

# Índice

## Capítulo I

### Historia de la Inteligencia Artificial y Machine Learning.....13

- 1. Introducción..... 13
- 2. Breve historia del Machine Learning ..... 14

## Capítulo II

### Conceptos básicos sobre Machine Learning.....17

- 1. ¿Qué es el Aprendizaje Automático o Machine Learning? ..... 17
- 2. Machine Learning y Big Data ..... 18
- 3. Machine Learning, Estadística y Data Mining ..... 19
- 4. Descomponiendo un problema complejo en tareas abordables ..... 20
- 5. Distintas técnicas de análisis de datos: tipos de aprendizaje Machine Learning 21
  - Aprendizaje supervisado ..... 22
  - Aprendizaje no supervisado ..... 23
  - Aprendizaje por refuerzo..... 24
  - Aprendizaje profundo (Deep Learning) ..... 25
- 6. Algoritmos ..... 25
  - Clasificación de algoritmos según su funcionamiento ..... 26
- 7. Evaluación del modelo ..... 29
  - Elección de las métricas ..... 30
  - Elección del método ..... 33
  - Comparación de resultados: Curvas ROC..... 33

## Capítulo III

### El proceso de Machine Learning y su aplicación en Ciberseguridad ...37

- 1. El papel del Machine Learning ..... 37
- 2. El proceso de Machine Learning ..... 38

Comprensión del problema .....	38
Comprensión de los datos .....	38
Preprocesamiento .....	39
Extracción de características .....	39
Selección de características .....	39
Entrenamiento .....	40
Evaluación .....	41
Análisis de resultados .....	41
Despliegue .....	41
Aplicaciones de Machine Learning en el mundo de la Ciberseguridad .....	41
Autenticación .....	42
Detección de amenazas en Internet .....	43
Antispam y antiphishing .....	44
Detección de malware .....	45
Detección de anomalías en redes .....	45
Criptografía .....	47
Resolución automática de CAPTCHAs .....	48
Detección de anomalías en dispositivos IoT .....	49
<b>3. Metodología del libro .....</b>	<b>50</b>
<b>Capítulo IV</b>	
<b>Preprocesamiento de datos.....</b>	<b>51</b>
<b>1. Preprocesamiento .....</b>	<b>51</b>
Fases del preprocesado de datos .....	51
<b>2. Tratamiento de nulos.....</b>	<b>53</b>
<b>3. Discretización .....</b>	<b>54</b>
Métodos de discretización de la información.....	55
<b>4. Overfitting y underfitting.....</b>	<b>58</b>
Cómo evitar el overfitting y el underfitting.....	60
<b>5. Jittering .....</b>	<b>62</b>
<b>6. Anonimización .....</b>	<b>63</b>
<b>7. Extracción de características.....</b>	<b>66</b>
<b>8. Selección de características .....</b>	<b>67</b>
El proceso de Análisis de Componentes Principales (PCA).....	67
Ejemplo práctico de aplicación .....	68
<b>Capítulo V</b>	
<b>Sistemas Expertos .....</b>	<b>73</b>

<b>1. Sistema Experto</b> .....	<b>73</b>
Componentes de un Sistema Experto.....	73
CLIPS, software open source para crear sistemas expertos.....	75
<b>2. LiLaS</b> .....	<b>75</b>
<b>3. Firewalls</b> .....	<b>79</b>
Iptables, el firewall más utilizado.....	80
Preparación de las reglas del Firewall para su procesamiento por un sistema experto.....	81
Aplicaciones de análisis de los ficheros de configuración del firewall .....	85

## Capítulo VI

<b>Regresión lineal. Detección y clasificación de SPAM</b> .....	<b>87</b>
<b>1. Introducción</b> .....	<b>87</b>
<b>2. Regresión lineal</b> .....	<b>87</b>
<b>3. Regresión logística</b> .....	<b>88</b>
<b>4. Caso práctico. Detección de SPAM</b> .....	<b>89</b>
Procesado de datos y extracción de características.....	89
Aplicación de Machine Learning .....	94

## Capítulo VII

<b>Aprendizaje supervisado. Detección de documentos maliciosos</b> .....	<b>99</b>
<b>1. Introducción</b> .....	<b>99</b>
<b>2. Caso de uso. Detección de malware en documentos ofimáticos RTF</b> .....	<b>99</b>
Extracción y codificación de características .....	100
Evaluación y procesamiento previo de los datos .....	104
Selección de modelos y algoritmos.....	105
Árboles de decisión .....	105
Redes neuronales. Perceptrón Multicapa .....	109
Aplicación de modelos a un problema concreto .....	110
Etapas de la creación de modelos supervisados y tipos de datasets utilizados .....	111
Generación de modelos .....	112
Evaluación de los modelos y extracción de conclusiones.....	113

## Capítulo VIII

<b>Aprendizaje no supervisado</b> .....	<b>117</b>
<b>1. ¿Qué son los algoritmos de aprendizaje no supervisado?</b> .....	<b>117</b>
<b>2. Algoritmos de clustering: K-means</b> .....	<b>118</b>
<b>3. Implementación y aplicaciones de K-means</b> .....	<b>120</b>

