

Contenido

PARTE I. FUNDAMENTOS DE CIENCIA DE DATOS

Capítulo 1. Ciencia de Datos: Conceptos, habilidades, herramientas y aplicaciones

Introducción	1
1.1. Definición de Ciencia de Datos ..	2
1.2. Disciplinas (Habilidades) de Ciencia de Datos	6
1.3. Dominios y Aplicaciones de Ciencia de Datos	8
1.3.1. Aplicaciones de Ciencia de Datos	9
1.4. El proceso de Ciencia de Datos: El ciclo de vida	14
1.5. El científico de datos	16
1.6. Herramientas y plataformas de la Ciencia de Datos	18
1.7. Roles profesionales relacionados con datos y Certificaciones Profesionales	20
1.8. Ética, sesgos y privacidad de los datos	22
1.9. Tendencias en Ciencia de Datos	24
Resumen	25

Capítulo 2. El ecosistema de datos: Big Data

Introducción	29
2.1. Crecimiento exponencial de datos	30
2.2. Fuentes de datos	31
2.2.1. Tipos de fuentes de datos	32
2.3. Definición, origen y evolución de Big Data	34
2.3.1. El origen moderno de Big Data	37
2.4. Taxonomía de datos	38
2.4.1. Datos estructurados	39
2.4.2. Datos no estructurados	40
2.4.3. Datos semiestructurados	40
2.4.4. Datos categóricos y numéricos: escalas de medida	41
2.5. Características de Big Data	43
2.5.1. Modelo 3 V de Big Data	43
2.5.2. Modelos 4 V y 5 V de IBM	46

2.5.3. Modelo de las 7 V	47
2.6. El tamaño de los Big Data	48
2.7. Datos en organizaciones y empresas	49
2.7.1. Datos internos	49
2.7.2. Datos externos	50
2.8. Arquitectura de Big Data	50
2.8.1. Identificación de las fuentes de datos	51
2.8.2. Recolección (ingesta) de datos	51
2.8.3. Almacenamiento de datos	52
2.8.4. Procesamiento y análisis de datos	52
2.8.5. Visualización de datos	53
2.9. Tecnologías y herramientas de Big Data	54
Resumen	55

Capítulo 3. El proceso ciclo de vida de la Ciencia de Datos y los científicos de datos

Introducción	59
3.1. Minería de Datos y Ciencia de Datos	60
3.2. Proceso tradicional de Ciencia de Datos	61
3.2.1. Comprensión del negocio	63
3.2.2. Comprensión de los datos	64
3.3.3. Preparación de los datos	64
3.3.4. Modelado de los datos	65
3.3.5. Evaluación	65
3.3.6. Despliegue	65
3.3. Preparación (preprocesamiento) de datos	66
3.3.1 Limpieza de datos	68
3.3.2. Integración de datos	71
3.3.3. Transformación de datos	72
3.3.4. Reducción de datos	73
3.3.5. Discretización de datos	74
3.4. Análisis exploratorio de datos ...	74
3.4.1 Objetivos de la exploración de datos	75
3.5. El proceso de Ciencia de Datos en equipo (TDSP) de Microsoft: ciclo de vida	76

3.6. El científico de datos: rol profesional	79
3.7. Habilidades del científico de datos	80
3.7.1. Habilidades técnicas/duras (hard skills)	80
3.7.2. Habilidades no técnicas/blandas (soft skills)	81
3.8. Herramientas de Ciencia de Datos y del científico de datos	83
3.8.1. Portal Kaggle (caso de estudio en Ciencia de Datos)	84
Resumen	86

PARTE II. ANALÍTICA DE DATOS EN CIENCIA DE DATOS

Capítulo 4. Analítica de datos

Introducción	89
4.1. ¿Qué es analítica de datos? (Data Analytics)	90
4.2. Tipos de analítica de datos	91
4.2.1. Analítica descriptiva (Descriptive Analytics)	92
4.2.2. Analítica de diagnóstico	93
4.2.3. Analítica predictiva (Predictive Analytics)	94
4.2.4. Analítica prescriptiva (Prescriptive Analytics)	95
4.3. Analítica aumentada	97
4.3.1. Democratización de los datos ..	98
4.3.2. Proveedores de soluciones de software de analítica aumentada	100
4.4. Una visión global de analítica de Big Data	100
4.5. Categorías de analítica utilizadas en la empresa	102
4.6. Tipos de análisis de datos	103
4.7. Proliferación de datos sociales ..	104
4.8. Analítica social	105
4.8.1. Métricas de social media	106
4.9. Analítica de sentimientos	107
Resumen	109

