



مقدمة في  
علم البيانات

الدكتور  
طلال ناظم الزهيري  
استاذ علم المعلومات

بغداد ٢٠٢٣

# مقدمة في علم البيانات

الدكتور

طلال ناظم الزهيري

أستاذ علم المعلومات

بغداد 2023

005.7

ز 994 الزهيري، طلال ناظم

مقدمة في علم البيانات/ طلال ناظم الزهيري

. - ط 1 . - بغداد: مطبعة زاكي، 2023.

205 ص ؛ 24 سم

1. البيانات، علم أ. العنوان

م. و.

2023/ 374

المكتبة الوطنية. / الفهرسة اثناء النشر

رقم الايداع في دارالكتب و الوثائق في بغداد

( 374 ) لسنة 2023

## قائمة المحتويات

10.....	المقدمة
12.....	الفصل الأول
12.....	المفاهيم الأساسية لعلم البيانات
12.....	تمهيد
13.....	مفهوم البيانات
14.....	أهمية وفوائد البيانات
16.....	مصادر البيانات
17.....	أنواع البيانات
19.....	تصنيف البيانات
20.....	جمع البيانات
22.....	فهم البيانات
23.....	إعداد البيانات
24.....	استكشاف البيانات
26.....	تحويل البيانات
28.....	نماذج البيانات

30.....	نشر النتائج.....
31.....	علم البيانات.....
33.....	تاريخ علم البيانات.....
35.....	أهمية علم البيانات.....
37.....	مجالات استخدام علم البيانات.....
39.....	منهجية علم البيانات.....
40.....	أدوات علم البيانات.....
42.....	تطبيق علم البيانات.....
43.....	فوائد علم البيانات.....
45.....	دور علماء البيانات.....
47.....	مهارات العمل في مجال البيانات.....
49.....	اسئلة الفصل الأول.....
50.....	الفصل الثاني.....
50.....	تحليل البيانات.....
50.....	تمهيد.....
50.....	فيئات تحليل البيانات.....

51.....	مراحل تحليل البيانات .....
54.....	اولاً: جمع البيانات.....
54.....	الاستطلاعات.....
56.....	المقابلات.....
57.....	الملاحظات.....
58.....	التجارب.....
60.....	السجلات وعمليات التتبع.....
61.....	تعقب البيانات.....
63.....	ثانياً: تنظيف البيانات.....
65.....	<i>Excel &amp; Google Sheets</i> .....
66.....	<i>OpenRefine</i> .....
67.....	<i>Trifacta</i> .....
67.....	<i>Talend</i> .....
68.....	<i>DataWrangler</i> .....
68.....	<i>RapidMiner</i> .....
69.....	<i>KNIME</i> .....
69.....	<i>Alteryx</i> .....
70.....	<i>DataRobot</i> .....
71.....	مهارات منظم البيانات.....

73.....	ثالثاً: استكشاف البيانات.....
75.....	مهارات مستكشف البيانات.....
76.....	رابعاً: تحويل البيانات.....
78.....	خامساً: نمذجة البيانات.....
79.....	سادساً: عينات البيانات.....
81.....	سابعاً: تصور البيانات.....
83.....	ثامناً: التحقق من البيانات.....
84.....	تاسعاً: تخزين البيانات.....
85.....	مهارات محلي البيانات.....
89.....	اسئلة الفصل الثاني.....
90.....	الفصل الثالث.....
90.....	البيانات الضخمة.....
90.....	تمهيد.....
91.....	مفهوم البيانات الضخمة.....
93.....	مصادر البيانات الضخمة.....
94.....	تحديات تنظيم البيانات الضخمة.....
96.....	أدوات إدارة البيانات الضخمة.....

96.....	<i>Hadoop</i>
97.....	<i>Apache Spark</i>
98.....	<i>NoSQL</i>
100.....	<i>Amazon Redshift</i>
101.....	<i>Google BigQuery</i>
103.....	<i>Google TensorFlow</i>
104.....	<i>Scikit-Learn</i>
105.....	<i>Apache Airflow</i>
107.....	<i>Apache Nifi</i>
108.....	<i>Apache Kafka</i>
111.....	تطبيقات البيانات الضخمة
111.....	الرعاية الصحية.
113.....	التمويل
115.....	البيع بالتجزئة
117.....	النقل
119.....	الطاقة
121.....	التصنيع
123.....	الاعلام
125.....	الأمن
127.....	اهمية دراسة البيانات الضخمة.

129	اسئلة الفصل الثالث
130	الفصل الرابع
130	إدارة البيانات وحوكمتها
130	تمهيد
131	إدارة قواعد البيانات
133	حوكمة البيانات
135	اخلاقيات البيانات
136	هندسة البيانات
138	تَسبب البيانات
140	مقاييس البيانات
141	أدوات إدارة البيانات
142	أولاً: منصات إدارة البيانات
143	ثانياً: أدوات جودة البيانات
145	إدارة مشاريع البيانات
146	أولاً: تحديد نطاق المشروع وأهدافه
147	ثانياً: تحديد متطلبات البيانات
148	ثالثاً: تطوير خطة المشروع
149	رابعاً: إدارة الميزانية والجدول الزمنية للمشروع

150	.....	خامساً: تنسيق واجبات الفريق
152	.....	سادساً: مراحل تقدم المشروع
153	.....	سابعاً: ضمان جودة البيانات ودقتها
154	.....	ثامناً: إدارة مخاطر المشروع
155	.....	تاسعاً: إدارة وثائق المشروع
156	.....	عوامل نجاح مشروع البيانات
158	.....	عوامل فشل مشروع البيانات
159	.....	مهارات مدير مشروع البيانات
160	.....	أهمية دراسة إدارة البيانات
163	.....	اسئلة الفصل الرابع
164	.....	الفصل الخامس
164	.....	علم البيانات: الأدوار والمسؤوليات
164	.....	تمهيد
165	.....	محلل البيانات
169	.....	مهندس البيانات
171	.....	عالم البيانات
175	.....	الخبير الاحصائي

178	.....	معماري البيانات
181	.....	مسؤول البيانات
184	.....	محلل الأعمال
186	.....	مدير البيانات
190	.....	لغات برمجة البيانات
191	.....	<i>Python</i>
193	.....	<i>R</i>
194	.....	<i>SQL</i>
197	.....	<i>MatLab</i>
198	.....	<i>SAS</i>
199	.....	<i>JAVA</i>
201	.....	اسئلة الفصل الخامس
202	.....	مشاريع بحوث و دراسات علمية مقترحة :
204	.....	المراجع

## المقدمة

في عالم اليوم، تؤدي البيانات دورًا مهمًا في تشكيل الطريقة التي نعيش بها ونفهم العالم من حولنا. إنها القوة الدافعة وراء العديد من التطورات التكنولوجية و الصناعية التي اسهمت في تحسين حياة المجتمعات، من الرعاية الصحية إلى التمويل إلى النقل. وفي الاقتصاد الحديث، أصبحت البيانات سلعة ثمينة، وهناك طلب مرتفع على القدرة على تحليلها والاستفادة منها بشكل فعال. و مع استمرار استخدام البيانات في مختلف القطاعات، تزداد أيضًا الحاجة إلى المهنيين المهرة القادرين على فهم الكميات الهائلة من البيانات التي يتم إنشاؤها كل يوم. خاصة في مجال تكنولوجيا المعلومات، حيث تكون البيانات في مقدمة التحديات التي تواجهها المنظمات.

لهذا السبب، من المهم أكثر من أي وقت مضى أن نزود طلاب جامعتنا بالأدوات والمعرفة التي يحتاجون إليها لفهم البيانات والعمل معها. ينطبق هذا بشكل خاص على الطلاب الذين يدرسون في قسم المعلومات، حيث يعد الفهم العميق للبيانات أمرًا ضروريًا للنجاح في المجال الذي يختارونه. نهدف من خلال هذا الكتاب إلى تقديم مقدمة في مجال علم البيانات وأهمية البيانات في حياتنا.

سنغطي المفاهيم والتقنيات الأساسية المستخدمة لتحليل البيانات واستخدامها، وسنستكشف الطرق التي تعمل بها البيانات على تحويل انماط الحياة. سنقوم أيضًا بفحص التحديات والفرص التي تنشأ عند التعامل مع البيانات، والاعتبارات الأخلاقية التي يجب أخذها في الاعتبار.

هدفنا هو تزويد طلاب قسم المعلومات بأساس متين في علم البيانات وتقدير لقوة البيانات في وظائفهم المستقبلية. يهدف هذا الكتاب إلى أن يكون مصدرًا قيمًا للطلاب، وكذلك لأي شخص مهتم بفهم دور البيانات في العالم الحديث.

يقدم الفصل الأول، "المفاهيم الأساسية لعلم البيانات"، لمحة عامة عن ماهية علم البيانات. ويغطي المفاهيم الأساسية مثل جمع البيانات وتنظيف البيانات واستكشاف البيانات و تصور البيانات.

ناقشنا في الفصل الثاني تحليل البيانات واعدادها الذي يعد فصلًا مهمًا للطلاب لأنه يضع الأساس لبقية موضوعات علم البيانات. بدون تحليل البيانات واعدادها بشكل صحيح، ستكون العمليات التي يتم إجراؤها لاحقًا ضعيفة وغير موثوقة.

الفصل الثالث يهدف إلى تقديم نظرة عامة شاملة عن مفهوم البيانات الضخمة وتطبيقاتها. الفصل الرابع يركز على أفضل الممارسات لإدارة البيانات وكيف يمكن للمؤسسات الاستفادة من إدارة البيانات الفعالة.

الفصل الخامس سنقوم بفحص أدوار محلل البيانات، ومهندس البيانات، ومهندس التعلم الآلي، ومحلل ذكاء الأعمال، ومهندس تصور البيانات، وعالم البيانات.

## الفصل الأول

### المفاهيم الأساسية لعلم البيانات.

#### تمهيد

مجال علم البيانات نما بسرعة كبيرة في السنوات الأخيرة، وأصبح تأثيره محسوسًا عبر مجموعة واسعة من المجالات الصناعية والتكنولوجية. وتجدر الإشارة إلى أن التوجهات الحديثة لدراسة علم البيانات غالبًا ما تهدف إلى فتح الرؤى من أجل اتخاذ قرارات أفضل. تم تصميم هذا الكتاب ليكون بمثابة مقدمة لمفاهيم وتعريفات علم البيانات للدارسين في هذا المجال. إنه مخصص للطلاب و المتخصصين في مجال علم المعلومات، وأي شخص آخر مهتم بفهم أساسيات علم البيانات وتطبيقاتها.

يقدم الفصل الأول، "المفاهيم الأساسية لعلم البيانات"، لمحة عامة عن ماهية علم البيانات. ويغطي المفاهيم الأساسية مثل جمع البيانات وتنظيف البيانات واستكشاف البيانات و تصور البيانات. وغيرها من المفاهيم ذات العلاقة المباشرة في تخصص علم المعلومات.

سنناقش أيضًا الأدوار والمسؤوليات المختلفة لعالم البيانات والأدوات والتقنيات المختلفة المستخدمة في إدارة البيانات. بنهاية هذا الفصل، سيكون لديك فهم قوي لماهية علم البيانات ومكوناته المختلفة وكيفية استخدامه لحل المشكلات في العالم الحقيقي. ومن

المتوقع يعطيك هذا الفصل استعداد جديد للتعلم في فصول الكتاب الأخرى.

## مفهوم البيانات

تقديم تعريف جامع مانع لمفهوم البيانات، لا يمكن ان يتم بعيدا عن الإحاطة بمفاهيم المعلومات والمعرفة، كون البيانات تشكل الركيزة التي تقوم عليها المفاهيم الموضوعية الأخرى. وهنا تجدر الإشارة الى وجود اتفاق شبه كامل بين المتخصصين في مجال المعلومات و نظرائهم في مجال الحاسوب، على ان البيانات تشير إلى مجموعة من القيم أو الملاحظات أو الحقائق التي يتم تمثيلها بتنسيق معين، مثل الجدول أو جدول البيانات أو قاعدة بيانات. ويمكن أن تأتي في أشكال عديدة، مثل البيانات العددية والفئوية والنمطية والنصية. كما يمكن أن تكون منظمة أو غير منظمة، اعتمادًا على كيفية تمثيلها وتخزينها. غالبًا ما تُستخدم البيانات لتمثيل ظاهرة في العالم الحقيقي، ويمكن جمعها وتحليلها لدعم اتخاذ القرار أو اختبار الفرضيات أو تحديد الأنماط والاتجاهات.

اما المعلومات فيُنظر لها على أنها مورد يتم إنشاؤه وجمعه وتنظيمه وإتاحته لدعم البحث العلمي والتعليم والأنشطة الأخرى. يمكن أن تتخذ المعلومات أشكال عدة، مثل الكتب والمجلات والمقالات والموارد الرقمية وغيرها من المواد. تؤدي مؤسسات المعلومات دورًا حاسمًا في توفير الوصول إلى هذه المعلومات، وفي كثير من الحالات، يكون متخصصو المعلومات مسؤولين عن إدارة هذه الموارد وتنظيمها.

قد يشاركون أيضًا في إنشاء وتنظيم المعلومات والبيانات الوصفية الرقمية، فضلًا عن الحفاظ على المعلومات بتدسيقات مختلفة.

في سياق متصل يرتبط مفهوم المعرفة ارتباطًا وثيقًا بفكرة المعلومات. اذ غالبًا ما يُنظر إلى المعرفة على أنها فهم عالي المستوى أو إتقان لموضوع أو منطق معين، مبني على أساس المعلومات. يُنظر إلى المعرفة أيضًا، على أنها أحد الأصول الحاسمة التي يجب إنشاؤها والتقاطها ومشاركتها وتطبيقها لدعم الأهداف الشخصية والتنظيمية. ترتبط المعرفة بعملية الاستفسار والتعلم، بناءً على جودة ومصداقية المعلومات والموارد المستخدمة كأساس. متخصصو المعلومات غالبًا هم المسؤولين عن تنسيق المعرفة والحفاظ عليها، لا سيما في التنسيق الرقمي. كما انهم يؤدون دورًا مهمًا في إنشاء وصيانة البيانات الوصفية والتصنيفات وأدوات تنظيم المعلومات الأخرى لمساعدة المستخدمين في العثور والوصول الى المعلومات التي يحتاجون إليها.

### **أهمية وفوائد البيانات**

البيانات بلاشك شك لها اهمية كبيرة لأنها تمكن الأفراد والمؤسسات من اتخاذ قرارات مستنيرة وتنفيذ إجراءات فعالة. و يمكن استخدامها لتحديد الأنماط والاتجاهات واكتساب الأفكار واختبار الفرضيات. يمكن أيضًا استخدام البيانات لتقييم أداء العمليات والمنتجات والخدمات، وقياس نجاح الاستراتيجيات والمبادرات. فضلًا عن ذلك، يمكن استخدام البيانات لتحسين دقة وكفاءة المهام المختلفة وأتمتة عملية صنع القرار.

بشكل عام، تؤدي البيانات دورًا مهمًا في العديد من جوانب الحياة، وهي ضرورية لاتخاذ قرارات مستنيرة وفعالة في مختلف الأعمال و البحث العلمي والمجالات المتنوعة الأخرى .