Implementando K-means para la detección de escáneres de red .....	121
Obtención del conjunto de datos .....	121
Importación y representación del conjunto de datos .....	122
Inicialización de los centroides .....	125
Agrupación de los datos .....	126
Reorientación de los centroides .....	128
Resultados del experimento .....	129
<b>4. Conclusiones .....</b>	<b>134</b>

## Capítulo IX

### **Detección de anomalías.....135**

<b>1. ¿Qué es la detección de anomalías? .....</b>	<b>135</b>
<b>2. Sistemas de detección de intrusiones .....</b>	<b>136</b>
<b>3. Detección de anomalías mediante técnicas de aprendizaje automático .....</b>	<b>137</b>
<b>4. Detección de anomalías basándose en la distribución Gaussiana.....</b>	<b>138</b>
Modelo estadístico .....	140
¿Qué es una distribución Gaussiana? .....	140
Caso de uso práctico: Detectando un ransomware.....	143
Representación gráfica del conjunto de datos .....	144
División del conjunto de datos .....	146
Construcción del modelo y selección del límite .....	147
Predicción y representación gráfica de los resultados.....	149
Conclusiones del ejercicio.....	150
<b>5. Detección de anomalías basándose en Isolation Forest.....</b>	<b>150</b>
Implementación de Isolation Forest .....	152
<b>6. La detección de anomalías y el aprendizaje supervisado .....</b>	<b>154</b>

## Capítulo X

### **Visualización de Datos .....**

<b>1. Importancia de la visualización de datos .....</b>	<b>155</b>
<b>2. Visualización de datos con Python .....</b>	<b>155</b>
Matplotlib .....	156
Pandas .....	158
Seaborn.....	160
<b>3. Visualización de datos con Scilab.....</b>	<b>162</b>
<b>4. Visualización de datos con R .....</b>	<b>165</b>
Paquete estándar R .....	165
Lattice.....	166

Ggplot2.....	169
<b>5. Data StoryTelling.....</b>	<b>172</b>
¿Por qué es importante el Data Storytelling?.....	173
¿Cuáles son los elementos fundamentales del Data Storytelling?.....	175
¿Cómo construir una buena historia?.....	176
Principales herramientas de Data Storytelling.....	177
Puntos clave de las visualizaciones de datos.....	178
Cómo aprender Data Storytelling.....	178

## Capítulo XI

<b>Conclusiones .....</b>	<b>181</b>
---------------------------	------------

## Anexo I

<b>Machine Learning para mejorar el mundo .....</b>	<b>187</b>
<b>1. Temores.....</b>	<b>187</b>
<b>2. Cómo el Machine Learning puede ayudarnos a mejorar el mundo .....</b>	<b>188</b>
Medio Ambiente.....	189
Educación.....	192
Medicina .....	193
Ocio .....	194

## Anexo II

<b>¿Hacia dónde van el Machine Learning y la Ciberseguridad?.....</b>	<b>195</b>
<b>1. Tendencias .....</b>	<b>195</b>
<b>2. Las nuevas amenazas .....</b>	<b>196</b>
Malware cada vez más difícil de detectar .....	196
Accesos no autorizados.....	197
Los “básicos”: phishing, fraudes, noticias falsas etcétera.....	197
Análisis de registros robados para detectar información valiosa .....	197
Malware sin fichero.....	197
Los retos de seguridad relacionados con tecnologías Blockchain .....	198
Infraestructuras críticas: Ciberseguridad industrial .....	198
Objetivo: Internet de las Cosas .....	198
Ataques escalables con botnets inteligentes.....	198
Videojuegos móviles .....	199
Cloud Computing .....	199

## Anexo III

<b>Aprendizaje Reforzado con OpenAI y Gym .....</b>	<b>201</b>
---	------------

<b>1. Introducción.....</b>	<b>201</b>
<b>2. Profundizando un poco más en el aprendizaje reforzado, aprendizaje por refuerzo o Reinforcement Learning (RL) .....</b>	<b>202</b>
Q-Learning .....	203
<b>3. OpenAI y Gym.....</b>	<b>207</b>
Instalación .....	208
Entornos .....	209
Ejemplo de resolución de un entorno OpenAI Gym: Taxi.....	209
<b>4. Deep Q-Learning y DQN (Deep Q-Networks).....</b>	<b>216</b>
<b>5. Conclusiones .....</b>	<b>218</b>
<b>Anexo IV</b>	
<b>Kaggle.....</b>	<b>221</b>
<b>1. Introducción.....</b>	<b>221</b>
<b>2. ¿Qué es Kaggle?.....</b>	<b>221</b>
<b>3. Participando en un reto Kaggle .....</b>	<b>223</b>
Comprendiendo los datos.....	225
Aplicación de Machine Learning y análisis de resultados .....	231
Subida de predicciones a la plataforma y evaluación de modelos .....	232
<b>Bibliografía .....</b>	<b>235</b>
<b>Índice alfabético .....</b>	<b>237</b>
<b>Otros títulos de la colección .....</b>	<b>243</b>