Capítulo 5. Minería de Datos

Introducción	113
--------------------	-----

Capítulo 6. Visualización de datos: informes y consultas, cuadros de mando (dashboards) y cuadro de mando integral (CMI)

Introducción	115
6.1. Conceptos generales de visualización de datos	116
6.2. Gráficos	117
6.3. Tipos de gráficos	118
6.4. Mapas	118
6.5. Infografías	124
6.6. Informes (reporting) y consultas (query)	126
6.6.1. Informes (reportes).....	126
6.6.2. Herramientas de reportes (informes)	127
6.6.3. Consultas	128
6.6.4. Informes GIS (Sistemas de Información Geográfica)	128
6.7. Cuadros de mando (dashboards)	129
6.7.1. ¿Qué es un Cuadro de Mando?	129
6.7.2. Características de un Cuadro de Mando	130
6.8. Narrativa de datos (data storytelling)	137
6.9. Cuadro de Mando Integral (CMI) / Balanced Scorecard	140
6.10. Herramientas de visualización de datos	140
Resumen	141

PARTE III. LOS PILARES TECNOLÓGICOS DE LA CIENCIA DE DATOS

Capítulo 7. Computación en la Nube y la Ciencia de Datos

Introducción	145
7.1. Computación en la Nube (Cloud Computing): Origen, definiciones y evolución	146
7.1.1. Antecedentes: el modelo tradicional de computación	146
7.1.2. Origen y evolución del a computación en la nube	147
7.2. Características de Cloud Computing	150
7.2.1. Características esenciales	151

7.2.2. Características de Computación en la Nube que facilitan su adopción	152
7.3. Modelos de Computación en la Nube	153
7.4. Modelos de servicio	154
7.4.1 Software como servicio (SAAS, Software As A Service)	155
7.4.2 Plataforma como servicio (PAAS, Platform As A Service)	155
7.4.3 Infraestructura como servicio (IAAS, Infrastructure As A Service)	155
7.5. Modelos de despliegue de la nube (tipos de nubes)	156
7.5.1. Nube privada	157
7.5.2. Nube pública	158
7.5.3. Nube híbrida	159
7.5.4. Nube comunitaria	159
7.6. La nube del futuro en organizaciones y empresas. La multinube (multicloud)	160
7.6.1. Selección del modelo de nube	163
7.7. La emergencia de la nube híbrida y la multinube	163
7.8. Proveedores de la nube	165
7.8.1. Implantación de la nube en organizaciones y empresas	169
7.8.2. Aspectos económicos	170
7.9. Seguridad en la nube: Retos y amenazas	171
7.10. Contratación de servicios de Cloud Computing	172
7.11. Los centros de datos como soportes de cloud computing	173
7.12. La computación en la Ciencia de Datos	174
Resumen	175
Capítulo 8. Computación en el borde, la nube nativa y la computación sin servidores	
Introducción	179
8.1. Modelos Edge Computing (Computación en el Borde) y Fog Computing (Computación en la Niebla)	180
8.2. Edge Computing (Computación en el Borde)	181
8.2.1. Características de la computación en el borde	182
8.2.2. Aplicaciones de computación en el borde	184
8.3. Fog Computing (Computación en la Niebla)	184
8.4. Edge Computing versus Fog Computing	185
8.5. La Computación en el Borde (Edge Computing) en 2022 y siguientes años	187
8.6. Computación sin servidor (Serverless computing)	188
8.6.1. Funcionamiento de la computación sin servidor	191
8.6.2. Función como servicio (FAAS)	191
8.6.3. Proveedores de servicios sin servidor	192
8.7. Metodología Devops	193
8.8. Microservicios: Arquitectura de microservicios	196
8.9. Contenedores	201
8.9.1. Docker	203
8.10. Orquestación de contenedores	203
8.10.1. Kubernetes	203
8.11. Orquestación de la nube	205
8.11.1. Orquestación versus coreografía	206
8.12. La nube del futuro/evolución de la nube	207
8.13. Aplicaciones nativas de la nube	209
8.13.1. ¿Qué es una aplicación nativa de la nube?	209
8.13.2. Aplicaciones tradicionales frente a aplicaciones nativas de la nube	210
8.14. La Nube Nativa (Cloud Native)	211
8.15. Cloud Native Computing Foundation (CNCF)	212
Resumen	214
Capítulo 9. Almacenamiento de datos: Data Warehouses y Data Lakes	
Introducción	219
9.1. Data Warehouse. Definición y conceptos importantes	219
9.1.1. Características de un Data Warehouse	220
9.2. Data Mart	222
9.2.1. Data Warehouse versus Data	

