

Puede solicitar los archivos complementarios
de este libro escribiendo a **comercial@ediciones-eni.com**

Prólogo

| | |
|--|----|
| 1. Público objetivo y requisitos previos | 11 |
| 2. Objetivos | 12 |
| 3. Metodología | 12 |
| 4. Agradecimientos | 13 |

Capítulo 1

Introducción a la inteligencia artificial

| | |
|---|----|
| 1. Introducción | 15 |
| 2. Definición de inteligencia artificial | 16 |
| 2.1 Sistemas basados en reglas | 16 |
| 2.2 Machine Learning aprendizaje automático | 16 |
| 2.3 Deep Learning - aprendizaje profundo | 17 |
| 3. Evolución de la inteligencia artificial | 18 |
| 4. Aplicaciones concretas de la IA | 19 |

Capítulo 2

PHP y la inteligencia artificial

| | |
|--|----|
| 1. Introducción | 21 |
| 2. El lugar de PHP en la IA | 22 |
| 2.1 Facilidad de uso | 22 |
| 2.2 Librerías y frameworks | 22 |
| 2.3 Comunidad | 23 |
| 2.4 Rendimiento | 23 |
| 2.5 Ecosistema de desarrollo | 24 |

2 _____ PHP e inteligencia artificial

Conceptos, herramientas y aplicaciones

| | |
|---|----|
| 3. Librerías PHP para la IA | 24 |
| 3.1 Librerías de creación de modelos | 24 |
| 3.1.1 PHP-ML | 25 |
| 3.1.2 Rubix ML | 25 |
| 3.2 Librerías de consumo de modelos | 25 |
| 4. Configurar el entorno de desarrollo para PHP | 26 |
| 4.1 Servidor web y PHP 7.4 | 26 |
| 4.1.1 Windows | 26 |
| 4.1.2 Linux | 26 |
| 4.1.3 Mac OS | 26 |
| 4.1.4 Extensiones PHP | 27 |
| 4.2 Librerías PHP orientadas a la IA. | 27 |

Capítulo 3

Entender y preparar los datos

| | |
|--|----|
| 1. Introducción | 31 |
| 2. Tipos de datos. | 32 |
| 2.1 Datos numéricos. | 32 |
| 2.2 Datos categóricos | 32 |
| 2.3 Datos ordinales | 33 |
| 2.4 Datos textuales | 33 |
| 2.5 Datos temporales | 33 |
| 2.6 Datos geoespaciales | 33 |
| 2.7 Datos multimedia | 34 |
| 2.8 Datos lógicos/binarios | 34 |
| 3. Tipos de aprendizaje. | 35 |
| 4. Soluciones de visualización de datos. | 36 |
| 4.1 Diagrama de barras. | 37 |
| 4.2 Diagrama de tarta | 38 |
| 4.3 Diagrama de nube de puntos | 40 |
| 4.4 Diagrama de cajas | 42 |

| | | |
|-------|---|----|
| 4.5 | Mapa de calor | 45 |
| 4.6 | Diagrama de mapa geográfico | 47 |
| 5. | Recopilación, limpieza e imputación de datos | 50 |
| 5.1 | Recopilación de datos en formato CSV | 50 |
| 5.2 | Recopilación de datos en formato NDJSON | 51 |
| 5.3 | Recopilación de datos desde una base de datos SQL | 51 |
| 5.4 | Extracción/filtrado de columnas | 51 |
| 5.4.1 | Extracción de columnas | 52 |
| 5.4.2 | Filtrado de columnas | 53 |
| 5.5 | Conjuntos de datos en Rubix ML | 54 |
| 5.5.1 | Creación de un conjunto Labeled | 54 |
| 5.5.2 | Creación de un conjunto sin etiquetar | 55 |
| 5.5.3 | Métodos de descripción de conjuntos de datos | 56 |
| 5.5.4 | Métodos de selección y manipulación de conjuntos de datos | 58 |
| 5.6 | Limpieza e imputación de datos | 61 |
| 5.6.1 | Eliminación de los datos duplicados | 61 |
| 5.6.2 | Procesamiento masivo de datos con transformers | 61 |
| 5.6.3 | Gestión de datos ausentes imputación de datos | 63 |
| 6. | Exploración y análisis de datos | 70 |
| 6.1 | Análisis univariante de las características numéricas continuas | 70 |
| 6.2 | Análisis univariante de las características categóricas | 75 |
| 6.3 | Análisis multivariado | 77 |
| 6.4 | Matriz de correlaciones | 81 |
| 7. | Preprocesamiento de los datos | 82 |
| 7.1 | Transformación de datos categóricos en datos numéricos | 82 |
| 7.2 | Transformación de datos ordinales en datos numéricos | 83 |
| 7.3 | Transformar datos binarios en datos numéricos | 84 |
| 7.4 | Transformar datos textuales en datos numéricos | 84 |
| 7.5 | Transformar los datos multimedia en datos numéricos explotables | 86 |
| 7.6 | Estandarización | 87 |