وتجدر الإشارة الى ان هناك العديد من الفوائد للبيانات على مستوى الافراد والمؤسسات، والتي يمكن ايجازها بالآتي:

1. صنع القرار: يمكن استخدام البيانات لإعلام وتحسين عملية صنع القرار من خلال تزويد الأفراد والمنظمات بفهم واضح للحقائق والاتجاهات المتعلقة بعملياتهم وصناعاتهم .
2. الكفاءة: يمكن استخدام البيانات لتحديد أوجه القصور ومجالات التحسين في العمليات والمنتجات والخدمات، وأتمتة المهام واتخاذ القرار .
3. الابتكار: يمكن استخدام البيانات لتحديد فرص جديدة للنمو والابتكار من خلال الكشف عن الأنماط والاتجاهات التي قد لا تكون واضحة على الفور .
4. قياس الأداء: يمكن استخدام البيانات لقياس وتتبع أداء العمليات والمنتجات والخدمات وتقييم نجاح الاستراتيجيات والمبادرات.
5. التخصيص: يمكن استخدام البيانات لإنشاء تجارب مخصصة للعملاء من خلال فهم تفضيلاتهم وسلوكهم .
6. التحليل التنبؤي: يمكن استخدام البيانات للتنبؤ بالاتجاهات والأحداث المستقبلية من خلال تحليل البيانات السابقة،

والتي يمكن أن تساعد الشركات والمؤسسات والأفراد على التخطيط والإعداد وفقاً لذلك.

7. خفض التكلفة: يمكن استخدام البيانات لتحديد المجالات التي يمكن خفض التكلفة فيها، من خلال تحديد أوجه القصور والمجالات الفوضوية.

8. الامتثال: يمكن استخدام البيانات لضمان الامتثال للقوانين واللوائح من خلال توفير سجل واضح للأنشطة والمعاملات.

بشكل عام، توفر البيانات للمؤسسات والأفراد رؤى قيمة يمكن استخدامها لدفع نمو الأعمال وتحسين العمليات وزيادة الكفاءة.

### مصادر البيانات

يمكن أن تأتي البيانات من مجموعة متنوعة ومتباينة من المصادر [Data Sources] ، بما في ذلك:

1. المعاملات: يمكن إنشاء البيانات من خلال المعاملات الإدارية والمالية، مثل شراء بطاقات الائتمان أو المعاملات المصرفية.
2. وسائل التواصل الاجتماعي: يمكن جمع البيانات من منصات وسائل التواصل الاجتماعي، مثل Facebook و Twitter ، لاكتساب نظرة ثاقبة حول سلوك العملاء ومشاعرهم.
3. مواقع الويب والأنظمة الأساسية عبر الإنترنت: يمكن جمع البيانات من مواقع الويب والأنظمة الأساسية عبر الإنترنت، مثل Google Analytics ، لتتبع حركة مرور موقع الويب وسلوك المستخدم.

4. أجهزة الاستشعار وأجهزة إنترنت الأشياء: يمكن جمع البيانات من أجهزة الاستشعار وأجهزة إنترنت الأشياء (IoT) ، مثل الأجهزة القابلة للارتداء والأجهزة المنزلية الذكية، لتتبع المعلومات الشخصية والظروف البيئية .

5. مجموعات البيانات الحكومية والعامّة: يمكن جمع البيانات من مختلف الوكالات الحكومية التي تجمع البيانات من المؤسسات العامة والخاصة، وكذلك البيانات من المنظمات غير الربحية والمؤسسات البحثية هي أيضًا مصادر قيمة للبيانات .

6. البيانات الداخلية: يمكن أيضًا إنشاء البيانات داخليًا بواسطة مؤسسة من خلال عملياتها الروتينية، مثل بيانات المبيعات وبيانات المخزون ومعلومات الموظف.

7. موفرو البيانات من الطرف الثالث: يمكن شراء البيانات من مزودي البيانات الخارجيين مثل شركات الأبحاث أو وسطاء البيانات أو مزودي المعلومات المالية أو الديموغرافية.

من المهم ملاحظة أن المصادر المختلفة للبيانات قد يكون لها مستويات مختلفة من الدقة والاكتمال والموثوقية، ومن المهم تقييم جودة البيانات قبل استخدامها في اتخاذ القرار أو التحليل.

## أنواع البيانات

هناك عدة أنواع مختلفة من البيانات، بما في ذلك :

1. البيانات المهيكلة: البيانات المنظمة بتنسيق معين، مثل الصفوف والأعمدة في جدول بيانات أو الحقول في قاعدة بيانات. تتضمن الأمثلة البيانات الرقمية والتواريخ والبيانات الفئوية .
2. البيانات غير المهيكلة: البيانات التي ليس لها تنسيق معين، مثل النص من منشورات وسائل التواصل الاجتماعي أو المقالات أو مراجعات العملاء. غالبًا ما يكون تحليله أصعب لأنه لا يحتوي على هيكل واضح .
3. البيانات شبه المنظمة: البيانات التي لها مستوى معين من التنظيم، ولكن ليس بقدر البيانات المنظمة. تتضمن الأمثلة ملفات JSON وXML ، التي تحتوي على عناصر وسمات متداخلة، لكن لا يوجد مخطط ثابت لها .
4. البيانات العددية: البيانات التي يمكن قياسها كميًا واستخدامها في العمليات الحسابية. يتضمن ذلك بيانات مثل العمر والوزن ودرجة الحرارة وما إلى ذلك .
5. البيانات الفئوية: البيانات التي يمكن تقسيمها إلى فئات . يتضمن ذلك بيانات مثل الجنس ونوع المنتج وما إلى ذلك .
6. البيانات الترتيبية: البيانات التي يمكن طلبها أو ترتيبها. إنها تشبه البيانات الفئوية، لكن لها ترتيب طبيعي. على سبيل المثال، تصنيف مسح مثل: ضعيف، عادل، جيد، ممتاز
7. البيانات الثنائية: البيانات التي يمكن أن تحتوي على قيمتين فقط. على سبيل المثال صح أو خطأ، 1 أو 0 .

8. بيانات السلاسل الزمنية: البيانات التي يتم جمعها خلال فترة زمنية، مثل أسعار الأسهم ودرجة الحرارة وما إلى ذلك. من المهم تحديد نوع البيانات لأنه يجب اختيار طريقة التحليل وأدوات الإحصاء وفقًا لذلك.

### تصنيف البيانات

تصنيف البيانات [Data Classification] هو عملية تنظيم البيانات في فئات أو اصناف بناءً على خصائص أو سمات معينة. هناك عدة طرق مختلفة لتصنيف البيانات، بما في ذلك:

1. تصنيف البيانات الرقمية: يعتمد هذا النوع من التصنيف على القيم العددية، مثل العمر والوزن ودرجة الحرارة. يمكن تجميع البيانات في فترات أو نطاقات، مثل الفئات العمرية (0-18، 18-35، إلخ) أو نطاقات درجة الحرارة (0-10 درجات، 10-20 درجة، إلخ).
2. تصنيف البيانات الفئوية: يعتمد هذا النوع من التصنيف على القيم الفئوية، مثل الجنس أو نوع المنتج أو الموقع الجغرافي. يمكن تجميع البيانات في فئات، مثل الرجال والنساء والأطفال أو أمريكا الشمالية وأوروبا وآسيا وما إلى ذلك.
3. تصنيف البيانات الهرمي: يعتمد هذا النوع من التصنيف على هيكل هرمي، حيث يتم تنظيم البيانات في هيكل يشبه الشجرة، مع مستويات متعددة من الفئات. على سبيل

المثال، قد يحتوي كتالوج منتجات الشركة على مستوى أعلى من الفئة مثل الإلكترونيات، ثم الفئات الفرعية مثل الهواتف وأجهزة الحاسوب المحمولة والأجهزة اللوحية وما إلى ذلك .

4. تصنيف البيانات النصية: يعتمد هذا النوع من التصنيف على البيانات النصية، مثل منشورات الوسائط الاجتماعية أو مراجعات العملاء. يمكن تحليل النص لتحديد بعض الكلمات الرئيسية أو الموضوعات، ثم تجميعها في فئات أو فئات، مثل المشاعر الإيجابية أو السلبية أو الحيادية .

5. تصنيف البيانات الصورية: يعتمد هذا النوع من التصنيف على بيانات الصورة، مثل الصور الفوتوغرافية أو مقاطع الفيديو. يمكن تحليل الصور لتحديد سمات أو أشياء معينة، ثم تجميعها في فئات أو فئات، مثل الحيوان أو السيارة أو المبنى.

تعتمد طريقة تصنيف البيانات على نوع البيانات والغرض من التحليل وسؤال البحث الذي تحتاج إلى إجابته. يمكن أن يساعد التصنيف في تحديد الأنماط والاتجاهات، لجعل البيانات أكثر قابلية للإدارة ومفيدة للتحليل واتخاذ القرار.

## جمع البيانات

هناك عدة طرق مختلفة لجمع البيانات [Data Collection] ، وستعتمد الطريقة التي تختارها على نوع البيانات التي تحتاجها

والموارد المتاحة لديك. تتضمن بعض الطرق الشائعة لجمع البيانات ما يلي:

1. الاستطلاعات: تعد الاستطلاعات طريقة شائعة لجمع البيانات من عدد كبير من الأشخاص. يمكن إجراء الاستطلاعات شخصيًا أو عبر الهاتف أو البريد أو عبر الإنترنت. يمكن أن تكون منظمة (بأسئلة محددة مسبقًا) أو غير منظمة (بأسئلة مفتوحة). (الاستطلاعات طريقة مفيدة لجمع البيانات حول الآراء والمواقف والسلوكيات).
2. التجارب: التجارب هي طريقة لجمع البيانات في بيئة مضبوطة. يسمح هذا للباحثين بعزل تأثيرات المتغيرات المختلفة على نتيجة الاهتمام. يمكن إجراء التجارب في المختبر أو في الميدان. يمكن أيضًا إجراؤها مع أشخاص أو حيوانات.
3. الملاحظات: الملاحظات هي طريقة لجمع البيانات من خلال مراقبة وتسجيل سلوك الأشخاص أو الحيوانات أو الأنظمة. يمكن إجراء الملاحظات بطريقة منظمة أو غير منظمة. يمكن أن تكون نوعية (بناءً على أوصاف ذاتية) أو كمية (بناءً على قياسات موضوعية).
4. المقابلات: تتضمن المقابلات طرح أسئلة على الأفراد في مقابلة وجهًا لوجه أو عبر الهاتف أو عبر الإنترنت. يمكن تنظيم المقابلات (بأسئلة محددة مسبقًا) أو غير منظمة (بأسئلة مفتوحة). (إنها طريقة مفيدة لجمع معلومات مفصلة حول تجارب الشخص ومعتقداته ومواقفه).

5. البيانات العامة: في بعض الأحيان تكون البيانات التي تحتاجها متاحة بالفعل من مصادر أخرى، مثل الوكالات الحكومية أو قواعد البيانات التجارية أو الباحثين الآخرين. يُعرف هذا بالبيانات الثانوية ويمكن استخدامه في البحث دون أي جمع أو تكلفة إضافية.

6. وسائل التواصل الاجتماعي: يمكن جمع بيانات الوسائط الاجتماعية من خلال تعقب الويب أو واجهة برمجة التطبيقات أو أي طريقة أخرى لاستخراج البيانات ذات الصلة من منصات مختلفة مثل Facebook و Twitter و Instagram أو إلى ذلك.

بمجرد جمع بياناتك، من المهم التأكد من أنها دقيقة وذات جودة عالية قبل أن تبدأ التحليل. قد يتضمن ذلك التحقق من الأخطاء أو التناقضات، وتنظيف البيانات وتنسيقها حسب الضرورة.

## فهم البيانات

فهم البيانات [Data Understanding] يراد به التعرف على البيانات والمشكلة التي يهدف مشروع علم البيانات إلى حلها. الهدف من هذه المرحلة هو تحديد المشكلة وفهم البيانات ووضع مجموعة واضحة من الأهداف والمتطلبات للمشروع. خلال هذه المرحلة، سيقوم عالم البيانات بما يلي:

1. تحديد المشكلة، وفهم سياق البيانات، وتحديد أهداف العمل للمشروع.

2. جمع البيانات وجمعها من مصادر مختلفة، مثل قواعد البيانات أو الملفات الثابتة أو تنقيب الويب.
3. فهم هيكل البيانات وتنسيقها، بما في ذلك عدد الملاحظات والميزات وأنواع البيانات.
4. حدد أي بيانات مفقودة أو تناقضات في البيانات يجب معالجتها في مرحلة الإعداد.
5. فهم مجال المشكلة واستكشف أي معلومات أساسية ذات صلة.
6. فهم المستخدمين النهائيين وأصحاب المصلحة، وكيف سيستخدمون النتائج.

هذه المرحلة حاسمة في إنشاء أساس قوي لبقية مشروع علم البيانات. من الضروري فهم البيانات والمشكلة والأهداف بشكل كامل من أجل اتخاذ قرارات مستنيرة طوال بقية دورة الحياة.

### إعداد البيانات

يقصد [Data preparation] عملية تحضير البيانات وتحويلها وتنسيقها بحيث يمكن استخدامها لمزيد من التحليل والنمذجة. هذه المرحلة حاسمة في ضمان سلامة وجودة البيانات، فضلاً عن جعلها مناسبة للمشكلة المحددة والهدف المطروح. خلال هذه المرحلة، سيقوم عالم البيانات بما يلي:

1. تنظيف البيانات، والتي تتضمن إزالة عدم الدقة والتناقضات والقيم المفقودة.

2. تحويل البيانات، والتي تتضمن تحويل البيانات من تنسيق أو بنية إلى أخرى وتطبيع البيانات.
3. هندسة الميزات، والتي تتضمن إنشاء ميزات جديدة أو دمج الميزات الموجودة أو تحديد مجموعة فرعية من الميزات.
4. معالجة البيانات الفئوية، والتي تتضمن ترميز المتغيرات الفئوية، وترميز المتغيرات الترتيبية.
5. قسّم البيانات إلى مجموعة تدريب ومجموعة اختبار، وهي ضرورية لتقييم أداء النموذج .

يعد إعداد البيانات عملية متكررة وتستغرق وقتاً طويلاً. خلال هذه المرحلة، سيقضي عالم البيانات الكثير من الوقت في فهم البيانات ومعالجتها والتحقق منها. سيتم استخدام مخرجات هذه المرحلة في المرحلة التالية مثل استكشاف البيانات والنمذجة والتقييم وما إلى ذلك. لذلك، فإن جودة وسلامة البيانات مهمة من أجل استخلاص النتائج الصحيحة منها.

