



GRADO EN ESTADÍSTICA

TRABAJO FIN DE GRADO

*Modelización estadística
de ventas
en el sector retail*

Marta Venegas Pardo

Sevilla, Noviembre de 2021

Índice general

Prólogo	III
Resumen	V
Abstract	VI
Introducción	VII
Índice de Figuras	IX
Índice de Tablas	XI
1. La ciencia de datos en el sector retail	1
1.1. Big Data en el sector retail (COMO TITULAR)	1
2. Análisis de cesta de la compra (Market basket análisis)	3
2.1. Introducción	3
2.2. Definición	3
2.3. Desarrollo	3
3. Modelos estadísticos	5
3.1. Análisis de series temporales	5
3.2. Modelo 2 (decidir)	5
4. Data science process	7
4.1. Conocimiento del negocio (Knowledge of Bussiness)	7
4.2. Adquisición de los datos (Collect the data)	7
4.3. Preparación de los datos (data preparation)	7
4.4. Análisis exploratorio de los datos (EDA)	7
4.5. Modelado	8
4.5.1. SVM	8
4.5.2. KNN	8
4.5.3. Árboles de regresión (XGBoost Model)	8
4.6. Evaluación y presentación de resultados (+análisis del error)	8
A. Apéndice: Título del Apéndice	9
A.1. Primera sección	9
B. Apéndice: Título del Apéndice	11
B.1. Primera sección	11
Bibliografía	13

Prólogo

Escrito colocado al comienzo de una obra en el que se hacen comentarios sobre la obra o su autor, o se introduce en su lectura; a menudo está realizado por una persona distinta del autor.

También se podrían incluir aquí los agradecimientos.

Resumen

Resumen. . .

Aquí comienza mi resumeeenenenneneneneneneennemmn

Abstract

Abstract...

Introduccion

Esta es la introducción de mi trabajo

- Definición

El sector retail, o también conocido como comercio minorista, es un sector económico que se encarga de proveer de bienes y servicios al consumidor final, ya que integra al grueso de comerciantes de un determinado lugar.

Este sector aglutina a comerciantes y empresas encargadas de la comercialización, ofreciendo de una gran variedad de productos y servicios a los consumidores.

(Una tienda, un supermercado, una librería, son claros ejemplos de lo que es el sector retail.)

- Objetivos
- Estimación de la demanda
- Asociación entre los dos productos

Índice de figuras

Índice de tablas

Capítulo 1

La ciencia de datos en el sector retail

Modelización estadística de ventas en el sector retail

- Objetivo: predicción de ventas, es decir, predicción de la demanda de productos

1.1. Big Data en el sector retail (COMO TITULAR)

Con R...

Capítulo 2

Análisis de cesta de la compra (Market basket análisis)

Este capítulo está por ver

2.1. Introducción

2.2. Definición

2.3. Desarrollo

Capítulo 3

Modelos estadísticos

3.1. Análisis de series temporales

3.2. Modelo 2 (decidir)

Capítulo 4

Data science process

4.1. Conocimiento del negocio (Knowledge of Business)

Problem definition (this is the introduction of my report) Definición de objetivos y metodología para cumplirlos Muy importante la definición del objetivo: VOLUMEN DE VENTAS

4.2. Adquisición de los datos (Collect the data)

Explicación de los datos, fuente, explicación de las variables,...

4.3. Preparación de los datos (data preparation)

Fase del data cleaning:

- Eliminación de duplicados (filas y columnas)
- Datos erróneos (ej: precios negativos)
- Detección de valores faltantes (¿eliminación o imputación?)
- Detección de outliers (decidir si mantener, quitar o tratar a parte)
- Unificación de variables (unificación de unidades,...)
- Creación de variables si fuera necesario

Preparación de los datos

- Reformato
- Categorización,...
- Feature selection (muy importante)

4.4. Análisis exploratorio de los datos (EDA)

- Distribución de las variables
- Relación entre las variables
- ¿Es necesario reducir la dimensionalidad?

4.5. Modelado

- Clasificación o regresión
- Librería caret
- Probar diferentes hiperparametrizaciones en los datos de entrenamiento
- Aplicar el modelo con la mejor hiperparametrización a los datos de validación y decidir un modelo final
- Evaluación del error con visualización (centrándome en los objetivos)

4.5.1. SVM

4.5.2. KNN

4.5.3. Árboles de regresión (XGBoost Model)

4.6. Evaluación y presentación de resultados (+análisis del error)

- Predicciones con el mejor modelo
- Final de la historia de una forma ordenada y resumida
- Señalar posibles mejoras y recomendaciones para proyectos futuros

Apéndice A

Apéndice: Título del Apéndice

A.1. Primera sección

Apéndice B

Apéndice: Título del Apéndice

B.1. Primera sección

Bibliografía

- JJ Allaire, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. *rmarkdown: Dynamic Documents for R*, 2021. URL <https://CRAN.R-project.org/package=rmarkdown>. R package version 2.11.
- Andrea Mesta Carmona. *Técnicas estadísticas en predicción de la demanda en el sector comercio*. Facultad de Matemáticas, Universidad de Sevilla, 2021.
- Pedro L. Luque-Calvo. *Escribir un Trabajo Fin de Estudios con R Markdown*, 2017.
- Pedro L. Luque-Calvo. *Cómo crear Tablas de información en R Markdown*, 2019.
- Francisco Coll Morales. "sector retail". Disponible en <https://economipedia.com/definiciones/sector-retail.html#:~:text=El%20sector%20retail%2C%20o%20comercio,comerciantes%20de%20un%20determinado%20lugar>.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL <https://www.R-project.org/>.
- RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA, 2015. URL <http://www.rstudio.com/>.
- Hadley Wickham, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, and Dewey Dunnington. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*, 2021a. URL <https://CRAN.R-project.org/package=ggplot2>. R package version 3.3.5.
- Hadley Wickham, Romain François, Lionel Henry, and Kirill Müller. *dplyr: A Grammar of Data Manipulation*, 2021b. URL <https://CRAN.R-project.org/package=dplyr>. R package version 1.0.7.
- Yihui Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2021. URL <https://yihui.org/knitr/>. R package version 1.36.