Mart	224	10.3. Componentes de la Inteligencia Artificial	250
9.3. Enfoques de desarrollo (modelos) de un sistema de Data Warehouse ...	224	10.4. Evolución de la Inteligencia Artificial	251
9.3.1. Modelo Inmon: modelo EDW ...	224	10.5. Inteligencia Artificial Aplicada .	253
9.3.2. Modelo Kimball: el enfoque de Data Mart	225	10.6. Taxonomía de la Inteligencia Artificial	254
9.3.3. Implantación práctica de un Data Warehouse empresarial	226	10.6.1. Tipos de Inteligencia Artificial basados en aspectos humanos y racionales	254
9.3.4. Data Warehouse 2.0 y data Warehouse en tiempo real	226	10.6.2. Tipos de Inteligencia Artificial basados en capacidades y funcionalidades de los sistemas	255
9.4. Data Lakes (Lagos de datos): los nuevos depósitos de almacenamiento (repositorio) de datos	227	10.7. Tipos de Inteligencia Artificial basados en capacidades	256
9.4.1. Origen del término Data Lake	228	10.8. Tipos de inteligencia artificial basados en funcionalidad	258
9.4.2. Definición de Data Lake	228	10.9. Aprendizaje Automático	259
9.4.3. Propiedades, ventajas e inconvenientes de los Data Lakes	229	10.9.1. Aprendizaje Supervisado	260
9.4.4. Proveedores de soluciones de Data Lake	230	10.9.2. Aprendizaje no Supervisado	260
9.5. Data Warehouse versus data Lake	230	10.9.3. Aprendizaje Semisupervisado	260
9.6. Big Data y la Inteligencia Artificial. Los pilares de los lagos de datos	234	10.9.4. Aprendizaje Reforzado	260
9.7. Data Lakehouse (almacén de lagos de datos)	236	10.10. Aprendizaje Profundo	261
9.7.1. Comparación de un data Warehouse, data Lake y data Lakehouse	237	10.11. Aprendizaje Automático frente a Aprendizaje Profundo	263
9.7.2. Arquitectura de un lago de datos y de un almacén de lagos de datos	240	10.12. El uso ético de la Inteligencia Artificial aplicada en la Unión Europea, España y Latinoamérica	265
9.8. Proveedores de Data Lake y de data Lakehouse en la nube	241	Resumen	267
Resumen	243		
Capítulo 10. Inteligencia Artificial en la Ciencia de Datos		Capítulo 11. Inteligencia Artificial del futuro: Inteligencia Artificial Conversacional e Inteligencia Artificial Generativa. El camino a la Inteligencia Artificial General (AGI)	
Introducción	245	Introducción	271
10.1. Origen de la Inteligencia Artificial	246	11.1. Procesamiento del Lenguaje Natural (NLP)	272
10.2. ¿Qué es la Inteligencia Artificial?	247	11.2. Técnicas utilizadas el análisis de NLP	274
10.2.1. Definiciones de la Unión Europea	248	11.2.1. Text Mining y Análisis de Texto	274
10.2.2. Definiciones de consultoras; Gartner de Information Technology Glossary y Mckinsey	250	11.3. Casos de uso de NLP	275
		11.4. Componentes del NLP	276