4 _____ PHP e inteligencia artificial

Conceptos, herramientas y aplicaciones

| | | |
|-----|---|----|
| 7.7 | Normalización | 87 |
| 8. | Reducción de dimensiones | 88 |
| 8.1 | Análisis en componentes principales | 89 |
| 8.2 | T-SNE - t-Distributed Stochastic Neighbor Embedding | 93 |
| 9. | En resumen | 94 |

Capítulo 4

Criterios de rendimiento y de selección

| | | |
|-------|---|-----|
| 1. | Introducción | 95 |
| 2. | El reto de crear el modelo ideal | 96 |
| 2.1 | Capacidad de generalización | 97 |
| 2.2 | Sobreaprendizaje | 98 |
| 3. | Evaluación del rendimiento | 99 |
| 3.1 | División en conjuntos de entrenamiento y prueba | 99 |
| 3.2 | Validación cruzada en K partes (K-folds) | 101 |
| 4. | Métricas de evaluación del rendimiento de los modelos | 101 |
| 4.1 | Métricas de evaluación de los modelos de clasificación | 102 |
| 4.1.1 | Matriz de confusión | 102 |
| 4.1.2 | Precisión global | 105 |
| 4.1.3 | Precisión por clase de predicción | 106 |
| 4.1.4 | Recuerdo de una clase | 107 |
| 4.1.5 | Score F1 | 107 |
| 4.1.6 | Las métricas de los modelos de clasificación con Rubix ML | 108 |
| 4.2 | Métricas de evaluación de modelos de regresión | 109 |
| 4.2.1 | Error absoluto medio (Mean Absolute Error) | 111 |
| 4.2.2 | Media del cuadrado de los errores (Mean Squared Error) | 112 |
| 4.2.3 | Raíz de la media de los errores al cuadrado (Root Mean Squared Error) | 113 |

| | |
|--|-----|
| 4.2.4 El error cuadrático relativo (RSE) y el coeficiente de determinación R^2 | 114 |
| 4.2.5 Las métricas de los modelos de regresión en Rubix ML | 115 |
| 5. Aprovechamiento de las métricas de evaluación en la optimización de los hiperparámetros | 116 |

Capítulo 5

Algoritmos de aprendizaje supervisado

| | |
|---|-----|
| 1. Introducción | 119 |
| 2. Algoritmos de regresión | 120 |
| 2.1 Regresión lineal y regresión | 120 |
| 2.1.1 Regresión lineal simple | 121 |
| 2.1.2 Regresión multilínea | 121 |
| 2.1.3 Puesta en práctica en PHP con el algoritmo Ridge | 122 |
| 2.2 Árbol de regresión | 131 |
| 2.2.1 Principio de funcionamiento de un árbol de regresión | 131 |
| 2.2.2 Puntos fuertes/débiles de los árboles de regresión | 132 |
| 2.2.3 Puesta en práctica en PHP | 133 |
| 2.2.4 El gradient boosting | 143 |
| 2.3 Algoritmo de regresión por los K vecinos más cercanos | 146 |
| 2.3.1 Principios de funcionamiento de la regresión por los K vecinos más cercanos (KNN) | 146 |
| 2.3.2 Puntos fuertes/débiles de KNN | 147 |
| 2.3.3 Nociones de distancia | 148 |
| 2.3.4 Puesta en práctica en PHP | 151 |
| 2.4 Algoritmo de regresión por máquinas de vector de soporte (SVR | 158 |
| 2.4.1 Principios de funcionamiento | 158 |
| 2.4.2 Puntos fuertes/débiles | 159 |
| 2.4.3 Puesta en práctica en PHP | 160 |