### استكشاف البيانات

يعد استكشاف البيانات [Data Exploration] هي المرحلة التي يتم فيها استخدام تصورات وإحصاءات موجزة لفهم البنية الأساسية وخصائص البيانات. الهدف من استكشاف البيانات هو التعرف على شكل البيانات وتحديد أي أنماط أو اتجاهات وتحديد أي مشكلات محتملة في البيانات. هناك العديد من الأساليب المختلفة التي يمكن استخدامها لاستكشاف البيانات، بما في ذلك :

1. التمثيل المرئي: تعد المرئيات أداة قوية لاستكشاف البيانات، لأنها تسمح لك برؤية الأنماط والعلاقات في البيانات بسرعة. تتضمن الأنواع الشائعة من المرئيات الرسوم البيانية ومخططات التبعثر ومخططات الخطوط ومخططات المربعات. يمكن إنشاء هذه باستخدام أدوات برمجية مثل R أو Python أو SAS أو Excel وغيرها.
2. إحصاءات موجزة: إحصائيات الملخص هي أوصاف رقمية للبيانات التي توفر طريقة سريعة وسهلة لفهم البنية الأساسية للبيانات. تتضمن إحصائيات الملخص الشائعة المتوسط والوسيط والانحراف المعياري والمربعات التي يمكن حسابها باستخدام أدوات برمجية أو لغات برمجة مثل R و Python.
3. جداول التكرار: تعرض جداول التكرار عدد المرات التي تظهر فيها كل قيمة فريدة في مجموعة البيانات. تساعد هذه الجداول في فهم توزيع بعض المتغيرات، وما هي القيم الأكثر شيوعًا وما إذا كان هناك أي قيم متطرفة.
4. الجدولة المتقاطعة: تُظهر الجداول المتقاطعة العلاقة بين متغيرين من خلال حساب عدد المرات التي تظهر فيها كل مجموعة من القيم في مجموعة البيانات. تُعرف هذه الطريقة أيضًا باسم جداول الطوارئ. تساعد هذه الجداول في فهم كيفية ارتباط متغيرين.

5. الارتباط والانحدار: يقيس الارتباط طبيعة العلاقة بين متغيرين، يمكن أن يكون موجبًا أو سلبياً. و الانحدار هو طريقة إحصائية تهدف إلى العثور على أفضل خط ملائم، والذي يصف كيف يعتمد أحد المتغيرات على الآخر. من المهم ملاحظة أن استكشاف البيانات هو عملية تكرارية وقد تحتاج إلى العودة وإجراء تغييرات على البيانات بناءً على ما تجده أثناء الاستكشاف .

الهدف من الاستكشاف ليس العثور على إجابة مثالية، ولكن الحصول على فهم واضح لبياناتك وتحديد أي مشكلات تحتاج إلى المعالجة قبل الانتقال إلى الخطوة التالية من التحليل.

## تحويل البيانات

تحويل او نقل البيانات [Data Transformation] يراد منها إنشاء متغيرات جديدة وإعادة ترميز المتغيرات ودمج البيانات من مصادر متعددة. الهدف من تحويل البيانات هو جعلها أكثر ملاءمة للتحليل الذي تريد إجراؤه، بالإضافة إلى فهم أفضل للأنماط والعلاقات الأساسية في البيانات. هناك العديد من الأساليب المختلفة التي يمكن استخدامها لتحويل البيانات، بما في ذلك :

1. إنشاء متغيرات جديدة: غالبًا ما يتم إنشاء متغيرات جديدة لاستخراج معلومات إضافية من المتغيرات الحالية. يمكن أن يشمل ذلك إنشاء متغيرات التفاعل، والتي هي نتاج اثنين أو

- أكثر من المتغيرات الحالية، أو إنشاء متغيرات وهمية، والتي تستخدم لتمثيل المتغيرات الفئوية في شكل رقمي .
2. متغيرات إعادة الترميز: تستخدم متغيرات إعادة الترميز لتغيير قيم أو ترميز المتغيرات الموجودة. يمكن أن يشمل ذلك تحويل المتغيرات المستمرة إلى متغيرات فئوية، أو تغيير قيم المتغيرات الفئوية لتناسب بشكل أفضل مع سؤال البحث.
3. الجمع بين البيانات من مصادر متعددة: يعد دمج البيانات من مصادر متعددة مفيدًا عندما تنتشر البيانات المطلوبة للتحليل عبر عدة قواعد بيانات أو مجموعات بيانات مختلفة. يمكن القيام بذلك عن طريق ضم البيانات الخاصة بالمتغيرات الشائعة، أو بدمج البيانات في مجموعة بيانات واحدة .
4. توحيد البيانات: تتكون هذه العملية من التأكد من أن البيانات على نفس المقياس، ويمكن إجراؤها عن طريق طرح المتوسط والقسمة على الانحراف المعياري. يمكن أن يساعد هذا في التأكد من أن جميع المتغيرات على نفس المقياس وقابلة للمقارنة .
5. تطبيع البيانات: تتكون هذه الطريقة من تحويل البيانات بحيث تتبع التوزيع الطبيعي، ويمكن أن تساعد هذه العملية في أن تكون البيانات قابلة للتفسير وسهلة المقارنة .
6. تنظيم البيانات: تتكون هذه العملية من تلخيص البيانات بطرق مختلفة عن طريق تجميعها حسب متغير معين. يمكن

أن يساعد ذلك في تقليل تعقيد البيانات وحجمها وتسهيل تفسيرها من المهم توثيق جميع خطوات تحويل البيانات، وإذا فقدت أي معلومات مهمة، فمن المهم تدوين ذلك. كأفضل ممارسة، غالبًا ما يكون من المفيد إنشاء مجموعة بيانات جديدة بالبيانات المحولة والاحتفاظ بنسخة منفصلة من البيانات الأصلية .

يمكن أن يكون تحويل البيانات عملية تكرارية وقد يحتاج الباحث إلى العودة وإجراء التغييرات بناءً على نتائج استكشاف البيانات وخطوات النمذجة.

## نماذج البيانات

نماذج البيانات [Data Models] تقوم على استخدام نماذج التعلم الإحصائي أو الآلي لتحليل البيانات وإجراء التنبؤات. الهدف من نمذجة البيانات هو فهم الأنماط والعلاقات الأساسية في البيانات، وكذلك لعمل تنبؤات أو قرارات بناءً على البيانات. هناك عدة أنواع مختلفة من النماذج التي يمكن استخدامها لنمذجة البيانات، بما في ذلك :

1. النماذج الإحصائية: النماذج الإحصائية هي نماذج رياضية تصف العلاقة بين المتغيرات. تشمل الأنواع الشائعة للنماذج الإحصائية الانحدار الخطي، والانحدار اللوجستي، و ANOVA، من بين أمور أخرى. تُستخدم هذه النماذج لعمل تنبؤات أو فهم العلاقة بين المتغيرات .

2. نماذج التعلم الآلي: نماذج التعلم الآلي هي نوع من النماذج التي يمكنها التعلم من البيانات والتنبؤ. تشمل الأنواع الشائعة لنماذج التعلم الآلي أشجار القرار، والغابات العشوائية، وآلات ناقلات الدعم، والشبكات العصبية، من بين أشياء أخرى. تُستخدم هذه النماذج لعمل تنبؤات وتصنيف البيانات وتحديد الأنماط في البيانات.
3. نماذج السلاسل الزمنية: نماذج السلاسل الزمنية هي نماذج يمكنها تحليل البيانات التي تحتوي على مكون زمني. هذه النماذج مفيدة لفهم الأنماط والتنبؤ بها في بيانات السلاسل الزمنية، مثل أسعار الأسهم أو الطقس أو بيانات المبيعات. تشمل الأمثلة ARIMA و Holt-winters.
4. تحليل الكتلة: تحليل الكتلة هو طريقة تقوم بتجميع البيانات في مجموعات فرعية متميزة، تسمى المجموعات أو الكتل، والتي تشبه بعضها البعض. الفكرة من وراء ذلك هي تقسيم البيانات إلى مجموعات ذات مغزى لها خصائص مشتركة.
5. تقليل الأبعاد: تقليل الأبعاد هو عملية تهدف إلى تقليل عدد الميزات في مجموعة البيانات، إما عن طريق اختيار الميزات الأكثر صلة أو الجمع بين الميزات المرتبطة. يمكن أن يساعد ذلك في تحسين أداء بعض النماذج، وجعل البيانات أكثر قابلية للتفسير. عند اختيار نموذج، من المهم مراعاة نوع البيانات وسؤال البحث والموارد المتاحة.

يمكن أن تكون النمذجة عملية تكرارية، وقد يحتاج الباحث إلى تجربة عدة نماذج مختلفة قبل العثور على النموذج الأنسب للبيانات. فضلاً عن ذلك، من المهم تقييم أداء النموذج وتفسيره ومقارنته بالنماذج الأخرى.

### نشر النتائج

يعد نشر النتائج الخطوة الأخيرة في دورة حياة البيانات. و تتضمن تقديم نتائج التحليل للآخرين بطريقة واضحة وفعالة. الهدف من توصيل النتائج هو جعل نتائج التحليل في متناول الجمهور المستهدف ومفهومة. هناك عدة طرق لتوصيل النتائج، بما في ذلك :

1. التقارير المكتوبة: التقارير المكتوبة هي طريقة شائعة لإيصال النتائج. يمكن أن تكون في شكل ورقة بحث أو تقرير فني أو تقرير أعمال، من بين أمور أخرى. تتضمن التقارير المكتوبة عادةً مقدمة وقسمًا للمنهجية وقسمًا للنتائج وخاتمة .
2. العروض التقديمية الشفوية: العروض التقديمية الشفوية هي طريقة شائعة لإيصال النتائج، خاصة في الأوساط الأكاديمية أو التجارية. تتضمن العروض التقديمية الشفوية عادةً عرض شرائح مع مساعدات بصرية وشرح منطوق للنتائج.
3. لوحات المعلومات التفاعلية: تُعد لوحات المعلومات التفاعلية طريقة لإيصال النتائج من خلال واجهة قائمة على الويب أو تفاعلية. يمكن استخدام هذه لتقديم البيانات

بطريقة تفاعلية وجذابة والسماح للمستخدمين باستكشاف البيانات بمزيد من العمق .

4. العرض المرئي للبيانات: يعد طريقة رائعة أخرى لتوصيل نتائج البيانات، ويمكن أن يساعد في تقديم البيانات بطريقة بديهية، مما يسمح للجمهور بفهم أفضل للنتائج، خاصةً عندما تكون هناك أنماط وعلاقات يصعب وصفها من خلال النص .

عند نشر النتائج، من المهم مراعاة الجمهور والغرض من التحليل والنتائج الرئيسية. من المهم أيضًا أن نكون صادقين بشأن أي قيود أو شكوك في التحليل، وأن نوفر سياقًا للنتائج. يجب أن يكون ناشر النتائج قادرًا على نقل القصة والأفكار الكامنة وراء البيانات، ويجب أن تكون هذه الأفكار واضحة وسهلة الفهم، ويجب تصميمها بطريقة تجعل من السهل التعامل مع البيانات والأفكار. فضلًا عن ذلك، من المهم الانتباه إلى تصميم التصورات واستخدام الألوان والتسميات والمقاييس، كل هذه العناصر يمكن أن يكون لها تأثير كبير على كيفية إدراك النتائج وتفسيرها من قبل الجمهور.

## علم البيانات

علم البيانات [Data Science] هو مجال متعدد التخصصات يتضمن استخدام الأساليب والعمليات والخوارزميات والأنظمة العلمية لاستخراج المعرفة والأفكار من البيانات المنظمة وغير المنظمة. يشمل على مجموعة واسعة من الأنشطة، بما في ذلك تنظيف

البيانات وإعدادها، وتصور البيانات، وتحليل البيانات والنمذجة، والتعلم الآلي، الخ.... الهدف من علم البيانات هو تحويل البيانات الأولية إلى رؤى وقرارات قابلة للتنفيذ. للقيام بذلك، يستخدم علماء البيانات مجموعة متنوعة من التقنيات، مثل: التحليل الإحصائي، والتعلم الآلي، ومعالجة اللغة الطبيعية، لاستخراج معلومات ذات مغزى من البيانات.

يعد تنظيف البيانات وإعدادها أحد المكونات الرئيسية لعلم البيانات، وهي عملية تحويل البيانات إلى تنسيق يمكن تحليله ونمذجته بسهولة. غالبًا ما يتضمن ذلك دمج البيانات من مصادر متعددة، والتعامل مع البيانات المفقودة أو غير الدقيقة، وتحويل البيانات إلى تنسيق يمكن تحليله بسهولة.

ويعد تصور البيانات أيضًا جانبًا مهمًا في علم البيانات، حيث يتم استخدامه لإنشاء المخططات والرسوم البيانية والتمثيلات المرئية الأخرى للبيانات التي تسهل فهم الرؤى والنتائج وتوصيلها.

أما تحليل البيانات والنمذجة فهما جوهر علم البيانات، وهذا يتضمن استخدام تقنيات التعلم الإحصائي والآلي المختلفة للكشف عن الأنماط والعلاقات في البيانات، ثم بناء النماذج لعمل تنبؤات أو تحديد الدوافع الرئيسية للنتيجة.

يتم استخدام علم البيانات في العديد من المجالات مثل: الرعاية الصحية، والتمويل وتجارة التجزئة، والنقل، والعديد من

المجالات الأخرى للكشف عن رؤى جديدة وتحسين عملية صنع القرار وتحسين العمليات التجارية.

بشكل عام، يعد علم البيانات مجالاً سريع النمو وله مجموعة واسعة من التطبيقات في مختلف المجالات ويزداد الطلب عليه بسبب زيادة كمية البيانات التي يتم إنشاؤها كل يوم.

### تاريخ علم البيانات

مجال علم البيانات له جذوره في عدد من التخصصات المختلفة، بما في ذلك الإحصاء والرياضيات وعلوم الحاسوب وعلوم المعلومات. يمكن إرجاع تاريخ علم البيانات إلى أوائل القرن العشرين، مع تطوير الأساليب والأدوات الإحصائية لتحليل البيانات.

كان الإحصائي رونالد أ. فيشر أحد رواد علم البيانات الأوائل، حيث طور العديد من التقنيات الإحصائية التي لا تزال مستخدمة في علم البيانات حتى اليوم. في سياق متصل طور بعض الإحصائيين مثل جورج إي بوكس، طريقة لتحديد القيم المتطرفة في البيانات، كما اسهم (دبليو إدواردز دمينغ)، المعروف بعمله في مراقبة الجودة ومراقبة العمليات الإحصائية من خلال نظريته المسماة بدائرة ديمينغ التي بناها على أربعة محاور (خطط - نفذ - افحص - باشر).

في أواخر الخمسينيات والستينيات من القرن الماضي، بدأ مجال علوم الحاسوب في التطور، وبدأت الأساليب المعتمدة على الحاسوب لتحليل البيانات تكتسب شعبية على المستوى العالمي. أتاح ظهور أجهزة الحاسوب المركزية وأنظمة تخزين البيانات واسعة

النطاق تحليل مجموعات بيانات أكبر بكثير مما كان ممكنًا في السابق . أدى ذلك إلى تطوير تقنيات جديدة لتحليل البيانات، مثل أشجار القرار وتحليل المجموعات، والتي تم تصميمها خصيصًا للاستخدام مع أجهزة الحاسوب.

استمر مجال علم البيانات في التطور في السبعينيات والثمانينيات، مع تطوير طرق تحليل البيانات الجديدة، مثل الشبكات العصبية الاصطناعية والخوارزميات الجينية. هذه التقنيات الجديدة، جنبًا إلى جنب مع التطورات في أجهزة وبرامج الحاسوب، جعلت من الممكن تحليل مجموعات بيانات أكبر وأكثر تعقيدًا .