11.4.1. Comprensión del lenguaje natural (Natural Language Understanding, NLU)	277	11.13.3. Transformadores (Transformers)	297
11.4.2. NLG (Natural Language Generation) o generación de lenguaje natural	278	11.14. Modelo de lenguaje	298
11.5. Sistemas de tecnologías de voz (speech technologies)	279	11.15. GPT-3: Modelo estándar de lenguaje de referencia	299
11.5.1. Aplicaciones de los sistemas de reconocimiento del habla	279	11.16. Tipos de modelos de lenguaje	301
11.6. Inteligencia Artificial Conversacional	280	11.16.1. DALL-E 2	301
11.6.1. ¿Qué es la inteligencia artificial conversacional?	281	11.16.2. LaMDA	303
11.7. Tecnologías de Inteligencia Artificial Conversacional	281	11.16.3. PaLM	303
11.7.1. Componentes de la inteligencia artificial conversacional	282	11.16.3. ChatGPT	304
11.8. Sistemas conversacionales	283	Resumen	305
11.8.1. Arquitectura clásica de un sistema conversacional (sistema de diálogo hablado)	284	Capítulo 12. Internet de las Cosas: Los datos de los objetos inteligentes en la Ciencia de Datos	
11.9. Chatbots y asistentes virtuales	287	Introducción	309
11.10. ¿Qué es un chatbot?	288	12.1. ¿Qué es Internet de las Cosas (Internet of Things, IoT)	310
11.10.1. Tipos de chatbots según su funcionalidad	289	12.2. Definición de Internet de las Cosas	312
11.10.2. Tipos de chatbots según la interacción	290	12.3. Objetos inteligentes en entornos inteligentes	314
11.10.3. Tipos de chatbots según el canal de comunicación	291	12.4. Historia de Internet de las Cosas	315
11.11. Asistentes virtuales	291	12.5. Aplicaciones y casos de estudio de Internet de las Cosas	316
11.12. Inteligencia Artificial Generativa	284	12.5.1. Casos de estudio de IoT	317
11.12.1. Asistentes virtuales más populares	293	12.6. Gemelos digitales (digital twins)	321
11.12.2. Asistentes virtuales del futuro	294	12.6.1. Definición	321
11.12.2. Asistentes de voz y altavoces inteligentes	294	12.6.2. Historia de los gemelos digitales	321
11.13. Inteligencia virtual generativa	295	12.6.3. Funcionamiento de los gemelos digitales	323
11.13.1. Redes GAN	296	12.6.4. Aplicaciones de gemelos digitales	323
11.13.2. Codificadores automáticos variacionales (Variational Auto-Encoders, VAE)	297	12.7. Tecnologías y tendencias habilitadoras del Internet de las Cosas	324
		12.7.1. Tecnologías de carga de baterías	326
		12.7.2. Otras tecnologías disruptivas de impacto en el Internet de las cosas	327
		12.7.3. Caso de estudio: Tecnologías Habilitadoras Digitales (THD)	327
		12.8. Internet de todo y las cuatro fuentes de generación de datos	328

