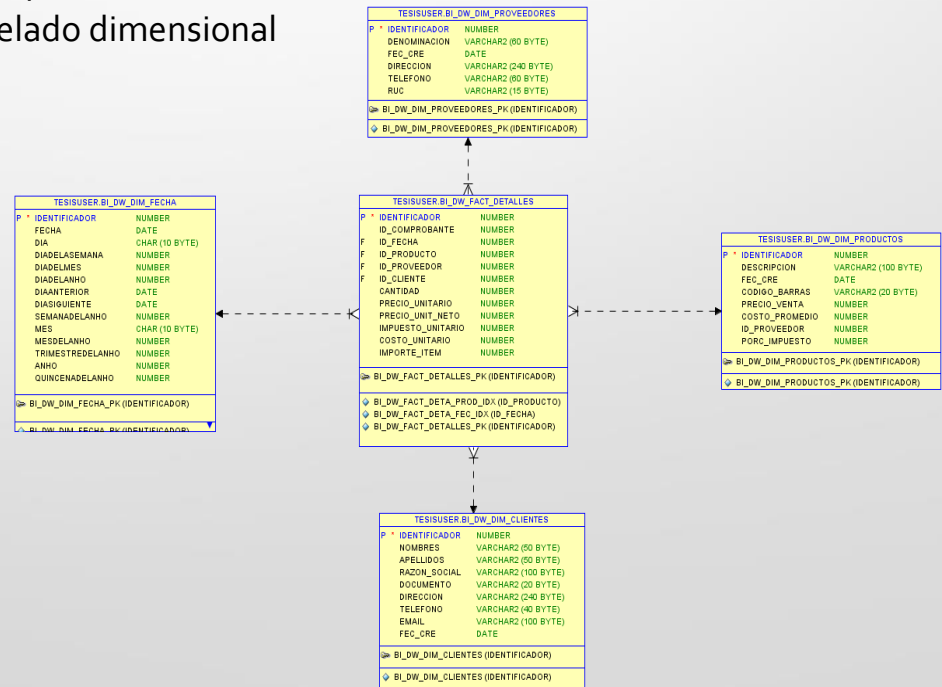
a) Base de datos relacional Oracle.

c) Registro de operaciones en el periodo de **NOV- 2013 a OCT-2016**

3. Datawarehouse: se utiliza el modelado dimensional

- Cabecera
- Detalles
- Stock.

- Fecha
- Productos
- Proveedores
- Clientes
- Cajas.



Business Intelligence

Indicadores Claves de Rendimiento

Los KPI (Key Performance Indicators) o Indicadores Clave de Rendimiento se tratan de indicadores que son decisivos para analizar de forma rápida la situación del negocio y que también facilitan la toma de decisiones

KPI Propuestos en el modelado.

1. Ticket Medio:
$$X_u = \frac{\sum (Cantidad)}{Total_Tickets_Periodo}$$
2. Cifra de Ventas:
$$X_u = \sum (Precio * Cantidad)$$
3. Margen Comercial:
$$X_u = \frac{\sum ((Precio - Costo) * Cantidad)}{\sum (Precio * Cantidad)} * 100$$
4. Rotación de Stock :
$$X_u = \frac{\sum (Total_Ventas_Periodo)}{\left(\frac{Stock_Inicial_u - Stock_Final}{2} \right)}$$
5. Coeficiente de Rentabilidad:
$$X = \left(\sum (Precio - Costo) * Cantidad \right) * Rotacion_Stock$$
6. Cobertura de Stock :
$$X_u = \frac{Stock_Actual}{Promedio_Cantidad_Venta_Ultimos_3_Periodos}$$

Business Intelligence

Asignación de etiquetas

A cada fila de valores KPI obtenidos para cada producto se asigna una etiqueta que debe ser realizada y revisada por el experto del área de compras de la empresa. 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.

Periodo : Semanal, quincenal o Mensual.

Etiquetas: [Nada, Poco, Medio, Mucho]

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

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 DE ETIQUETADO POR EL EXPERTO

aeimq	Nada	bejq	Poco	bejq	Poco
aeimr	Nada	bejnr	Poco	bejq	Medio
aeims	Nada	bejns	Nada	bejns	Poco
aeimt	Nada	bejnt	Nada	bejnt	Poco
aeimu	Nada	bejnu	Nada	bejnu	Nada
...