بدأ استخدام مصطلح "علم البيانات" نفسه على نطاق واسع في التسعينيات والعقد الأول من القرن الحادي والعشرين. واستمر المجال في التطور مع توفر أجهزة حاسوب أكثر قوة وأدوات تحليل بيانات أكثر تقدمًا . ومع صعود الإنترنت والكمية المتزايدة من البيانات التي يتم إنشاؤها عبر الإنترنت، أصبح علم البيانات ذا أهمية متزايدة للشركات التي تتطلع إلى اكتساب رؤى والتنبؤات من هذه البيانات.

في السنوات الأخيرة، استمر مجال علم البيانات في التطور والتوسع، مدفوعًا في جزء كبير منه بنمو البيانات الضخمة وزيادة توافر أدوات وأطر تحليل البيانات القوية مثل Hadoop و Spark و Tensorflow وغيرها من مكتبات التعلم الآلي الشهيرة. أصبح علم البيانات أداة أساسية للمنظمات التي تتطلع إلى اكتساب ميزة تنافسية في الاقتصاد القائم على البيانات، من المهم أيضًا ملاحظة

أن مجال علم البيانات هو مجال متعدد التخصصات، يتطور باستمرار مع دمج التقنيات والمنهجيات والأدوات الجديدة، مما يجعله مجالاً صعباً ولكنه مثير للعمل عليه.

### أهمية علم البيانات

أصبح علم البيانات مجالاً مهماً بشكل متزايد في السنوات الأخيرة، بسبب انفجار البيانات والطلب المتزايد على الرؤى والتنبؤات من تلك البيانات. فيما يلي بعض الأسباب التي تجعل علم البيانات مهماً للغاية:

1. تحسين عملية صنع القرار: يمكن لعلم البيانات أن يساعد المؤسسات على اتخاذ قرارات أفضل من خلال توفير الرؤى والتنبؤات التي يمكن أن توجه الاستراتيجيات والعمليات. على سبيل المثال، قد تستخدم شركة بيع بالتجزئة علم البيانات للتنبؤ بطلب العملاء لمنتجات مختلفة، وتعديل مخزونها وفقاً لذلك.
2. الأتمتة والكفاءة: يمكن أن يساعد علم البيانات في أتمتة العمليات وجعلها أكثر كفاءة. على سبيل المثال، يمكن للصيانة التنبؤية في التصنيع استخدام البيانات من المستشعر للتنبؤ بالوقت المحتمل لفشل الجهاز، بحيث يمكن جدولة الصيانة قبل حدوث العطل.
3. توفير التكاليف: باستخدام البيانات لإبلاغ القرارات، يمكن للمؤسسات توفير المال. يمكن أن تساعد النمذجة التنبؤية

الشركات على تحديد العملاء الذين من المرجح أن يتضاربوا بحيث يمكن تركيز الموارد على الاحتفاظ بهؤلاء العملاء، ويمكن للشركات التوفير في تكاليف الاستحواذ للعملاء الجدد.

4. الابتكار: يمكن لعلم البيانات أيضًا أن يقود الابتكار من خلال تحديد الفرص والرؤى الجديدة التي لم يكن من الممكن اكتشافها من خلال الأساليب التقليدية. على سبيل المثال، يمكن استخدام علم البيانات لتحليل البيانات من وسائل التواصل الاجتماعي ومراجعات العملاء والمصادر الأخرى لتحديد الاتجاهات والفرص الجديدة.

5. منتجات وخدمات أفضل: باستخدام علم البيانات لاكتساب فهم أعمق لاحتياجات العملاء وتفضيلاتهم، يمكن للمؤسسات إنشاء منتجات وخدمات أكثر فاعلية. على سبيل المثال، قد تستخدم الشركة علم البيانات لتحليل ملاحظات العملاء وتحديد المجالات التي يمكن تحسين منتجاتها فيها.

6. تحسين العمليات: يمكن استخدام علم البيانات لتحسين العمليات المختلفة وتعزيزها، مثل التصنيع واللوجستيات وسلسلة التوريد وخدمة العملاء والدعم وغير ذلك الكثير. يساعد على أتمتة العملية اليدوية وتحسين الكفاءة العامة للمؤسسة. بشكل عام، أصبح علم البيانات أداة مهمة للمؤسسات التي تتطلع إلى اكتساب ميزة في الاقتصاد القائم على البيانات اليوم. مع توفر المزيد من البيانات، سيستمر علم

البيانات في لعب دور حاسم في مساعدة المؤسسات على فهم تلك البيانات واستخدامها لتحسين عملياتها والابتكار واتخاذ قرارات أفضل.

### مجالات استخدام علم البيانات

علم البيانات هو مجال متعدد التخصصات يستخدم الأساليب العلمية والعمليات والخوارزميات والأنظمة لاستخراج المعرفة والرؤى من البيانات المنظمة وغير المنظمة. يمكن استخدامه لمجموعة واسعة من التطبيقات، بما في ذلك:

1. ذكاء الأعمال: يمكن استخدام علم البيانات لتحليل بيانات العملاء وبيانات المبيعات وغيرها من البيانات المتعلقة بالأعمال لاكتساب رؤى حول سلوك العملاء والتنبؤ بالاتجاهات المستقبلية وإبلاغ قرارات العمل.
2. كشف الاحتيال: يمكن استخدام علم البيانات لتحديد الأنماط والشذوذ في المعاملات المالية التي قد تشير إلى الاحتيال.
3. الرعاية الصحية: يمكن استخدام علم البيانات لتحليل السجلات الطبية والبيانات الأخرى المتعلقة بالصحة لتحسين نتائج المرضى، وتحديد المخاطر الصحية المحتملة، وإعلام سياسة الصحة العامة.

4. وسائل التواصل الاجتماعي: يمكن استخدام علم البيانات لتحليل بيانات وسائل التواصل الاجتماعي لفهم الرأي العام وتحديد الاتجاهات وتتبع انتشار المعلومات عبر الإنترنت .
5. معالجة اللغة الطبيعية : يمكن استخدام علم البيانات لمعالجة وفهم وعمل تنبؤات بناءً على اللغة البشرية، بما في ذلك روبوتات المحادثة والتعرف على الكلام وتلخيص النصوص الكبيرة .
6. رؤية الحاسوب: يمكن استخدام علم البيانات لتحليل الصور ومقاطع الفيديو ولتطوير السيارات ذاتية القيادة والتطبيقات المرئية الأخرى .
7. الصيانة التنبؤية: يمكن استخدام علم البيانات لتحليل بيانات المستشعر للتنبؤ بموعد تعطل المعدات، بحيث يمكن إصلاحها قبل تعطلها.
8. أنظمة التوصية: يمكن استخدام علم البيانات للتوصية بمنتجات أو مقالات أو خدمات للمستخدمين بناءً على تفاعلاتهم وسلوكهم السابق.

هذه مجرد أمثلة قليلة للطرق العديدة التي يمكن من خلالها تطبيق علم البيانات . ويتطور المجال باستمرار، يتم تطوير تطبيقات جديدة طوال الوقت .

باختصار، يتم استخدام علم البيانات لاستخراج الرؤى والتنبؤات من البيانات، والتي يمكن استخدامها لتحسين عملية صنع القرار وأتمتة العمليات.

## منهجية علم البيانات

منهجية علم البيانات هي إطار عام لحل المشكلات والإجابة على الأسئلة باستخدام البيانات. تتضمن العملية عادةً عدة خطوات، بما في ذلك :

1. حدد المشكلة: افهم مشكلة العمل أو سؤال البحث الذي تحاول حله. حدد بوضوح الأهداف ومعايير النجاح وأي قيود.
2. جمع وإعداد البيانات: جمع البيانات اللازمة وتجهيزها للتحليل بتنقية البيانات وتحويلها وتطبيعها.
3. استكشاف البيانات وتحليلها: استكشف البيانات لتحديد الأنماط والاتجاهات والعلاقات. استخدم تقنيات التعلم الإحصائي والآلي لتحليل البيانات وإنشاء النماذج.
4. توصيل النتائج وعرضها: قم بتوصيل نتائج التحليل من خلال أشكال مختلفة من التقارير، مثل تصور البيانات أو لوحات المعلومات أو العروض التقديمية.
5. نشر الحل وصيانتته: انشر الحل في بيئة إنتاج، حيث يمكن استخدامه لإنشاء رؤى أو عمل تنبؤات في الوقت الفعلي. راقب أداء الحل وقم بتحديثه حسب الحاجة.

ضع في اعتبارك أن هذا إطار عمل عام، وقد تختلف الخطوات والأدوات المحددة المستخدمة في مشروع علم البيانات اعتمادًا على المشكلة والبيانات والموارد المتاحة. أيضًا، إنها ليست دائمًا عملية خطية، ومن الشائع التكرار بين الخطوات أو حتى العودة وإعادة النظر في الخطوات السابقة عندما يتم الكشف عن معلومات جديدة أو

تغيير الأهداف باختصار، عملية علم البيانات هي عملية دورية لفهم المشكلة، وجمع البيانات وإعدادها، وتحليل البيانات ونمذجةها، وإيصال النتائج وتصورها، ونشر الحل والحفاظ عليه.

## أدوات علم البيانات

هناك العديد من الأدوات المختلفة التي يتم استخدامها بشكل شائع في علم البيانات، اعتمادًا على المهمة أو المشروع المحدد في متناول اليد. تتضمن بعض الأدوات الأكثر شيوعًا ما يلي:

1. لغات البرمجة Python و R : هما أكثر لغات البرمجة شيوعًا لعلم البيانات، ولكن يمكن أيضًا استخدام لغات أخرى مثل Julia أو Scala. توفر هذه اللغات مجموعة واسعة من المكتبات وأطر العمل لمعالجة البيانات وتحليلها وتصورها. معالجة البيانات وتنظيفها: توفر أدوات مثل pandas و dplyr في Python و R إمكانيات فعالة لمعالجة البيانات وتنظيفها، مما يتيح لك تبادل البيانات وإعادة تشكيلها استعدادًا للتحليل.
2. تخزين البيانات: تُستخدم قواعد البيانات مثل MySQL و PostgreSQL و MongoDB ، بالإضافة إلى أدوات تخزين البيانات مثل Hadoop و Spark ، بشكل شائع لتخزين وإدارة مجموعات كبيرة من البيانات.

3. التمثيل المرئي للبيانات: تتيح لك أدوات مثل matplotlib و ggplot و seaborn في Python و ggplot2 in R إنشاء تصورات غنية لإيصال النتائج والأفكار.
4. التعلم الآلي: توفر المكتبات مثل scikit-Learn و TensorFlow و Keras مجموعة واسعة من خوارزميات التعلم الآلي والأدوات لبناء النماذج التنبؤية .
5. معالجة اللغة الطبيعية: أدوات مثل NLTK و spaCy في python أو OpenNLP في Java ، لمعالجة نص أو كلام بلغة طبيعية واستخراج المعنى منه .
6. الأنظمة الأساسية السحابية: يوفر موفرو السحابة مثل AWS و GCP مجموعة من الخدمات لتخزين البيانات وتحليلها وتصورها على نطاق واسع، بما في ذلك تخزين البيانات المُدارة وتخزين البيانات وخدمات التعلم الآلي .
7. التعاون والمشاركة: تسمح أدوات مثل Jupyter و RStudio و Zeppelin بالمشاركة والتعاون في الكود والتصورات والنتائج، مما يسهل على علماء البيانات العمل في فرق .

هذه أمثلة قليلة على العديد من الأدوات المتاحة، والأدوات والتقنيات الجديدة تظهر باستمرار في هذا المجال. من المهم لعلماء البيانات أن يكون لديهم فهم جيد للأدوات الأكثر شيوعًا وأن يكونوا قادرين على تحديد الأداة المناسبة لمهمة معينة، ولكن أيضًا يكونون منفتحين لتعلم أدوات جديدة حيث يتطور المجال باستمرار.

## تطبيق علم البيانات

علم البيانات يمكن تطبيقه في العديد من الصناعات والمجالات المختلفة. تتضمن بعض الأمثلة على مجالات تطبيق علم البيانات ما يلي :

1. الخدمات المالية: علم البيانات يستخدم لاكتشاف الاحتيال وتحديد اتجاهات السوق وعمل تنبؤات حول أسعار الأسهم والمتغيرات المالية الأخرى .
2. الرعاية الصحية: يتم استخدام علم البيانات لتحليل السجلات الطبية وتحديد المخاطر الصحية المحتملة وتحسين نتائج المرضى.
3. البيع بالتجزئة: يتم استخدام علم البيانات لتحليل بيانات العملاء وسجل الشراء للتنبؤ بسلوك المستهلك وإبلاغ قرارات التسويق والمخزون.
4. التصنيع: يتم استخدام علم البيانات لتحليل بيانات المستشعر من معدات التصنيع للتنبؤ بموعد تعطل المعدات، بحيث يمكن إصلاحها قبل تعطلها .
5. التسويق: يتم استخدام علم البيانات لتحليل بيانات العملاء ونشاط وسائل التواصل الاجتماعي لتحديد الأنماط والاتجاهات، وإبلاغ استراتيجيات التسويق .
6. النقل: يُستخدم علم البيانات لتحسين التوجيه والجدولة لشركات التوصيل وشركات الطيران وخدمات مشاركة الرحلات .

7. الزراعة: تُستخدم علوم البيانات لتحليل بيانات الاستشعار من معدات الزراعة وأنماط الطقس لتحسين غلة المحاصيل وتحسين كفاءة العمليات الزراعية .
  8. الألعاب: تُستخدم علوم البيانات لتحسين تصميم اللعبة والتنبؤ بسلوك اللاعب وتحليل بيانات اللاعب لتحسين تجربة اللعب .
  9. الطاقة: تُستخدم علوم البيانات لتحسين استهلاك الطاقة والتنبؤ بأسعار الطاقة والطلب عليها .
  10. الرياضة: يتم استخدام علم البيانات لتحليل أداء اللاعب والتنبؤ بنتائج الألعاب وتحديد أنماط أداء الفريق واللاعب .
- هذه مجرد أمثلة قليلة للطرق العديدة التي يتم بها استخدام علم البيانات لحل مشاكل العالم الحقيقي ودفع الابتكار عبر مختلف الصناعات. مع إنشاء المزيد من البيانات، ستستمر حالات استخدام علم البيانات في النمو، مع إمكانية تحسين صنع القرار وأتمتة العمليات وإنشاء منتجات وخدمات جديدة.

### فوائد علم البيانات

يمكن أن يقدم علم البيانات مجموعة واسعة من الفوائد للشركات، بما في ذلك:

1. تحسين عملية صنع القرار: من خلال تحليل البيانات، يمكن للشركات اكتساب رؤى يمكن أن تساعد في اتخاذ قرارات أكثر استنارة. على سبيل المثال، يمكن استخدام علم البيانات

للتنبؤ بسلوك العميل وتحديد اتجاهات السوق وتحسين العمليات .

2. زيادة الكفاءة والأتمتة: يمكن استخدام علم البيانات لأتمتة المهام المتكررة وتحديد أوجه القصور في العمليات التجارية .

هذا يمكن أن يؤدي إلى توفير التكاليف وزيادة الإنتاجية .

3. رؤى أفضل للعملاء: من خلال تحليل بيانات العملاء، يمكن للشركات اكتساب فهم أعمق لاحتياجات عملائها وتفضيلاتهم، مما يمكن أن يساعدهم في تصميم منتجاتهم وخدماتهم لتلبية تلك الاحتياجات بشكل أفضل .

