





# **MACHINE LEARNING**

**Fundamentos, algoritmos y aplicaciones para  
los negocios, industria y finanzas**



**JACINTO VELASCO REBOLLEDO**

# **MACHINE LEARNING**

**Fundamentos, algoritmos y aplicaciones para  
los negocios, industria y finanzas**



Madrid • Buenos Aires • México • Bogotá

© Jacinto Velasco Rebolledo, 2024

Reservados todos los derechos.

«No está permitida la reproducción total o parcial de este libro, ni su tratamiento informático, ni la transmisión de ninguna forma o por cualquier medio, ya sea electrónico, mecánico por fotocopia, por registro u otros métodos, sin el permiso previo y por escrito de los titulares del Copyright.»

Ediciones Díaz de Santos  
Internet: <http://www.editdiazdesantos.com>  
E-mail: [ediciones@editdiazdesantos.com](mailto:ediciones@editdiazdesantos.com)

ISBN: 978-84-9052-530-2  
Depósito Legal: M-17696-2024

Fotocomposición y diseño de cubiertas: P55 Servicios Culturales

Printed in Spain Impreso en España

# ÍNDICE

---

Prólogo.....	XI
<b>MÓDULO I. Conceptos generales.....</b>	<b>1</b>
1.1. Conceptos generales de aprendizaje automático .....	3
1.2. Infraestructura y tipos de analítica de un proyecto de ciencia de datos.....	5
1.2.1. Fundamentos sobre infraestructura de datos .....	5
1.2.2. Diseño de una infraestructura de datos .....	8
1.2.3. Tipos de analítica .....	13
1.3. Tipos de aprendizaje automático .....	15
1.3.1. Aprendizaje supervisado.....	16
1.3.2. Aprendizaje no supervisado.....	17
1.3.3. Aprendizaje por refuerzo.....	17
1.4. Etapas de un proyecto de aprendizaje automático.....	19
1.4.1. <i>Feature engineering</i> .....	19
1.4.2. Etapas en un proyecto de aprendizaje automático .....	20
1.5. Técnicas de optimización en aprendizaje automático.....	25
1.6. Python y buenas prácticas de programación.....	31
1.6.1. Conceptos fundamentales de Python.....	31
1.6.2. Buenas prácticas en la programación de modelos de aprendizaje automático.....	41
1.7. Análisis exploratorio de datos (EDA).....	42
<b>MÓDULO II. Aprendizaje supervisado. Clasificación.....</b>	<b>47</b>
2.1. Técnicas de validación de modelos.....	49
2.1.1. Matriz de confusión para modelos de clasificación.....	49
2.1.2. Curva ROC.....	53
2.1.3. Validación cruzada o <i>cross validation</i> .....	57
2.1.4. Entropía e índice Gini. Curvas de aprendizaje.....	59
2.1.5. Medidas de distancias .....	63
2.2. Teoría de la decisión: clasificación bayesiana.....	65
2.2.1. Teorema de Bayes.....	65
2.2.2. Algoritmo de Naive Bayes.....	67
2.3. Sistemas de clasificación básicos .....	71
2.3.1. Árbol de decisión simple .....	71
2.3.2. Algoritmo KNN.....	73

2.4. Sistemas de clasificación avanzados .....	76
2.4.1. <i>Random forest</i> .....	76
2.4.2. <i>Boosting</i> .....	80
2.5. Máquinas de soporte vectorial .....	87
2.6. Introducción a redes neuronales.....	92
2.6.1. Definición y tipos de redes neuronales .....	92
2.6.2. Redes neuronales multicapa .....	96
2.6.3. Redes neuronales convolucionales (CNN) .....	102
2.6.4. Redes neuronales recurrentes (RNN).....	106
2.7. Procesamiento del lenguaje natural (NLP) .....	108
2.8. Estrategias de resolución de problemas en modelos de clasificación .....	115
<b>MÓDULO III. Modelos de regresión y series temporales .....</b>	<b>119</b>
3.1. Validación en modelos de regresión .....	121
3.2. Regresión lineal múltiple .....	128
3.3. Regresión logística .....	134
3.4. Modelos de árboles con regresión .....	139
3.5. Regresión con vector soporte .....	144
3.6. Series temporales con aprendizaje automático .....	147
3.6.1. Introducción a las series temporales y análisis de componentes .....	147
3.6.2. Modelos ARIMA y SARIMA.....	150
3.6.3. Series temporales con redes neuronales .....	155
<b>MÓDULO IV. Aprendizaje no supervisado y por refuerzo .....</b>	<b>159</b>
4.1. Análisis de componentes principales (PCA).....	161
4.2. Modelos de aprendizaje no supervisado.....	163
4.2.1. Agrupación en clústeres: K-means .....	163
4.2.2. DBSCAN .....	167
4.2.3. Algoritmos a priori.....	169
4.2.4. Técnicas de validación de modelos no supervisados .....	173
4.3. Aprendizaje por refuerzo .....	177
<b>MÓDULO V. Casos de uso .....</b>	<b>181</b>
5.1. Caso de uso de una segmentación de prospectos .....	183
5.1.1. Definición del modelo.....	183
5.1.2. Descripción del caso de uso.....	184
5.1.3. Resolución técnica.....	184
5.1.4. Análisis de resultados .....	188