Resultado de la asignación de etiquetas

KPI TICKET MEDIO	KPI CIFRA VENTAS	KPI MARGEN COMERCIAL	KPI ROTACION STOCK	KPI COEF RENTABILIDAD	KPI COBERTURA STOCK	CANTIDAD	AÑO	MES	SEMANA	RESULTADO
4667	28000	12008	0.483	5797	2.571	7	2013	12	49	Mucho
4000	4000	1715	0.061	104	3.4	1	2013	12	50	Nada
4000	20000	8577	0.27	2318	4.364	5	2013	12	51	Nada
4000	16000	6862	0.211	1445	4.846	4	2013	12	52	Nada
4000	8000	3431	0.125	429	5.1	2	2013	12	53	Nada
4000	20000	8577	0.4	3431	4.091	5	2014	1	1	Medio
6000	12000	5146	0.353	1816	2.727	3	2014	1	2	Nada
4000	12000	5146	0.353	1816	2.1	3	2014	1	3	Nada
5600	28000	12008	1.077	12932	2.727	7	2014	1	4	Medio

Machine Learning

"Machine Learning es un campo de estudio que da a las computadoras la capacidad de aprender sin ser explícitamente programadas".

1. Categoría de los algoritmos

- 1.1 Aprendizaje supervisado
- 1.2 Aprendizaje no supervisado

2. Tipos de problemas

- 2.1 Regresión
- 2.2 Clasificación
- 2.3 Segmentación
- 2.4 Análisis de red

3. Problemas de clasificación

- 3.1 Clasificación binaria
- 3.2 Clasificación multiclase:

$y \in \{Nada, Poco, Medio, Mucho\}$

Machine Learning

Instancias, conjuntos de entrenamiento y teste

La entrada de un esquema de aprendizaje automático es un conjunto de instancias o ejemplos (**examples**). Estas instancias son las cosas que deben ser clasificadas, asociadas o agrupadas. Las instancias son caracterizadas mediante los valores de un conjunto predeterminado de atributos (**features**). Para el problema de estudio el proceso de Business Intelligence es quien provee las instancias. El grupo de ejemplos o instancias utilizados en el proceso de entrenamiento de los algoritmos de aprendizaje automático constituyen el conjunto de entrenamiento (**training set**). 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 (**test set**).

Machine Learning

Evaluación del aprendizaje

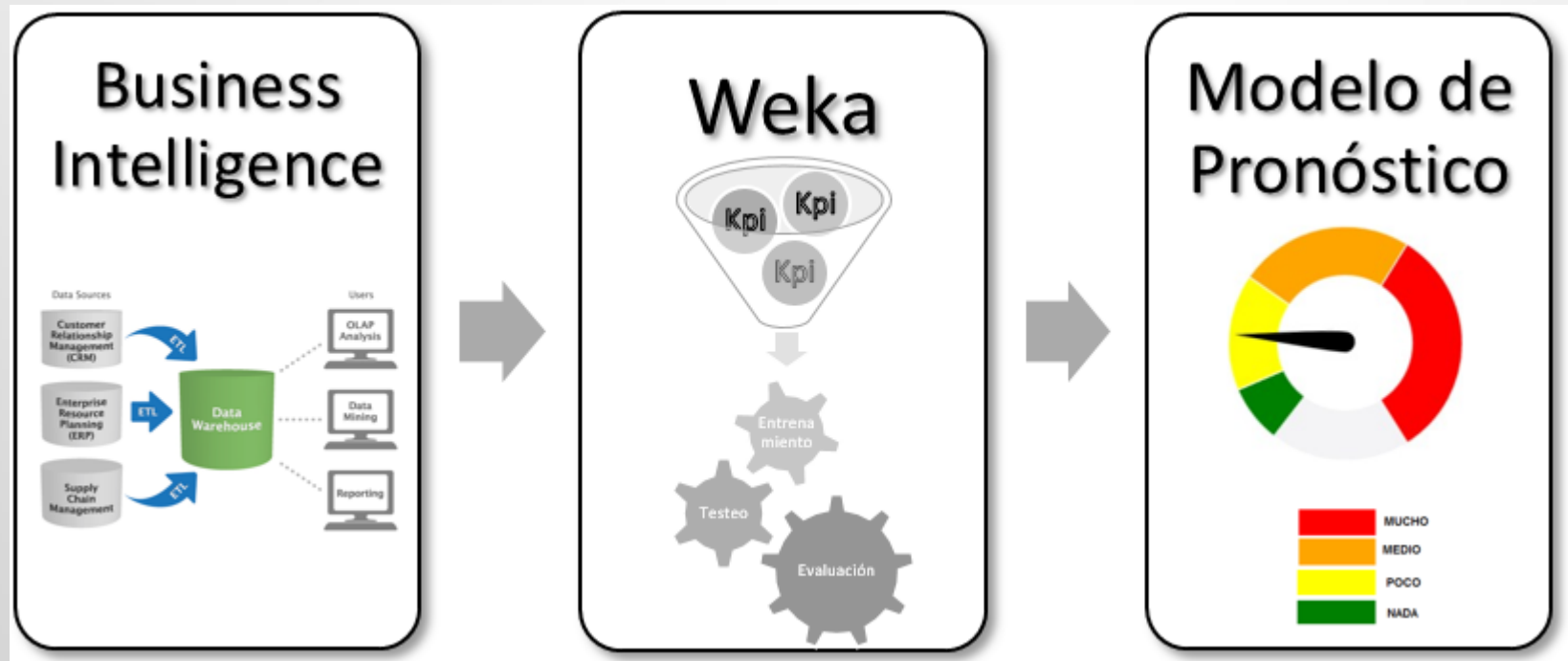
- ✓ Cross-Validation
- ✓ Stratified k-fold Cross-Validation
- ✓ Percentage Split