6 _____ PHP e inteligencia artificial

Conceptos, herramientas y aplicaciones

| | | |
|-------|---|-----|
| 2.5 | Redes neuronales aplicadas a la regresión | 166 |
| 2.5.1 | Principios de funcionamiento | 167 |
| 2.5.2 | Funciones de activación | 173 |
| 2.5.3 | Puntos fuertes/débiles de las redes neuronales | 174 |
| 2.5.4 | Puesta en práctica en PHP | 175 |
| 3. | Algoritmos de clasificación | 183 |
| 3.1 | Regresión logística (clasificación binaria) | 184 |
| 3.1.1 | Principio de funcionamiento | 184 |
| 3.1.2 | Puntos fuertes y débiles de la regresión logística | 184 |
| 3.1.3 | Aplicación en PHP | 185 |
| 3.1.4 | One versus all | 191 |
| 3.2 | Árbol de clasificación | 196 |
| 3.2.1 | Principio de funcionamiento | 196 |
| 3.2.2 | Puntos fuertes/débiles de los árboles de clasificación. | 197 |
| 3.2.3 | Puesta en práctica en PHP | 198 |
| 3.3 | Bosque aleatorio | 203 |
| 3.3.1 | Principios de funcionamiento | 203 |
| 3.3.2 | Ventajas/desventajas en comparación con un árbol simple | 203 |
| 3.3.3 | Puesta en práctica en PHP | 204 |
| 3.4 | AdaBoost | 206 |
| 3.4.1 | Principios de funcionamiento | 206 |
| 3.4.2 | Puntos fuertes/débiles de AdaBoost | 209 |
| 3.4.3 | Ilustración de los modelos basados en árboles de clasificación | 209 |
| 3.4.4 | Puesta en práctica en PHP | 210 |
| 3.5 | SVC (clasificadores de vectores de soporte | 212 |
| 3.5.1 | Principios de funcionamiento | 212 |
| 3.5.2 | Puntos fuertes/débiles del SVC | 213 |
| 3.5.3 | Puesta en práctica en PHP | 214 |

| | | |
|-------|---|-----|
| 3.6 | Clasificador por los K vecinos más cercanos (clasificador KNN) | 221 |
| 3.6.1 | Principio de funcionamiento | 221 |
| 3.6.2 | Puntos fuertes y débiles de los algoritmos de clasificación KNN | 224 |
| 3.6.3 | Aplicación práctica en PHP | 225 |
| 3.7 | Redes neuronales de clasificación | 230 |
| 3.7.1 | Principio de funcionamiento de las redes neuronales de clasificación | 230 |
| 3.7.2 | Aplicación práctica de las redes neuronales de clasificación | 231 |
| 3.8 | Naive Bayes | 239 |
| 3.8.1 | Puntos fuertes y débiles de los modelos basados en Naive Bayes | 242 |
| 3.8.2 | Puesta en práctica con PHP | 243 |
| 4. | Resumen | 246 |

Capítulo 6

Algoritmos de aprendizaje no supervisado

| | | |
|-------|---|-----|
| 1. | Introducción | 247 |
| 2. | Algoritmos de partición de datos/agrupación | 248 |
| 2.1 | Algoritmo K-Means | 248 |
| 2.1.1 | Principio de funcionamiento del algoritmo K-Means | 248 |
| 2.1.2 | Puntos fuertes/débiles del clustering con K-Means | 250 |
| 2.1.3 | Puesta en práctica en PHP | 250 |
| 2.2 | Algoritmo DBSCAN | 259 |
| 2.2.1 | Principio de funcionamiento de DBSCAN | 259 |
| 2.2.2 | Puntos fuertes y débiles de DBSCAN | 260 |
| 2.2.3 | Aplicación práctica en PHP | 260 |