12.9. La capa física de la Red (sensores, actuadores, controladores)329

12.10. Ciudades inteligentes (smart cities)331

12.10.1. Ciudades inteligentes y sostenibles333

12.11. La seguridad en Internet de las cosas333

12.11.1. Estado del arte333

12.12. La privacidad en Internet de las cosas334

Resumen336

Capítulo 13. Blockchain: Seguridad y transparencia en los datos

Introducción339

13.1. Blockchain (Cadenas de bloques): La nueva revolución de Internet340

13.2. Fundamentos de Blockchain342

13.2.1. ¿Qué es Blockchain?343

13.2.2. Definición de Blockchain343

13.3. Blockchain: La visión del NIST344

13.3.1. ¿Qué es Blockchain según el NIST?344

13.4. Funcionamiento de Blockchain346

13.4.1. Proceso de las transacciones en una cadena de bloques347

13.4.2. Componentes de Blockchain: funcionamiento347

13.5. ¿Cuáles son las principales aplicaciones de la tecnología Blockchain?349

13.6. Tipos de Blockchain: pública, privada e híbrida350

13.7. Contratos inteligentes352

13.8. Trazabilidad355

13.9. Identidad digital357

13.10. Blockchain en Ciencia de Datos358

13.10.1. El Internet de las cosas: un registro de todas las cosas360

13.10.2. Aplicaciones del registro de todas las cosas360

13.11. Convergencia de Blockchain, Inteligencia Artificial e Internet de las Cosas en la Ciencia de Datos361

13.11.1. Hacia un modelo de convergencia de Blockchain-IoT-AI ... 361

13.12. Tecnologías NFT (Non Fungible Token) 364

Resumen 369

PARTE IV. SEGURIDAD, PRIVACIDAD, PROTECCIÓN Y ÉTICA DE LOS DATOS

Capítulo 14. Seguridad y ciberseguridad en Ciencia de Datos

Introducción 373

14.1. Seguridad y privacidad 374

14.2. ¿Qué es ciberseguridad? 374

14.3. La ciberseguridad en tiempo real 377

14.4. La seguridad de la información y la privacidad en la Computación en la Nube 379

14.5. Infraestructuras críticas 379

14.6. Ciclo de vida de la ciberseguridad 381

14.7. Ciberamenazas: agentes y tipos 383

14.8. Ciberataques: Herramientas utilizadas por los atacantes 384

14.9. Resiliencia 385

14.10. La web profunda, la web invisible (deep web) 386

14.11. Los ciberriesgos 387

14.12. Directiva de ciberseguridad de la Unión Europea: NIS/NIS 2 388

14.12.1. Nueva normativa NIS 2 de la Unión Europea 389

14.13. Estrategias de ciberseguridad de la UE: 2013 y 2020 389

14.13.1. Estrategias de ciberseguridad de la UE: 2013 390

14.13.2. Estrategias de ciberseguridad de la UE: 2020 390

14.13.3. Estrategias nacional de ciberseguridad 2019 de España 390

14.14. Reglamento europeo de ciberseguridad 392

14.15. Organizaciones internacionales 393

14.16. Tendencias en ciberseguridad en 2022 y años siguientes 395

Resumen 396

Capítulo 15. Ética, privacidad, protección de datos y compliance en la Ciencia de Datos: Normas legales y regulaciones, consorcios y organizaciones

Introducción	401
15.1. La ética y la responsabilidad social de las empresas en el ecosistema de Ciencia de Datos	402
15.2. Evaluación de impacto en la protección de datos	403
15.3. Reglamento de Protección de Datos y Privacidad de la Unión Europea	404
15.3.1. Aspectos más sobresalientes del reglamento vigente	404
15.3.2. Recomendaciones de la AEPD sobre el reglamento GDPR	405
15.4. El delegado de protección de datos (DPD)	406
15.5. Evaluación de impacto de la privacidad	407
15.5.1. Privacidad desde el diseño ...	407
15.6. La privacidad en la Ciencia de Datos	410
15.7. El escudo de privacidad: Unión Europea-EE.UU.	411
15.8. Compliance (cumplimiento normativo)	412
Resumen	413

PARTE V. EL FUTURO DE LA CIENCIA DE DATOS

Capítulo 16. Tendencias tecnológicas disruptivas en Ciencia de Datos en el horizonte 2025

Introducción	417
16.1. Tendencias tecnológicas estratégicas para 2023 (Gartner)	418
16.1.1. Principales tendencias tecnológicas estratégicas para 2023 (Gartner)	418
16.2 Tendencias tecnológicas en Ciencias de Datos de Forbes (Bernard Marr) 2022 y 2023	421
16.2.1. Las cinco grandes tendencias tecnológicas en ciencia de datos para 2022 según Forbes	421

16.2.2. Las cinco grandes tendencias tecnológicas en ciencia de datos para 2023 según Forbes	421
16.3 Panorámica general de tendencias tecnológicas en Ciencia de Datos para 2022-2025	423
16.4. La hiperconectividad en la Ciencia de Datos del futuro: Redes 6G y Wi-Fi 7	424
16.4.1. Futuras redes Wi-Fi: Wi-Fi 7	424
16.4.2. Futuras redes móviles 6G	424
16.5. Computación cuántica	426
16.6. El metaverso: el nuevo universo digital	427
16.7. Web 3.0	428
16.8. Los NFT en convergencia con la Ciencia de Datos	430
16.9. El futuro de las tecnologías emergentes y su convergencia con la Ciencia de Datos (Hype Cycle de Gartner, Agosto 2022)	431
16.10. El futuro del científico de datos	434
16.11. El futuro de la Ciencia de Datos	436
Resumen	437