Métricas de desempeño

- ✓ Porcentaje de Aciertos
- ✓ Estadística Kappa:

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

Experimentación



Experimentación

Datos proveídos por Business Intelligence

- 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 definido 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 definido 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, Cobertura de Stock, Cantidad, Año, Semana. La clase de cada instancia está definido por:

$y \in \{Nada, Poco, Medio, Mucho\}$

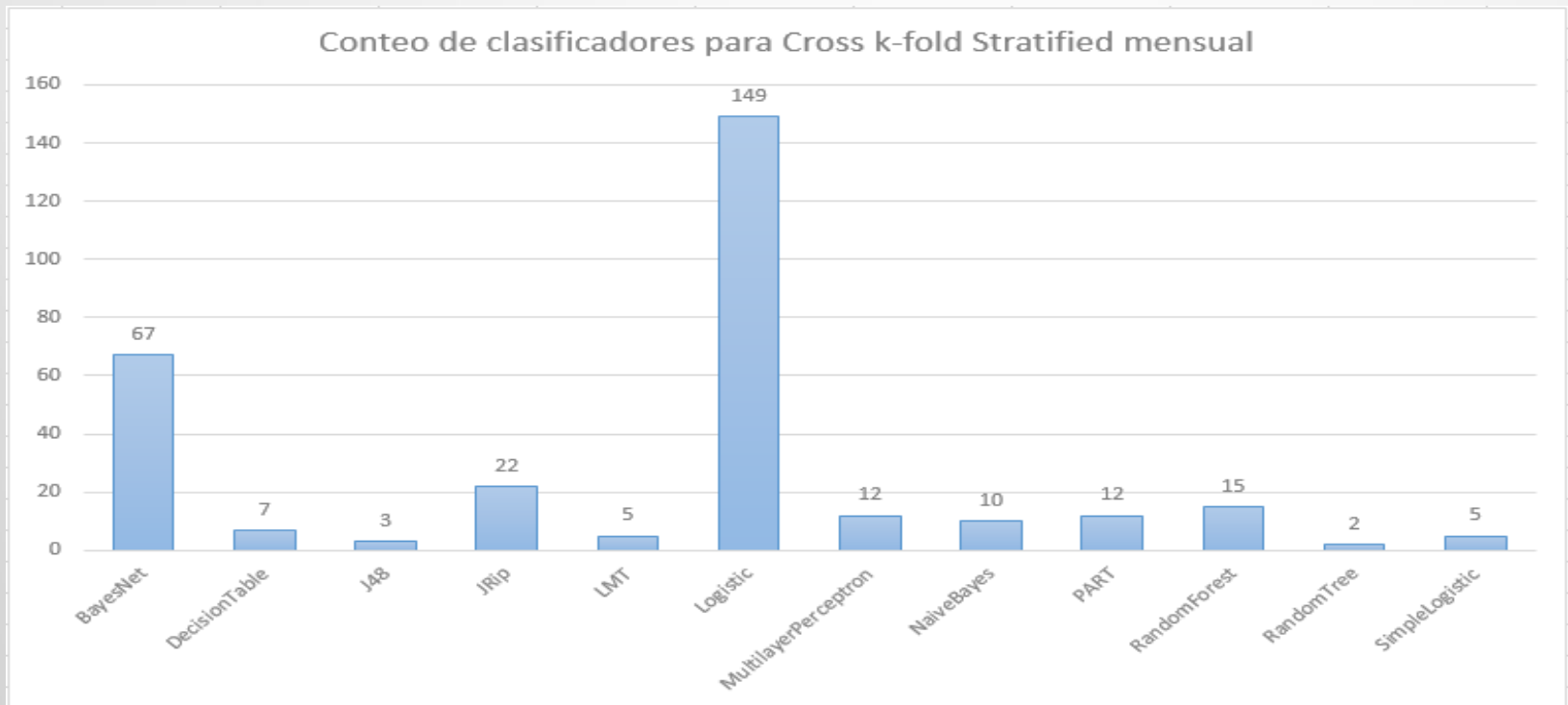
Experimentación

Entrenamiento y evaluación

```
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);  
            evaluar clasificador (conjunto de testeo);  
            obtener métricas de evaluación;  
        endfor;  
        criterios de línea de base (ZeroR o criterio experto);  
        seleccionar mejor clasificador (max(Kappa));  
        guardar clasificador;  
    endfor;  
endfor;
```