8 _____ PHP e inteligencia artificial

Conceptos, herramientas y aplicaciones

| | | |
|-------|--|-----|
| 3. | Algoritmos de detección de anomalías | 265 |
| 3.1 | El algoritmo One-Class SVM. | 265 |
| 3.1.1 | Principio de funcionamiento | 265 |
| 3.1.2 | Puntos fuertes/débiles de One-Class SVM | 266 |
| 3.1.3 | Puesta en práctica en PHP | 266 |
| 3.2 | Los bosques de aislamiento | 271 |
| 3.2.1 | Principio de funcionamiento de los bosques de aislamiento | 271 |
| 3.2.2 | Puntos fuertes y débiles de los bosques de aislamiento | 272 |
| 3.2.3 | Puesta en práctica en PHP | 272 |
| 3.3 | Algoritmo del Z-Score robusto | 277 |
| 3.3.1 | Principio de funcionamiento | 277 |
| 3.3.2 | Puntos fuertes/débiles del Z-Score robusto | 278 |
| 3.3.3 | Aplicación práctica en PHP | 279 |
| 4. | En resumen | 284 |

Capítulo 7

Persistir y reutilizar los modelos

| | | |
|----|--|-----|
| 1. | Introducción | 285 |
| 2. | Los objetos persistentes y los serializadores de Rubix ML. | 286 |
| 3. | Descubrir las codificaciones y los persistentes en Rubix ML. | 288 |
| 4. | Puesta en práctica de la copia de seguridad de un modelo. | 289 |
| 5. | Puesta en práctica de la carga de un modelo. | 293 |
| 6. | Soluciones de despliegue de modelos en producción | 294 |

Capítulo 8

Elección, optimización y selección de modelos

| | |
|--|-----|
| 1. Introducción | 301 |
| 2. Preselección de los modelos más adecuados | 302 |
| 2.1 Preselección de modelos de regresión | 304 |
| 2.2 Preselección de modelos de clasificación | 306 |
| 2.3 Estrategias de elección y de optimización de hiperparámetros | 308 |
| 2.3.1 Favorecer la optimización de los hiperparámetros más críticos | 309 |
| 2.3.2 Ajustes mediante variaciones de granularidad | 309 |
| 2.3.3 Utilizar valores aleatorios | 310 |
| 3. Selección de modelos | 311 |
| 3.1 Elección del modelo más eficaz | 311 |
| 3.2 Elección de una combinación de modelos | 311 |
| 3.3 Rechazo de modelos | 313 |
| 4. Resumen | 313 |

Capítulo 9

Utilizar modelos de terceros a través de APIs

| | |
|---|-----|
| 1. Introducción | 315 |
| 2. Creación de accesos al API OpenAI | 316 |
| 3. Consumo de un LLM para responder a una pregunta textual | 319 |
| 3.1 Generación de una respuesta a una pregunta textual sin contexto especificado | 319 |
| 3.2 Generación de una respuesta a una pregunta textual con contexto | 325 |

10 PHP e inteligencia artificial

Conceptos, herramientas y aplicaciones

- 4. Uso de un modelo de visión por ordenador junto con un LLM para recuperar información sobre imágenes 328
- 5. Uso de un modelo de transcripción automática de un archivo de audio a texto Speech-To-Text) junto con un LLM 334
 - 5.1 Transcripción automática del audio a texto 335
 - 5.2 Extracción de datos de negocio mediante un LLM 337
- 6. Creación de un archivo de audio a partir de un texto utilizando un modelo de generación de voz 338
- 7. Uso de un embedding para vectorizar texto 341
 - 7.1 Uso de embedding para transformar texto en vectores 342
 - 7.2 Pruebas sobre las similitudes semánticas entre vectores 344
- Índice 347