4. تسويق مستهدف أفضل: باستخدام تحليل البيانات، يمكن للشركات تحديد استراتيجيات التسويق الأكثر فعالية وتوجيه جهودهم التسويقية بشكل أكثر فعالية للوصول إلى العملاء الأكثر قيمة .

5. كشف الاحتيال والوقاية منه: يمكن استخدام علم البيانات لتحليل المعاملات المالية وتحديد السلوك المشبوه، مما يساعد الشركات على اكتشاف ومنع الاحتيال .

6. تحسين إدارة المخاطر: من خلال تحليل البيانات التاريخية، يمكن للشركات تحديد المخاطر المحتملة واتخاذ قرارات أكثر استنارة حول كيفية التخفيف من هذه المخاطر .

7. تنبؤ أكثر دقة: باستخدام نماذج إحصائية متقدمة، يمكن للشركات أن تقدم تنبؤات أكثر دقة حول الأحداث

والاتجاهات المستقبلية، والتي يمكن أن تساعد في التخطيط للمستقبل واتخاذ قرارات أفضل .

8. المزايا على المنافسة: مع الاستخدام المتزايد لعلوم البيانات عبر الصناعات، يمكن أن يؤدي امتلاك الخبرة واستخدام الأدوات والتقنيات إلى ميزة إستراتيجية على المنافسين والبقاء في المقدمة في السوق .

بشكل عام، يمكن أن يساعد علم البيانات الشركات على اتخاذ قرارات أفضل وتحسين الكفاءة وزيادة الإيرادات.

### دور علماء البيانات

يتطلب استخراج رؤى ذات مغزى من البيانات العمل على تحليل البيانات، حيث يأتي دور علماء البيانات. عالم البيانات متخصص في جمع البيانات وتنظيمها وتحليلها وتفسيرها للعثور على الاتجاهات والأنماط والارتباطات. يؤدي علماء البيانات دورًا أساسيًا في ضمان اتخاذ المنظمات قرارات مستنيرة. إنهم يعملون بشكل وثيق مع قادة الأعمال لتحديد أهداف محددة، مثل تحديد تقسيم العملاء وقيادة التحسينات في المنتجات والخدمات. باستخدام خوارزميات التعلم الآلي المتقدمة والنماذج الإحصائية، يمكن لعلماء البيانات فحص مجموعات البيانات الضخمة للكشف عن الأنماط والرؤى التي تساعد المؤسسات على اتخاذ قرارات سليمة. يتمتع علماء البيانات عمومًا بمجموعة من المهارات التقنية ومعرفة تفسير

البيانات وتصورها. يجب أن يكون لديهم خبرة في التحليل الإحصائي ولغات البرمجة وخوارزميات التعلم الآلي وأنظمة قواعد البيانات .  
دعنا نلقي نظرة عامة على المسؤوليات التي يتولاها علماء البيانات المحترفين .

1. جمع البيانات وتنقيتها وتنظيمها لاستخدامها في النماذج التنبؤية والتعليمية تحليل كميات هائلة من المعلومات لاكتشاف الاتجاهات والأنماط.
2. استخدام لغات البرمجة في هيكلية البيانات وتحويلها إلى معلومات قابلة للاستخدام.
3. العمل مع أصحاب المصلحة لفهم مشاكل العمل وتطوير حلول تعتمد على البيانات.
4. تطوير النماذج التنبؤية باستخدام النماذج الإحصائية للتنبؤ بالاتجاهات المستقبلية.
5. بناء نماذج التعلم الآلي وصيانتها ومراقبتها.
6. تطوير واستخدام خوارزميات التعلم الآلي المتقدمة والأساليب التحليلية الأخرى لإنشاء حلول تعتمد على البيانات وتوصيل الحلول القائمة على البيانات لأصحاب المصلحة.
7. الكشف عن الأنماط والاتجاهات المخفية في مجموعات البيانات الضخمة باستخدام مجموعة متنوعة من أدوات التنقيب عن البيانات.
8. تطوير حلول البيانات والتحقق من صحتها من خلال تصورات البيانات والتقارير ولوحات المعلومات والعروض التقديمية

في الختام، يعد دور علماء البيانات أمرًا بالغ الأهمية للشركات التي تتطلع إلى اتخاذ قرارات تعتمد على البيانات. علماء البيانات مسؤولون عن جمع البيانات وتنظيمها وتحليلها وتفسيرها لتحديد الاتجاهات والارتباطات. كما يقومون بتطوير خطوط نقل معالجة البيانات، وتقارير التصميم، ولوحات المعلومات، وتطوير النماذج للتنبؤ بالاتجاهات المستقبلية. للنجاح في هذا المجال، يحتاجون إلى فهم سياق العمل واحتياجات العميل.

### مهارات العمل في مجال البيانات

تختلف المهارات المحددة المطلوبة لأدوار البيانات اعتمادًا على الوظيفة، ولكن هناك بعض المهارات الشائعة المطلوبة عادةً للعديد من الوظائف المتعلقة بالبيانات. وتشمل هذه:

1. مهارات تحليلية قوية: القدرة على جمع وتنظيف وتحليل مجموعات كبيرة من البيانات، وتحديد الأنماط والاتجاهات، واستخلاص النتائج.
2. المهارات التقنية: إتقان لغات البرمجة مثل Python و R و SQL وغيرها من اللغات التي يشيع استخدامها لتحليل البيانات ومعالجتها. يعد الإلمام بتقنيات البيانات الضخمة مثل Hadoop و Spark ميزة إضافية للعديد من الأدوار.
3. خبرة في تمثيل البيانات: القدرة على تقديم البيانات بتنسيق سهل الفهم وجذاب بصريًا، باستخدام أدوات مثل Tableau و Power BI وأدوات تصور البيانات الأخرى.

4. مهارات الاتصال والتعاون القوية: القدرة على العمل بشكل جيد مع الآخرين، بما في ذلك أصحاب المصلحة غير التقنيين، وشرح المفاهيم والنتائج التقنية المعقدة بشكل واضح لمجموعة متنوعة من الجماهير.
5. الإلمام بتقنيات التعلم الإحصائي والآلي: معرفة الأساليب الإحصائية، مثل اختبار الفرضيات وتحليل الانحدار، والخبرة مع مكتبات وأطر التعلم الآلي مثل scikit-Learn و TensorFlow، ستكون ذات قيمة عالية .
6. الإلمام بخدمات الحوسبة السحابية مثل AWS أو Azure أو GCP.
7. اساس في الرياضيات مثل الجبر الخطي والإحصاء لفهم المفاهيم الأساسية لنماذج التعلم الآلي.
8. فهم سياسات إدارة البيانات وفقا لأفضل الممارسات.
9. الفطنة في الأعمال لتكون قادرًا على فهم كيفية استخدام البيانات لاتخاذ قرارات العمل .

باختصار، يحتاج العاملون في مجال البيانات إلى مزيج من المهارات الفنية والتحليلية والتواصلية للعمل بفعالية مع البيانات واستخراج الأفكار. يجب أن يكونوا مرتاحين للرياضيات والإحصاءات ولديهم معرفة بمجموعة متنوعة من الأدوات والتقنيات المستخدمة بشكل شائع لإدارة البيانات وتحليلها.

## اسئلة الفصل الأول

1. ما المقصود بالبيانات وما هي اوجه الاختلاف بينها وبين المعلومات.؟
2. تاتي البيانات من مصادر مختلفة تكلم عن تلك المصادر بنقاط مفصلة.؟
3. تصنيف البيانات هو عملية تنظيم البيانات في فئات او اصناف بناءً على خصائص أو سمات معينة. ما هي اهم تلك التصنيفات.؟
4. ما هي الأنواع المختلفة من النماذج التي يمكن استخدامها لنمذجة البيانات.
5. منهجية علم البيانات هي إطار عام لحل المشكلات والإجابة على الأسئلة باستخدام البيانات. تكلم عنها بنقاط متسلسلة.
6. للعمل في قطاع البيانات تحتاج الى اكتساب مهارات وخبرات محدد. بين بنقاط اهم تلك المهارات.
7. يعد دور علماء البيانات مهما للشركات التي تتطلع إلى اتخاذ قرارات تعتمد على البيانات. بين اهم ما يقوم به علماء البيانات من مهام.
8. تلکم بالتفصيل عن الاسباب التي ساهمت في زيادة اهمية علم البيانات.
9. يقدم علم البيانات مجموعة واسعة من الفوائد للشركات والمنظمات. بين ما هي هذه الفوائد.
10. ماهي الأدوات الأكثر شيوعا والتي تستخدم في تحليل ومعالجة البيانات.

## الفصل الثاني

### تحليل البيانات

#### تمهيد

يعد تحليل البيانات [Data Analysis] واعدادها فصلاً مهماً للطلاب لأنه يضع الأساس لبقية موضوعات علم البيانات. بدون تحليل البيانات واعدادها بشكل صحيح، ستكون العمليات التي يتم إجراؤها لاحقاً ضعيفة وغير موثوقة. وعليه يعد تحضير البيانات واعدادها خطوة أساسية في عملية علم البيانات، سيتم التركيز في هذا الفصل على عمليات تحليل البيانات و كل ما يرتبط بها من عمليات تهدف في المحصلة النهائية الى التوصل الى نتائج واقعية يمكن على اساسها اتخاذ القرارات المناسبة لحل المشكلات او تحقيق الاهداف.

#### فيئات تحليل البيانات

تحليل البيانات هي عملية فحص وتنظيف وتحويل ونمذجة البيانات بهدف اكتشاف معلومات مفيدة واستخلاص النتائج ودعم اتخاذ القرار. يتضمن استخدام التقنيات الإحصائية والحسابية لتحديد الأنماط والاتجاهات والرؤى في البيانات. يمكن تقسيم تحليلات البيانات إلى عدة فئات:

1. التحليلات الوصفية: يركز هذا النوع من التحليلات على تلخيص البيانات ووصفها لفهمها بشكل أفضل. يتضمن ذلك إنشاء

تقارير وتصورات لتقديم ملخص للبيانات وتحديد الأنماط والاتجاهات.

2. التحليلات التشخيصية: يستخدم هذا النوع من التحليلات لتحديد سبب حدوث المشكلة. يتضمن البحث في البيانات لفهم الأسباب الكامنة وراء نتيجة معينة .

3. التحليلات التنبؤية: يستخدم هذا النوع من التحليلات تقنيات التعلم الإحصائي والآلي لتحليل البيانات والتنبؤ بالأحداث المستقبلية. يمكن أن يشمل ذلك أشياء مثل التنبؤ بالمبيعات أو تحديد العملاء المحتملين .

4. التحليلات الوصفية: هذا النوع من التحليلات يتجاوز التوقعات ويقترح الإجراءات التي يجب اتخاذها بناءً على البيانات التي تم تحليلها. فهو يجمع بين التحليلات الوصفية والتشخيصية والتنبؤية ويقترح ما يجب على المنظمة فعله بعد ذلك .

تُستخدم تحليلات البيانات في صناعات مختلفة مثل التمويل والرعاية الصحية وتجارة التجزئة والتصنيع. يمكن تطبيق تحليلات البيانات على أنواع مختلفة من البيانات، بما في ذلك البيانات المنظمة (مثل البيانات المخزنة في قاعدة بيانات) والبيانات غير المنظمة (مثل النصوص والصور ومقاطع الفيديو).

## مراحل تحليل البيانات

تتضمن العملية المتضمنة في تحليل البيانات عدة خطوات مختلفة، وكالاتي:

- تتمثل الخطوة الأولى في تحديد متطلبات البيانات أو كيفية تجميع البيانات. قد يتم تصنيف البيانات حسب العمر أو الديموغرافية أو الدخل أو الجنس. قد تكون قيم البيانات عددية أو مقسمة حسب الفئة .
- الخطوة الثانية في تحليلات البيانات هي عملية جمعها. يمكن القيام بذلك من خلال مجموعة متنوعة من المصادر مثل أجهزة الحاسوب أو المصادر عبر الإنترنت أو الكاميرات أو المصادر البيئية أو من خلال الموظفين .
- بمجرد جمع البيانات، يجب تنظيمها بحيث يمكن تحليلها. قد يحدث هذا في جدول بيانات أو أي شكل آخر من البرامج التي يمكن أن تأخذ بيانات إحصائية.
- ثم يتم تنظيف البيانات قبل التحليل. هذا يعني أنه قد تم مسحه وفحصه للتأكد من عدم وجود ازدواجية أو خطأ، وأنه ليس غير مكتمل. تساعد هذه الخطوة في تصحيح أي أخطاء قبل أن تنتقل إلى محلل البيانات لتحليلها.

ويمكن تتبع هذه الخطوات بسلسلة متصلة من المهام التي يمكن تنفيذها بالتتابع وصولاً إلى النتائج النهائية وكالاتي:

1. جمع البيانات: تتضمن هذه المهمة الحصول على البيانات من مصادر مختلفة، مثل قواعد البيانات أو الملفات أو الأنظمة الأساسية عبر الإنترنت. هذه الخطوة حاسمة لأن جودة وكمية البيانات ستؤثر بشكل مباشر على نتائج التحليل .

2. تنظيف البيانات: تتضمن هذه المهمة تحديد وتصحيح الأخطاء والتناقضات والبيانات المفقودة في البيانات التي تم جمعها. هذه الخطوة مهمة لأنها تضمن أن البيانات دقيقة وموثوقة، وتزِيل أي قيم متطرفة وأخطاء قد تؤثر على نتائج التحليل.
  3. استكشاف البيانات: تتضمن هذه المهمة استكشاف البيانات لاكتساب فهم أعمق لها ولتحديد الأنماط والاتجاهات والعلاقات. هذه الخطوة مهمة لأنها تساعد في تحديد المشكلة التي يجب حلها والأسئلة التي يجب الإجابة عليها.
  4. تحويل البيانات: تتضمن هذه المهمة تحويل البيانات إلى تنسيق أكثر ملاءمة للتحليل. قد يشمل ذلك تسوية البيانات أو تحجيمها أو ترميزها، بالإضافة إلى إنشاء متغيرات جديدة أو تجميع البيانات.
  5. أخذ عينات البيانات: تتضمن هذه المهمة تحديد مجموعة فرعية من البيانات لاستخدامها في التحليل. هذه الخطوة مهمة لأنها تضمن أن البيانات تمثل السكان وتقلل من حجم البيانات المراد معالجتها.
- يعد استكشاف البيانات وإعدادها عملية تستغرق وقتًا طويلاً، ولكنها ضرورية لنجاح التحليل لأنها تضمن أن البيانات دقيقة وموثوقة ومناسبة للتحليل. وفيما يأتي عرض أكثر تفصيلاً لهذه الخطوات:

## اولا: جمع البيانات

يشير جمع البيانات [Data Collection] إلى عملية جمع وقياس المعلومات من مصادر مختلفة. يمكن أن يشمل ذلك الاستطلاعات والمقابلات والملاحظات والتجارب. يمكن بعد ذلك تحليل البيانات واستخدامها لاتخاذ قرارات مستنيرة أو اكتساب رؤى .