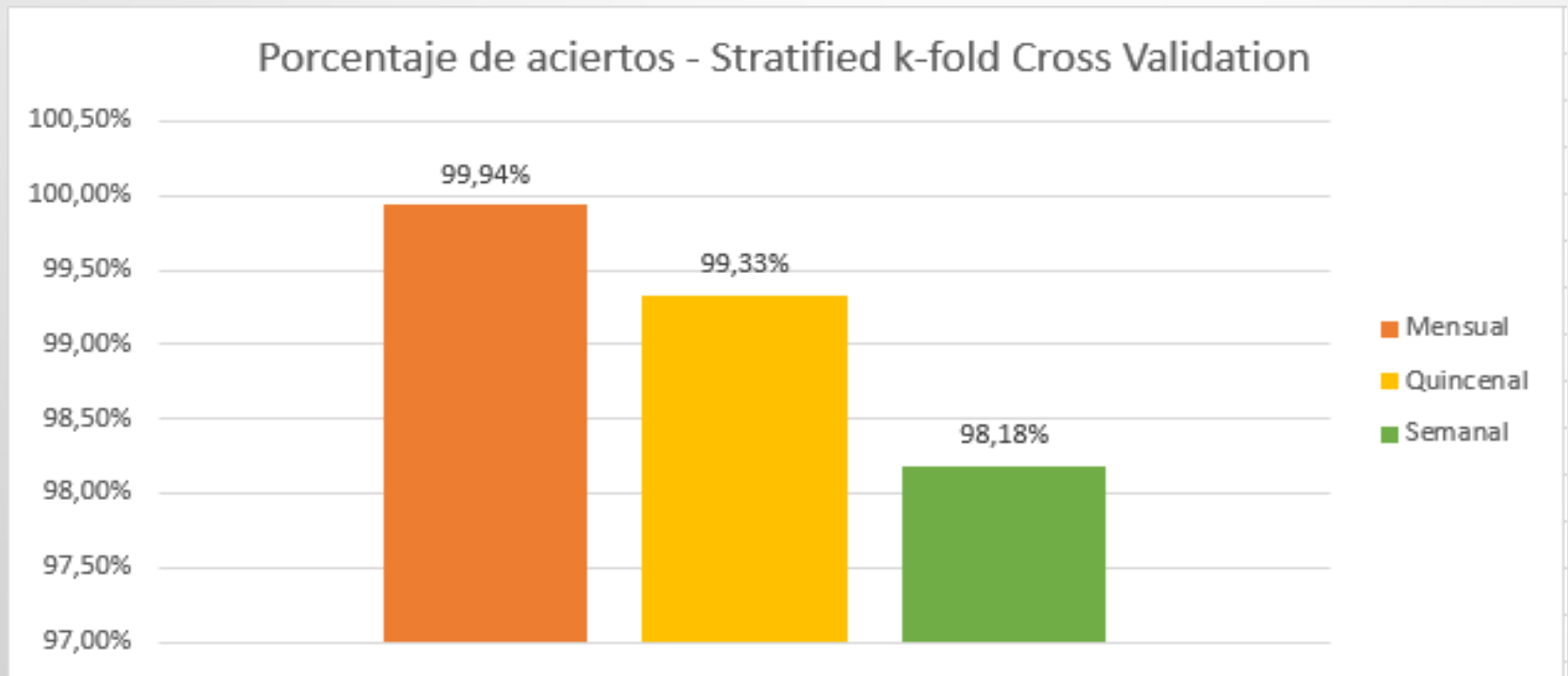
Experimentación

Clasificadores para período mensual



Experimentación

Promedio del porcentaje de aciertos



Experimentación

Cómo hacer las predicciones

```
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;
```

Discusión

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, etc.

En el presente trabajo, por tratarse de períodos de tiempo muy cercanos (1, 2 y 4 semanas) no se observan diferencias significativas en el porcentaje 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.

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”.***

Conclusiones

Este trabajo se enfocó en proponer una **nueva técnica** de estimación de la demanda de productos, para reposición de stock en empresas retail. Como se mencionó en la Sección 2, 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 el modelado 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 evaluando con un método ampliamente aceptado. La técnica propuesta pretende que este nuevo modelo se convierta en una ***“herramienta de apoyo en la toma de decisiones del gerente de compras en el proceso de reposición de stock”***.



Gracias!!