5.2. Caso de uso de un modelo de propensión.....	192
5.2.1. Definición del modelo.....	192
5.2.2. Descripción del caso de uso.....	193
5.2.3. EDA.....	194
5.2.4. Resolución técnica.....	196
5.2.5. Cálculo del tiempo de retorno.....	204
5.2.6. Distribución de nuevas probabilidades.....	208
5.2.8. Conclusiones y análisis de negocio.....	210
5.3. Análisis de series temporales para predecir el número de via- jeros en el transporte urbano.....	210
5.3.1. Descripción del caso de uso.....	210
5.3.2. Resolución técnica usando SARIMA.....	211
5.3.3. Conclusiones y análisis de negocio.....	217
5.4. Caso de uso de un modelo de abandono ( <i>churn</i> ) para explicar la pérdida de clientes en el sector financiero. Optimización de parámetros.....	218
5.4.1. Definición del modelo.....	218
5.4.2. Descripción del caso de uso.....	219
5.4.3. EDA.....	220
5.4.4. Resolución técnica.....	223
5.4.5. Conclusiones y análisis de negocio.....	229
5.5. Caso de uso para el análisis de sentimiento en redes sociales mediante el teorema de Bayes.....	231
5.5.1. Definición del modelo.....	231
5.5.2. Descripción del caso de uso.....	232
5.5.3. Resolución técnica.....	233
5.5.4. Conclusiones y análisis de negocio.....	235
<b>Bibliografía .....</b>	<b>237</b>



# PRÓLOGO

---

En la intersección entre la innovación tecnológica y las estrategias empresariales, emerge el mundo del aprendizaje automático. En la era digital, donde los datos son valorados como el nuevo petróleo y la habilidad para tomar decisiones precisas y ágiles se considera la clave del éxito empresarial, el dominio de las herramientas y técnicas del aprendizaje automático se ha vuelto imprescindible tanto para líderes empresariales consolidados como para estudiantes universitarios.

Este campo en constante evolución no solo revoluciona la forma en que las organizaciones abordan sus operaciones y estrategias comerciales, sino que también abre un vasto espectro de oportunidades para aquellos que están dispuestos a explorarlo. Desde la personalización de experiencias para los clientes hasta la detección de patrones ocultos en grandes conjuntos de datos, el aprendizaje automático se ha convertido en un motor fundamental de la innovación y el progreso en el ámbito empresarial moderno.

En este contexto, comprender los fundamentos y aplicaciones del aprendizaje automático se convierte en una ventaja competitiva crucial. Los líderes empresariales deben estar equipados con el conocimiento necesario para aprovechar al máximo el potencial de esta tecnología, mientras que los estudiantes universitarios buscan adquirir las habilidades esenciales que les permitirán prosperar en el mercado laboral del futuro.

El autor, con más de 20 años de experiencia en la aplicación de estas metodologías para resolver casos de uso en diversas áreas empresariales, desde la financiera e industrial hasta la comercial y de marketing, ha destilado su conocimiento en este libro. La obra ha sido diseñada para ser una guía completa y accesible sobre aprendizaje automático aplicado a los negocios, representando un recurso invaluable tanto para aquellos que buscan mejorar su comprensión de las complejidades del aprendizaje automático en el contexto empresarial, como para profesores que desean impartir conocimientos actualizados y relevantes en el aula universitaria.

Dentro de las páginas de este compendio, los lectores serán guiados en un fascinante viaje que parte desde los cimientos elementales hasta las sofisticadas técnicas del aprendizaje automático, todo ello expuesto mediante un lenguaje matemático comprensible por cualquier lector con formación básica en economía, finanzas o tecnología. A lo largo de este recorrido, se explorará minuciosamente cómo aplicar algoritmos de aprendizaje supervisado y no supervisado para abordar una amplia variedad de desafíos empresariales.

Desde la anticipación de la demanda hasta la optimización de la cadena de suministro, así como el análisis de sentimientos y la personalización de ex-

perencias para los clientes, este libro se convierte en un recurso esencial para aquellos que buscan ampliar su comprensión y dominio en el ámbito del aprendizaje automático aplicado al entorno empresarial.

Este texto va más allá de meras teorías abstractas; su enfoque se concentra en proporcionar a los lectores las habilidades prácticas necesarias para afrontar los desafíos cambiantes del mundo empresarial contemporáneo. Además, está meticulosamente estructurado para ser utilizado tanto por estudiantes universitarios en busca de conocimientos fundamentales como por profesionales que aspiran a integrar el aprendizaje automático en sus actividades cotidianas.

En el libro se incluyen ejemplos en Python para ayudar a comprender mejor la teoría presentada. Estos ejemplos prácticos muestran cómo aplicar los conceptos de aprendizaje automático a situaciones empresariales concretas. No solo clarifican la teoría expuesta en el texto, sino que también proporcionan a los lectores herramientas prácticas que pueden aplicar en sus propios proyectos empresariales. Al ver cómo se implementan los conceptos de aprendizaje automático en código, los lectores pueden comprender mejor cómo adaptar y aplicar estas técnicas a sus propios casos de uso específicos dentro de sus organizaciones. Esta combinación de teoría y ejemplos prácticos en Python asegura que los lectores no solo adquieran conocimientos conceptuales sobre aprendizaje automático, sino que también desarrollen habilidades prácticas que puedan utilizar para abordar desafíos empresariales reales.

En última instancia, este libro tiene como objetivo capacitar a sus lectores para que se conviertan en líderes visionarios y tomadores de decisiones informadas, capaces de aprovechar el poder del aprendizaje automático para impulsar la innovación, la eficiencia y el crecimiento en sus organizaciones y en la sociedad en general.

**DR. JACINTO VELASCO REBOLLEDO**