يعد جمع البيانات خطوة حاسمة في عملية البحث ويستخدم في مجموعة واسعة من المجالات، بما في ذلك الأعمال والعلوم البحتة والعلوم الاجتماعية. هناك العديد من المهارات المهمة المتعلقة بجمع البيانات التي يجب على الطلاب تعلمها منها على سبيل المثال:

- تصميم البحث: يتضمن ذلك فهم كيفية تصميم دراسة تسفر عن نتائج صحيحة وموثوقة .
- تقنيات أخذ العينات: يجب أن يتعلم الطلاب كيفية اختيار عينة تمثيلية من مجموعة سكانية .
- طرق جمع البيانات: يجب أن يكون الطلاب على دراية بمجموعة متنوعة من طرق جمع البيانات، مثل الاستطلاعات والمقابلات والتجارب .

وتجدر الاشارة الى ان هناك العديد من الأدوات والتقنيات المستخدمة بشكل شائع لجمع البيانات، بما في ذلك:

## الاستطلاعات.

الاستطلاعات او الاستبيانات هي طريقة مستخدمة على نطاق واسع لجمع البيانات في البحث. تسمح الاستطلاعات للباحثين بجمع

البيانات من عدد كبير من الأشخاص في وقت واحد، ويمكن إجراؤها بعدة طرق، بما في ذلك شخصيًا أو عبر الهاتف أو عبر الإنترنت. هناك عدة أنواع من الاستطلاعات، بما في ذلك:

- الاستبيانات ذاتية الإدارة: يقوم المشاركون بملء الاستبيان بأنفسهم، إما عبر الإنترنت أو على الورق .
- الاستبيانات التي يديرها المحاور: يسأل المحاور المدرب المشاركين عن أسئلة الاستطلاع شخصيًا أو عبر الهاتف .
- استطلاعات الويب: الاستطلاعات التي تتم إدارتها عبر الإنترنت باستخدام منصة على شبكة الإنترنت
- الاستبيانات عبر البريد: الاستطلاعات التي يتم إرسالها إلى المشاركين عبر البريد وإعادتها بالبريد أو عبر الإنترنت .
- الاستطلاعات ذات الوضع المختلط: المسوحات التي تستخدم طرقًا متعددة للإدارة، مثل مجموعة من الاستطلاعات عبر البريد والإنترنت .

و يمكن أن تكون الاستطلاعات مغلقة أو مفتوحة تحتوي الاستطلاعات المغلقة على مجموعة من الأسئلة المحددة مسبقًا التي يتم طرحها على جميع المشاركين، بينما تسمح الاستطلاعات المفتوحة بمزيد الحرية في الردود. كما يمكن أن تكون الاستطلاعات كمية أو نوعية، اعتمادًا على نوع الأسئلة وطريقة التحليل.

تتمتع الاستطلاعات بالعديد من المزايا، مثل القدرة على الوصول إلى عدد كبير من الأشخاص، وتوحيد الأسئلة، والقدرة على جمع

البيانات الفئوية والرقمية. ومع ذلك، فإن الاستطلاعات لها أيضًا بعض القيود، مثل التحيز في الاستجابة، ومعدلات الاستجابة المنخفضة، وتحيز الرغبة الاجتماعية.

بشكل عام، تعد الاستطلاعات أداة مفيدة لجمع البيانات، ولكن من المهم استخدام تصميم ومنهجية المسح المناسبين لسؤال البحث ومراعاة قيود طريقة المسح.

### المقابلات

المقابلات هي طريقة لجمع البيانات تتضمن قيام الباحث بطرح أسئلة على أحد المشاركين من أجل جمع المعلومات. يمكن إجراء المقابلات شخصيًا أو عبر الهاتف أو عبر الإنترنت. هناك عدة أنواع من المقابلات، بما في ذلك:

- المقابلات المنظمة: يتم طرح مجموعة من الأسئلة المحددة مسبقًا على جميع المشاركين بنفس الترتيب.
- المقابلات شبه المنظمة: يتم طرح مجموعة من الأسئلة المحددة مسبقًا، ولكن يتمتع القائم بإجراء المقابلة ببعض المرونة في طرح أسئلة المتابعة أو التحقيق للحصول على مزيد من المعلومات.
- المقابلات غير المنظمة: لدى القائم بإجراء المقابلة موضوع عام أو سؤال بحثي، لكن لا توجد قائمة محددة من الأسئلة. بدلاً من ذلك، يتابع القائم بإجراء المقابلة المحادثة فور وقوعها.

يمكن أن تكون المقابلات نوعية أو كمية حسب نوع الأسئلة وطريقة التحليل. وتتمتع المقابلات بالعديد من المزايا، مثل القدرة على جمع معلومات مفصلة وطرح أسئلة متابعة، والقدرة على استكشاف وجهات نظر المشاركين وخبراتهم، والقدرة على بناء علاقة مع المشاركين. ومع ذلك، فإن المقابلات لها أيضًا بعض القيود، مثل تحيز الباحث، وتكلفة إجراء المقابلات، وإمكانية تحيز الرغبة الاجتماعية.

بشكل عام، تعد المقابلات أداة قيمة لجمع البيانات، ولكن من المهم استخدام تصميم المقابلة المناسب ومنهجية لسؤال البحث، والنظر في قيود طريقة المقابلة.

### الملاحظات

الملاحظات هي طريقة لجمع البيانات يقوم الباحث من خلالها بمراقب الأشخاص أو الأحداث أو المواقف من أجل جمع البيانات. يمكن إجراء الملاحظات في البيئات الطبيعية أو في البيئات الخاضعة للرقابة، مثل المختبر. هناك عدة أنواع من الملاحظات، منها:

- الملاحظات الطبيعية: يلاحظ الباحثون الأشخاص أو الأحداث في بيئتهم الطبيعية، دون التدخل أو التلاعب بالموقف.
- الملاحظات المقيدة: يتلاعب الباحثون أو يتحكمون في متغيرات معينة من أجل دراسة التأثير على متغير تابع.

- ملاحظات بالمشاركة: يشارك الباحثون بنشاط في الموقف أو المجموعة التي تتم ملاحظتها، غالبًا من أجل اكتساب فهم أكثر تعمقًا للثقافة أو الموقف .
- ملاحظات دون مشاركة: يلاحظ الباحثون الحالة أو المجموعة دون المشاركة الفعالة فيها .

يمكن أن تكون الملاحظات كمية أو نوعية حسب نوع البيانات التي تم جمعها وطريقة التحليل. تتمتع الملاحظات بالعديد من المزايا، مثل القدرة على جمع المعلومات في ظروف العالم الحقيقي، والقدرة على مراقبة السلوكيات التي قد يكون من الصعب قياسها من خلال طرق أخرى، والقدرة على دراسة الظواهر التي قد يكون من الصعب تكرارها في المختبر. ومع ذلك، فإن الملاحظات لها أيضًا بعض القيود، مثل تحيز المراقب، وتكلفة إجراء الملاحظات، وإمكانية فقدان السلوكيات أو الأحداث المهمة .

بشكل عام، تعد الملاحظات أداة قيمة لجمع البيانات، ولكن من المهم استخدام تصميم ومنهجية الملاحظة المناسبين لسؤال البحث، والنظر في قيود طريقة المراقبة.

### التجارب

التجارب هي طريقة لجمع البيانات تتضمن معالجة واحد أو أكثر من المتغيرات المستقلة من أجل ملاحظة التأثير على المتغير التابع. غالبًا ما تُستخدم التجارب لإنشاء علاقات السبب والنتيجة، ويمكن

إجراؤها في ظروف طبيعية أو معملية. هناك عدة أنواع من التجارب،  
منها:

- التجارب الحقيقية: يتعامل الباحثون مع متغير واحد أو أكثر من المتغيرات المستقلة ويلاحظون التأثير على المتغير التابع، بينما يتحكمون في المتغيرات الخارجية الأخرى .
- أشباه التجارب: لا يمتلك الباحثون سيطرة كاملة على التلاعب بالمتغيرات المستقلة، لكنهم ما زالوا يحاولون إقامة علاقات السبب والنتيجة .
- التجارب الميدانية: يجري الباحثون تجارب في ظروف واقعية، بدلاً من المختبر .
- التجارب المعملية: يقوم الباحثون بإجراء التجارب في بيئة خاضعة للرقابة، مثل المختبر .

يمكن أن تكون التجارب كمية أو نوعية حسب نوع البيانات التي تم جمعها وطريقة التحليل. للتجارب العديد من المزايا، مثل القدرة على إقامة علاقات السبب والنتيجة، والقدرة على التحكم في المتغيرات الخارجية، والقدرة على معالجة المتغيرات من أجل دراسة تأثيرها. ومع ذلك، فإن التجارب أيضاً لها بعض القيود، مثل إمكانية التصنع، وصعوبة تكرار المواقف في العالم الحقيقي، والاعتبارات الأخلاقية للتلاعب بالمتغيرات .

بشكل عام، تعد التجارب أداة قيمة لجمع البيانات، ولكن من المهم استخدام التصميم التجريبي المناسب والمنهجية لسؤال البحث، ومراعاة قيود الطريقة التجريبية.

### السجلات وعمليات التتبع

هي البيانات التي يتم تجميعها تلقائيًا بواسطة الأنظمة، مثل ملفات سجل الخادم، وتتبع الشبكة، وتتبع التطبيقات. غالبًا ما تُستخدم هذه الأنواع من البيانات للمراقبة واستكشاف الأخطاء وإصلاحها والتحليل في مجالات مختلفة مثل تكنولوجيا المعلومات والأمن السيبراني وهندسة البرمجيات. وتنقسم إلى :

- سجلات الخادم: هي سجلات للنشاط على الخادم، بما في ذلك معلومات حول الاتصالات والطلبات والأخطاء الواردة.
- تتبع الشبكة: تتبع الشبكة عبارة عن سجلات لنشاط الشبكة، بما في ذلك معلومات حول الحزم والبروتوكولات وأداء الشبكة.
- تتبع التطبيق: تتبع التطبيق عبارة عن سجلات للنشاط داخل أحد التطبيقات، بما في ذلك معلومات حول استدعاءات الوظائف والأداء والأخطاء.

يمكن تحليل هذه الأنواع من البيانات باستخدام أدوات متخصصة مثل أدوات تحليل السجلات، وأجهزة تحليل الشبكة، وأدوات مراقبة الأداء. يمكن توفير نظرة ثاقبة لأداء النظام والأمان وسلوك المستخدم. تتمتع السجلات والتتبعات بالعديد من المزايا مثل

توفير بيانات في الوقت الفعلي، وتوفير معلومات مفصلة، وسهولة جمعها. ومع ذلك، فإن لديهم أيضًا بعض القيود مثل تكلفة تخزين وتحليل كميات كبيرة من البيانات، والحاجة إلى أدوات وخبرات متخصصة لتحليلها .

بشكل عام، تعد السجلات والتتبعات أداة مفيدة لجمع البيانات، ولكن من المهم مراعاة قيود هذه الطريقة. من المهم أيضًا أن يكون لديك سياسة استبقاء مناسبة وبروتوكولات أمان لحماية البيانات الحساسة التي يتم جمعها.

### تعقب البيانات

تعقب البيانات هي تقنية تستخدم لتعقب البيانات من مواقع الويب ومنصات الوسائط الاجتماعية ومصادر أخرى للتحليل والبحث. يتضمن استخدام برامج أو نصوص برمجية متخصصة لتعقب البيانات وجمعها تلقائيًا من موقع ويب أو مصدر آخر عبر الإنترنت. هناك عدة أنواع من تعقب البيانات، بما في ذلك :

- تعقب الويب: عملية تعقب البيانات من مواقع الويب باستخدام البرامج النصية أو البرامج الآلية .
- تعقب مواقع التواصل الاجتماعي: عملية تعقب البيانات من منصات التواصل الاجتماعي كالتغريدات والمشاركات والتعليقات .

- تعقب API : عملية تعقب البيانات من واجهة برمجة التطبيقات (API) التي توفر الوصول إلى مجموعة بيانات أو خدمة معينة .

يمكن استخدام تعقب البيانات لجمع مجموعة كبيرة من البيانات مثل أسعار المنتجات وبيانات الطقس والمقالات الإخبارية ومنشورات الوسائط الاجتماعية. يستخدم على نطاق واسع في مختلف المجالات مثل التجارة الإلكترونية والتمويل والتسويق. يتميز تعقب البيانات بالعديد من المزايا، مثل القدرة على جمع كميات كبيرة من البيانات بسرعة، والقدرة على أتمتة المهام المتكررة، والقدرة على الوصول إلى البيانات التي قد لا تكون متاحة للجمهور. ومع ذلك، فإن تعقب البيانات له أيضًا بعض القيود، مثل الحاجة إلى المهارات والأدوات المتخصصة، وخطر انتهاك شروط خدمة موقع الويب أو إلغاء القوانين واللوائح، واحتمال وجود بيانات غير دقيقة أو غير كاملة .

بشكل عام، يعد تعقب البيانات أداة مفيدة لجمع البيانات، ولكن من المهم استخدام الطريقة المناسبة لسؤال البحث، ومراعاة قيود طريقة تعقب البيانات، والامتثال للقوانين واللوائح.

هناك العديد من المهارات المهمة المتعلقة بجمع البيانات التي يجب على الطلاب تعلمها:

- تصميم البحث: يتضمن ذلك فهم كيفية تصميم دراسة تسفر عن نتائج صحيحة وموثوقة .

- تقنيات أخذ العينات: يجب أن يتعلم الطلاب كيفية اختيار عينة تمثيلية من مجموعة سكانية .
- طرق جمع البيانات: يجب أن يكون الطلاب على دراية بمجموعة متنوعة من طرق جمع البيانات، مثل الاستطلاعات والمقابلات والتجارب .
- إدخال البيانات وإدارتها: يجب أن يتعلم الطلاب كيفية إدخال البيانات وإدارتها بشكل صحيح لضمان دقتها وسلامتها .
- تحليل البيانات: يجب أن يتعلم الطلاب كيفية تحليل البيانات وتفسيرها باستخدام الأساليب الإحصائية المناسبة .
- الاعتبارات الأخلاقية: يجب على الطلاب التعرف على المبادئ الأخلاقية التي توجه البحث، مثل الموافقة المستنيرة وحماية سرية المشاركين .

### ثانياً: تنظيف البيانات

يشير تقنية او تنظيف البيانات [Data Cleaning] إلى عملية تحديد وتصحيح أو إزالة عدم الدقة والتناقضات والأخطاء في مجموعة البيانات. يمكن أن يشمل ذلك مهام مثل إزالة السجلات المكررة، وتصحيح أخطاء إدخال البيانات، ومعالجة القيم المفقودة أو الفارغة . و يعد تنظيف البيانات، المعروف أيضًا باسم المعالجة المسبقة للبيانات، خطوة مهمة في عملية تحليل البيانات. إنها عملية تحديد وتصحيح الأخطاء والتناقضات والقيم المفقودة في مجموعة

البيانات. يعد تعلم مهارات تنظيف البيانات أمرًا مهمًا للطلاب لعدة أسباب:

- تحسين جودة البيانات: يساعد تنظيف البيانات على ضمان دقة وموثوقية البيانات التي يتم تحليلها، مما يحسن جودة البحث واتخاذ القرار.
- توفير الوقت: قد يستغرق تنظيف البيانات وقتًا طويلاً، ولكن من الضروري توفير الوقت والموارد عندما يتعلق الأمر بتحليل البيانات وتفسيرها.
- تحديد الأنماط والاتجاهات: يساعد تنظيف البيانات على تحديد الأنماط والاتجاهات التي يمكن أن تكون مفيدة في عمل التنبؤات وتحديد فرص التحسين.
- تحسين تصور البيانات: يساعد تنظيف البيانات على إنشاء تصورات واضحة وموجزة تكون أكثر فاعلية في توصيل المعلومات.
- التحضير للتعلم الآلي: يعد تنظيف البيانات خطوة حاسمة في إعداد البيانات للتعلم الآلي، وهو أمر مهم بشكل متزايد في العديد من المجالات اليوم.
- الاعتبارات الأخلاقية: يعد تنظيف البيانات أيضًا خطوة مهمة لضمان عدم تحيز البيانات وحماية خصوصية بيانات المشاركين.

بشكل عام، يعد تعلم مهارات تنظيف البيانات أمرًا ضروريًا للطلاب الذين يرغبون في العمل باستخدام البيانات واتخاذ قرارات مستنيرة بناءً على تلك البيانات. وهناك العديد من الأدوات التي يمكن استخدامها لتنظيف البيانات، وفيما يأتي عرض لهذه الأدوات :

### Excel & Google Sheets

عبارة عن برامج جداول بيانات يمكن استخدامها لتنظيم البيانات وتنظيفها يدويًا. يتيح لك كلا البرنامجين إنشاء جداول البيانات وتحريرها وإجراء العمليات الحسابية وإنشاء المخططات والرسوم البيانية

Excel. هو برنامج تجاري متاح كجزء من Microsoft Office Suite، بينما Google Sheets هو برنامج مجاني قائم على الويب وهو جزء من مجموعة مستندات Google. يحتوي كلا البرنامجين على مجموعة واسعة من الميزات التي يمكن استخدامها لتنظيف البيانات بما في ذلك :

فرز البيانات وتصفيتها، إزالة التكرارات، البحث عن البيانات واستبدالها، تأكيد صحة البيانات، تحويل نص إلى أعمدة، تنسيق البيانات، نَسب البيانات وظائف VLookup و HLookup الجداول المحورية والمخططات، يحتوي Excel على ميزات ووظائف أكثر تقدمًا والتي تشمل وحدات الماكرو، والبرمجة النصية لـ VBA، والتكامل مع برامج Microsoft Office الأخرى، بينما تعد جداول بيانات Google أكثر أساسية وتستخدم في الغالب لمهام إدارة البيانات البسيطة .

بشكل عام، يعد كل من Excel و Google Sheets أدوات متعددة الاستخدامات يمكن استخدامها لتنظيف البيانات، ويعتمد الاختيار بين الاثنين على احتياجاتك الخاصة وتعقيد مهام تنظيف البيانات الخاصة بك.

### OpenRefine

كانت تعرف سابقا باسم Google Refine. هي أداة مجانية ومفتوحة المصدر لتنظيف البيانات وتحويلها. إنها أداة قوية للعمل مع مجموعات البيانات الضخمة، وهي مفيدة بشكل خاص للعمل مع البيانات غير المنظمة. يتيح لك OpenRefine استكشاف البيانات وتنظيفها وتحويلها بسهولة من خلال توفير واجهة سهلة الاستخدام تسهل أداء مهام تنظيف البيانات الشائعة. تتضمن بعض الميزات الرئيسية لبرنامج OpenRefine ما يلي :

- توجيه البيانات: يسمح لك بتحديد الأنماط والقيم المتطرفة في بياناتك بسرعة.
- تجميع البيانات: تجميع البيانات المتشابهة معًا، مما يسهل تحديد الأخطاء وتصحيحها.
- تسوية البيانات: تطابق بياناتك مع مصادر البيانات الخارجية، مثل ويكيبيديا، للمساعدة في تحسين جودة البيانات
- تحويل البيانات: يسمح لك بإجراء تحويلات معقدة للبيانات باستخدام واجهة بسيطة للتأشير والنقر.

- تصدير البيانات: تصدير البيانات بتنسيقات متنوعة، بما في ذلك CSV و JSON و RDF OpenRefine هو تطبيق قائم على Java يعمل على أنظمة التشغيل Windows و Mac و Linux ويمكن استخدامه للعمل مع مجموعة متنوعة من تنسيقات البيانات، بما في ذلك CSV و TSV و Excel و XML هي أداة قوية لتنظيف البيانات وتحويلها، وهي مفيدة بشكل خاص للعمل مع البيانات الفوضوية أو غير المهيكلة. يمكن أن يساعدك في التعرف بسرعة على الأنماط والقيم المتطرفة في بياناتك، كما أنه يجعل من السهل أداء مهام تنظيف البيانات الشائعة.

### Trifacta

هي شركة برمجيات لمعالجة البيانات وإعدادها، توفر منصة لمستخدمي الأعمال لاكتشاف وتنظيف وتنظيم مجموعات البيانات الضخمة. يستخدم برنامجهم خوارزميات التعلم الآلي لأتمتة عملية إعداد البيانات، مما يقلل الوقت والجهد اللازمين لتنظيف البيانات يدويًا. يتم استخدام تقنية Trifacta من قبل المؤسسات في مجموعة متنوعة من المجالات، بما في ذلك التمويل والرعاية الصحية وتجارة التجزئة، لتحسين سرعة ودقة تحليل البيانات.

### Talend

هي شركة توفر حلول برمجية لتكامل البيانات وإدارتها. يمكن برنامج الشركة المؤسسات من جمع البيانات وإدارتها وتحويلها ومشاركتها. تتضمن منصة Talend مجموعة واسعة من الأدوات

لتكامل البيانات وجودة البيانات وإدارة البيانات الرئيسية والبيانات الضخمة. يمكن استخدام البرنامج لتوصيل مجموعة متنوعة من مصادر البيانات والأنظمة، ويستخدم واجهة السحب والإفلات لإنشاء خطوط نقل البيانات ومهام سير العمل. يمكن نشر حلول Talend في أماكن العمل أو في السحابة، ويتم استخدامها على نطاق واسع في مختلف المجالات، بما في ذلك البيع بالتجزئة والرعاية الصحية والتمويل.

### DataWrangler

هي أداة لتنظيف وتحويل البيانات على شبكة الإنترنت تم تطويرها بواسطة مجموعة التصور المرئي بجامعة ستانفورد. تم تصميم الأداة لمساعدة المستخدمين على تنظيف البيانات وإعادة تشكيلها بسرعة وسهولة لاستخدامها في التطبيقات الأخرى. يوفر DataWrangler واجهة بسيطة تشبه جدول البيانات للعمل مع البيانات، ويتضمن مجموعة متنوعة من وظائف تنظيف البيانات وتحويلها مثل التصفية والفرز وإعادة التشكيل. كان DataWrangler متاحًا مجانًا عبر الإنترنت، ولكن تم إيقافه ولم يعد متاحًا للتنزيل. تتضمن بعض أدوات تنقية البيانات المماثلة Trifacta و OpenRefine و Talend Data Preparation .

### RapidMiner

عبارة عن نظام أساسي لعلوم البيانات يوفر مجموعة واسعة من الأدوات للمعالجة المسبقة للبيانات والنمذجة والنشر. تم تصميمه ليكون سهل الاستخدام ومرنًا، مما يسمح للمستخدمين

بالاتصال بمصادر البيانات المختلفة، وإجراء عمليات تحويل البيانات المختلفة، وإنشاء تقارير ولوحات معلومات تفاعلية. غالبًا ما يستخدم RapidMiner في مجال علوم البيانات وذكاء الأعمال والتعلم الآلي. وهو يدعم التكامل مع مجموعة واسعة من لغات البرمجة، بما في ذلك R و Python و SQL. كما يسمح باستخدام النماذج والمكتبات المبنية مسبقًا، وواجهة سهلة الاستخدام للسحب والإفلات ومجموعة واسعة من خوارزميات التعلم الآلي. وهي متوفرة في كل من الإصدارات مفتوحة المصدر والتجارية.

### KNIME

هو نظام أساسي لتحليل البيانات مفتوح المصدر يوفر مجموعة واسعة من الأدوات لاستكشاف البيانات والمعالجة المسبقة للبيانات والنمذجة وإعداد التقارير. تم تصميمه ليكون سهل الاستخدام ومرنًا، مما يسمح للمستخدمين بربط مصادر البيانات المختلفة، وإجراء عمليات تحويل وتحليل البيانات المختلفة، وإنشاء تقارير ولوحات معلومات تفاعلية. غالبًا ما يتم استخدام KNIME في مجال علوم البيانات وذكاء الأعمال والتعلم الآلي. وهو يدعم التكامل مع مجموعة واسعة من لغات البرمجة، بما في ذلك R و Python و SQL.

### Alteryx

عبارة عن نظام أساسي لتحليل البيانات يوفر مجموعة واسعة من الأدوات للمعالجة المسبقة للبيانات والنمذجة والنشر. تم تصميمه ليكون سهل الاستخدام ومرنًا، مما يسمح للمستخدمين بالاتصال بمصادر البيانات المختلفة، وإجراء عمليات تحويل البيانات المختلفة،

وإنشاء تقارير ولوحات معلومات تفاعلية. غالبًا ما يستخدم Alteryx في مجال علم البيانات وذكاء الأعمال والتعلم الآلي. يدعم التكامل مع مجموعة واسعة من لغات البرمجة، بما في ذلك R و Python ، مع واجهة السحب والإفلات. يحتوي Alteryx أيضًا على ميزة تسمى Alteryx Promote، والتي تتيح للمستخدمين نشر نماذج التعلم الآلي ومراقبتها وإدارتها Alteryx. متاح في كل من الإصدار المفتوح المصدر والتجاري. وهي معروفة بشكل خاص بقدراتها على مزج البيانات وإعداد البيانات

### DataRobot

عبارة عن نظام أساسي لعلوم البيانات يوفر مجموعة واسعة من الأدوات للمعالجة المسبقة للبيانات والنمذجة والنشر. تم تصميمه ليكون سهل الاستخدام ومرنًا، مما يسمح للمستخدمين بالاتصال بمصادر البيانات المختلفة، وإجراء عمليات تحويل البيانات المختلفة، وإنشاء تقارير ولوحات معلومات تفاعلية. غالبًا ما يتم استخدام DataRobot في مجال علم البيانات وذكاء الأعمال والتعلم الآلي، لا سيما لأتمتة عملية تطوير النموذج. وهو يدعم التكامل مع مجموعة واسعة من لغات البرمجة، بما في ذلك R & Python ، وله واجهة سحب وإفلات سهلة الاستخدام. يتيح DataRobot أيضًا للمستخدمين نشر نماذج التعلم الآلي ومراقبتها وإدارتها DataRobot. متاح في كلا الإصدارين مفتوح المصدر والتجاري. وهي معروفة بإمكانياتها الأتمتة المتقدمة والتعلم الآلي، وقدرتها على التعامل مع مجموعات البيانات واسعة النطاق وبيئات البيانات الضخمة.

هذه بعض الأدوات شائعة الاستخدام، ولكن هناك أدوات أخرى أيضاً. يعتمد اختيار الأداة على نوع البيانات وحجم مجموعة البيانات ومهام التنظيف المحددة التي يجب تنفيذها.

### مهارات منظم البيانات

الشخص الذي يعمل على تنظيف البيانات، والمعروف أيضاً باسم منظم البيانات أو المعالج المسبق للبيانات، مسؤول عن إعداد البيانات للتحليل عن طريق تحديد وتصحيح الأخطاء والتناقضات والقيم المفقودة. يمكن أن تشمل المهارات المحددة المطلوبة لهذا الدور ما يلي:

- الاهتمام الشديد بالتفاصيل: يجب أن يكون منظفو البيانات قادرين على تحديد وتصحيح الأخطاء والتناقضات والقيم المفقودة في البيانات.
- مهارات مناقشة البيانات: يجب أن يكون منظفو البيانات ماهرين في الأدوات والتقنيات لمعالجة البيانات وإعادة تشكيلها، مثل مكتبة الباندا في Excel و SQL و Python.
- معرفة تقنيات التحقق من صحة البيانات: يجب أن يكون منظفو البيانات على دراية بتقنيات فحص البيانات بحثاً عن الأخطاء والتناقضات، مثل القيم المفقودة والسجلات المكررة والقيم المتطرفة.

- مهارات تنسيق البيانات: يجب أن يكون منظفو البيانات قادرين على التأكد من أن البيانات بالتنسيق الصحيح للتحليل، مثل تحويل أنواع البيانات وتطبيع البيانات .
- مهارات تحويل البيانات: يجب أن يكون منظفو البيانات قادرين على تطبيق العمليات الحسابية والإحصائية على البيانات من أجل إعدادها للتحليل، مثل القياس والتوسيط وإنشاء متغيرات جديدة .
- مهارات إزالة البيانات المكررة: يجب أن يكون منظفو البيانات قادرين على تحديد وإزالة البيانات المكررة من مجموعة البيانات .
- مهارات تقييم جودة البيانات: يجب أن يكون منظفو البيانات قادرين على تقييم الجودة الإجمالية للبيانات وتحديد ما إذا كانت مناسبة للتحليل .
- مهارات توثيق البيانات: يجب أن يكون منظفو البيانات قادرين على توثيق عملية التنظيف، بما في ذلك الخطوات المتخذة والمشكلات التي تمت مواجهتها ومجموعة البيانات النهائية التي تم تنظيفها .
- مهارات إدارة البيانات: يجب أن يكون لمنظفو البيانات معرفة بأفضل ممارسات إدارة البيانات، مثل خصوصية البيانات والأمان والامتثال .

- مهارات اتصال قوية: يجب أن يكون منظفو البيانات قادرين على التواصل مع أصحاب المصلحة لفهم احتياجاتهم من البيانات وشرح عملية التنظيف وأي مشكلات تواجههم .
  - القدرة على التكيف: يجب أن يكون منظفو البيانات قادرين على التكيف مع أنواع مختلفة من البيانات وطرق التنظيف لأن البيانات التي يعملون معها يمكن أن تختلف كثيرًا .
  - المهارات الفنية: يجب أن يتمتع منظفو البيانات بفهم جيد لهياكل البيانات وقواعد البيانات، بالإضافة إلى لغات البرمجة مثل Python أو SQL أو R للتعامل مع مهام تنظيف البيانات .
- يعد تنظيف البيانات خطوة حاسمة في عملية تحليل البيانات، ويؤدي الشخص الذي يعمل على تنظيف البيانات دورًا حيويًا في ضمان أن البيانات جاهزة للتحليل ويمكن أن تؤدي إلى نتائج دقيقة وموثوقة.

### ثالثًا: استكشاف البيانات

استكشاف البيانات [Data Exploration] هو عملية تحليل وفهم خصائص مجموعة البيانات. هذه الخطوة مهمة في عملية علم البيانات لأنها تساعد على تحديد الأنماط والاتجاهات والعلاقات في البيانات، وتساعد على تحديد المشكلة التي يجب حلها والأسئلة التي يجب الإجابة عليها. تتضمن بعض المهام الرئيسية التي ينطوي عليها استكشاف البيانات ما يلي :

1. تلخيص البيانات: تتضمن هذه المهمة إنشاء إحصائيات وصفية مثل المتوسط والوسيط والانحراف المعياري وما إلى ذلك لفهم توزيع البيانات .
  2. تصور البيانات: تتضمن هذه المهمة إنشاء تمثيلات مرئية للبيانات مثل الرسوم البيانية والمخططات المبعثرة ومخططات الصندوق للمساعدة في تحديد الأنماط والعلاقات .
  3. تحديد القيم المتطرفة: تتضمن هذه المهمة تحديد وتحليل نقاط البيانات البعيدة عن نقاط البيانات العادية. يمكن أن تحدث هذه القيم المتطرفة بسبب أخطاء في جمع البيانات أو إدخالها .
  4. التحقيق في البيانات المفقودة: تتضمن هذه المهمة تحديد وتحليل البيانات المفقودة في مجموعة البيانات. من المهم أن تفهم سبب فقدان البيانات وكيف يمكن أن تؤثر على نتائج التحليل
  5. فحص الارتباطات: تتضمن هذه المهمة تحديد وتحليل العلاقات بين المتغيرات المختلفة في البيانات. يمكن القيام بذلك باستخدام تقنيات مثل مصفوفات الارتباط وخرائط الحرارة .
- يعد استكشاف البيانات عملية تكرارية، وقد يتطلب جولات متعددة من تنظيف البيانات وتحويلها وتصورها. يمكن لنتائج عملية استكشاف البيانات أن تفيد في اتخاذ القرارات في مراحل إعداد البيانات والنمذجة.

## مهارات مستكشف البيانات

يجب ان يمتلك الشخص الذي يعمل على استكشاف البيانات مجموعة من المهارات التي تشمل :

- تصور البيانات: يجب أن يكونوا بارعين في إنشاء تصورات واضحة وفعالة يمكن استخدامها لإيصال النتائج إلى أصحاب المصلحة. يتم استخدام أدوات مثل Tableau و Power BI و Ggplot وغيرها بشكل شائع لهذا الغرض .
- البيانات الفوضوية: يجب أن يكونوا قادرين على العمل مع مجموعات البيانات الضخمة ومعالجتها من أجل إجراء التحليل. يتضمن ذلك معالجة القيم المفقودة والقيم المتطرفة والمتغيرات الفئوية .
- المهارات التحليلية: يجب أن يكونوا قادرين على تحديد الأنماط والاتجاهات في البيانات، واستخدام التقنيات الإحصائية وتقنيات التعلم الآلي لعمل تنبؤات واستنتاجات .
- معرفة قوية بالمجال: يجب أن يكون لديهم فهم جيد للصناعة و سياق العمل الذي يتم فيه تحليل البيانات، ومدى ارتباطها بالأهداف العامة للمؤسسة .
- التواصل: يجب أن يكونوا قادرين على توصيل نتائجهم إلى أصحاب المصلحة غير التقنيين بطريقة واضحة وموجزة .
- المهارات التقنية: يجب أن يكون لديهم فهم جيد لهياكل البيانات وقواعد البيانات، بالإضافة إلى لغات البرمجة مثل

Python أو R أو SQL أو غيرها للتعامل مع مهام استكشاف البيانات .

- إدارة المشروع: يجب أن يكونوا قادرين على إدارة المشاريع من البداية إلى النهاية، بما في ذلك تحديد نطاق المشروع، وجمع البيانات، وتحليلها، وتقديم النتائج إلى أصحاب المصلحة .
- حل المشكلات: يجب أن يكون لديهم القدرة على تحديد وحل المشكلات المتعلقة باستكشاف البيانات وتحليلها، وتقديم توصيات بناءً على النتائج .

يعد استكشاف البيانات خطوة مهمة في عملية تحليل البيانات ويؤدي الشخص الذي يعمل على استكشاف البيانات دورًا حيويًا في فهم خصائص مجموعة البيانات وتحديد الرؤى التي يمكن أن تقود قرارات العمل.

#### رابعاً: تحويل البيانات

تحويل البيانات [ Data Transformation ] هو عملية تحويل البيانات من تنسيق أو بنية إلى أخرى لجعلها أكثر ملاءمة لغرض معين أو لدمجها مع أنظمة أخرى. هناك العديد من الأدوات والتقنيات التي يمكن استخدامها لإجراء تحويل البيانات، بما في ذلك :

1. أدوات تكامل البيانات: يمكن استخدام هذه الأدوات، مثل Talend و Informatica ، لاستخراج وتحويل وتحميل بيانات (ETL) من مصادر متعددة إلى نظام مستهدف .

2. أدوات تعيين البيانات: يمكن استخدام هذه الأدوات، مثل MapForce، لتحديد التعيين بين هياكل البيانات المصدر والهدف وإنشاء منطق التحويل تلقائيًا.
  3. لغات البرمجة: يمكن أيضًا إجراء تحويل البيانات باستخدام لغات البرمجة مثل Python و R، مع استخدام مكتبات مثل Pandas و Dplyr و tidyR و SQL.
  4. جداول البيانات: يمكن أيضًا إجراء تحويل البيانات باستخدام برامج جداول البيانات مثل Microsoft Excel و Google Sheets باستخدام الصيغ والفلاتر والجداول المحورية.
  5. أدوات تجاذب البيانات: تعد Trifacta و OpenRefine و Talend Data Preparation بعضًا من أدوات معالجة البيانات التي يمكن استخدامها لإجراء تحويل البيانات، فهي توفر واجهة بسيطة وسهلة الاستخدام وأتمتة بعض عمليات تنظيف البيانات وإعادة تشكيلها.
  6. أدوات جودة البيانات: يمكن استخدام أدوات مثل Informatica Data Quality لإجراء التحقق من صحة البيانات وتوحيدها وإلغاء البيانات المكررة، ويمكن أيضًا استخدام هذه الأدوات لتحويل البيانات عند إجراء عمليات تنظيف البيانات.
- يعتمد اختيار الأداة أو التقنية على المتطلبات المحددة لمشروع تحويل البيانات ونوع البيانات التي يتم تحويلها.

### خامساً: نمذجة البيانات

ضمن عملية التحليل تعد نمذجة البيانات [Data Modeling] عملية إنشاء تمثيل مبسط للبيانات من أجل فهمها والتنبؤ بها. هناك عدة أنواع من النماذج التي يمكن استخدامها في تحليلات البيانات :

- النماذج الإحصائية: تستخدم هذه النماذج تقنيات إحصائية لوصف العلاقات بين المتغيرات في مجموعة البيانات. تشمل الأمثلة الانحدار الخطي والانحدار اللوجستي ونماذج السلاسل الزمنية .
- نماذج التعلم الآلي: تستخدم هذه النماذج الخوارزميات للتعلم من البيانات وإجراء التنبؤات. تشمل الأمثلة أشجار القرار والغابات العشوائية والشبكات العصبية .
- نماذج المحاكاة: تستخدم هذه النماذج معادلات رياضية لمحاكاة كيفية عمل النظام. تشمل الأمثلة ديناميكيات الأنظمة والنماذج القائمة على الوكيل .
- نماذج التحسين: تستخدم هذه النماذج تقنيات رياضية لإيجاد أفضل حل لمشكلة ما. تتضمن الأمثلة البرمجة الخطية والبرمجة ذات الأعداد الصحيحة المختلطة .

تتضمن بيانات النمذجة عدة خطوات، بما في ذلك :

1. المعالجة المسبقة للبيانات: يتضمن ذلك تنظيف البيانات وإعدادها للتحليل، مثل التعامل مع القيم المفقودة، والقيم المتطرفة، والمتغيرات الفئوية .

2. اختيار النموذج: يتضمن اختيار النموذج المناسب بناءً على خصائص البيانات وأهداف التحليل .
  3. تدريب النموذج: يتضمن ذلك ملاءمة النموذج للبيانات وتعديل معلمات النموذج لتحسين الأداء .
  4. تقييم النموذج: يتضمن ذلك تقييم أداء النموذج باستخدام تقنيات مثل التحقق المتقاطع ومقاييس التقييم مثل الدقة والدقة والتذكر ودرجة F1.
  5. نشر النموذج: يتضمن ذلك وضع النموذج في الإنتاج، وإتاحته للاستخدام من قبل مستخدمي الأعمال أو دمجها في نظام أكبر .
- تجدر الإشارة إلى أن نمذجة البيانات هي عملية تكرارية وقد تتطلب عدة تكرارات للعثور على أفضل نموذج يناسب البيانات وأهداف التحليل.

### سادساً: عينات البيانات

أخذ عينات البيانات هو عملية اختيار مجموعة فرعية من البيانات من مجموعة بيانات أكبر للتحليل. يتم ذلك لجعل التحليل أكثر كفاءة وقابلية للإدارة، وكذلك لتقليل تكلفة تخزين البيانات ومعالجتها. هناك عدة طرق مختلفة لأخذ عينات البيانات، بما في ذلك:

1. أخذ عينات عشوائية بسيطة: طريقة بسيطة ومباشرة لاختيار عينة من مجموعة بيانات أكبر، حيث يكون لكل نقطة بيانات فرصة متساوية في الاختيار .

2. أخذ العينات الطبقيّة: طريقة لأخذ العينات حيث يتم تقسيم البيانات إلى مجموعات فرعية (طبقات) بناءً على خاصية معينة، ويتم اختيار عينة من كل طبقة. هذه الطريقة مفيدة عندما يكون للبيانات بنية معروفة أو عندما تحتاج مجموعات فرعية معينة إلى زيادة تمثيلها في العينة .
3. أخذ العينات العنقودية: طريقة لأخذ العينات حيث يتم تقسيم البيانات إلى مجموعات (مجموعات) ويتم اختيار عينة من المجموعات، ويتم تضمين جميع نقاط البيانات داخل تلك المجموعات في العينة. هذه الطريقة مفيدة عندما يتم تنظيم البيانات في مجموعات أو عناقيد تحدث بشكل طبيعي .
4. أخذ العينات المنتظم: طريقة لأخذ العينات حيث يتم اختيار البيانات على فترات منتظمة. هذه الطريقة مفيدة عندما يتم ترتيب البيانات بطريقة ما، ويتم اختيار نقطة بداية عشوائية ويتم تحديد البيانات في فاصل زمني ثابت.
5. أخذ العينات الملائمة: طريقة لأخذ العينات حيث يتم اختيار البيانات على أساس الملائمة. تُستخدم هذه الطريقة عندما يتم جمع البيانات بطريقة غير عشوائية، مثل سؤال الأشخاص الأوائل الذين يمشون بجوار المتجر . وتسمى أحيانا العينة العرضية.
6. أخذ عينات التمهيدي: طريقة لأخذ العينات حيث يتم اختيار عينات متعددة بشكل عشوائي مع الاستبدال، فهي تسمح بتقدير توزيع أخذ العينات للمقدر .

من المهم ملاحظة أن أخذ عينات البيانات عملية معقدة، وأن اختيار طريقة أخذ العينات سيعتمد على سؤال البحث المحدد، وخصائص البيانات، والموارد المتاحة. فضلاً عن ذلك، يعد حجم العينة جانباً مهماً آخر يجب مراعاته، فكلما زاد حجم العينة زادت موثوقية النتائج.

### سابعاً: تصور البيانات

تصور البيانات [Data Visualization] هو عملية إنشاء تمثيلات بيانية للبيانات من أجل توصيل المعلومات بفعالية وكفاءة. تتضمن بعض المبادئ الأساسية لتصور البيانات الفعال ما يلي:

1. اختر نوع المخطط المناسب: يتم تمثيل الأنواع المختلفة من البيانات بشكل أفضل من خلال أنواع مختلفة من المخططات. على سبيل المثال، تعد المخططات الشريطية هي الأفضل لمقارنة البيانات الفئوية، بينما تعد المخططات الخطية هي الأفضل لعرض الاتجاهات بمرور الوقت.
2. استخدم تصميمًا واضحًا وبسيطًا: حافظ على تصميم التصور الخاص بك نظيفًا وبسيطًا، مع تسلسل هرمي واضح للمعلومات.
3. تجنب استخدام الكثير من الألوان أو عناصر المخطط، لأن ذلك قد يجعل من الصعب تفسير التصور.
4. استخدم المقاييس المناسبة: استخدم المقاييس المناسبة للبيانات التي يتم تمثيلها. على سبيل المثال، لا تستخدم مقياسًا

- كبيرًا لقيم البيانات الصغيرة، لأن هذا يمكن أن يجعل الاختلافات بين نقاط البيانات تبدو أكبر مما هي عليه في الواقع.
5. اختر التسمية بوضوح: قم بتسمية عناصر المخطط الخاص بك بوضوح واستخدم أسلوب تصنيف متسق في جميع أنحاء التصور. سيسهل هذا على جمهورك فهم البيانات.
6. توفير السياق: وفر سياقًا لبياناتك من خلال تضمين المعلومات ذات الصلة مثل وحدات القياس والنطاقات الزمنية وأي معلومات أخرى من شأنها مساعدة الجمهور على فهم البيانات.
7. استخدام التصورات التفاعلية: تسمح التصورات التفاعلية للمستخدمين باستكشاف البيانات والتفاعل معها، وتوفر مرونة ورؤى أكثر من التصورات الثابتة.
8. الاختبار والتكرار: اختبر التصور الخاص بك مع عينة من جمهورك وقم بالتكرار بناءً على ملاحظاتهم، وهذا سيساعدك على تحديد أي مشكلات وتحسين فعالية التصور.
9. أخبر قصة: تصور البيانات هو طريقة لإخبار قصة بالبيانات، يجب تصميم تصور البيانات لنقل رسالة واضحة أو نظرة ثاقبة للجمهور.

هذه بعض المبادئ الأساسية لتصور البيانات الفعال. الهدف من تصور البيانات هو جعل البيانات أكثر سهولة وفهمًا للجمهور، حتى يتمكنوا من اتخاذ قرارات مستنيرة بشكل أفضل.

### ثامناً: التحقق من البيانات

التحقق من صحة البيانات هو عملية التأكد من أن البيانات التي يتم إدخالها في النظام دقيقة وكاملة ومتسقة. إنها خطوة مهمة في عملية إدارة البيانات وتساعد على ضمان أن البيانات مناسبة للاستخدام المقصود. هناك عدة طرق للتحقق من صحة البيانات، بما في ذلك :

1. التحقق من صحة بناء الجملة: التحقق من أن البيانات تتوافق مع تنسيق أو بنية معينة، مثل حقل التاريخ الذي يقبل فقط التواريخ الصالحة.
2. التحقق من صحة النطاق: التحقق من أن البيانات تقع ضمن نطاق محدد من القيم، مثل حقل درجة الحرارة الذي يقبل فقط القيم بين -40 و 100 درجة مئوية.
3. التحقق من صحة الأرقام: التحقق من أن رقمًا أو حرفًا معينًا في البيانات مشتق رياضيًا من بقية البيانات، مثل رقم بطاقة الائتمان أو رقم ISBN.
4. التحقق من صحة البحث: التحقق من وجود البيانات في قائمة مرجعية، مثل التحقق من وجود رمز منتج في قائمة أكواد المنتج الصالحة.
5. التحقق من صحة قاعدة العمل: التحقق من أن البيانات تتوافق مع قواعد عمل معينة، مثل التحقق من أن تاريخ الميلاد ليس في المستقبل.

6. التحقق من الصحة عبر الحقول: التحقق من تناسق البيانات عبر الحقول المختلفة، مثل التحقق من أن الكمية المطلوبة أقل من الكمية الموجودة في المخزون .

يتم إجراء التحقق من صحة البيانات عادةً في وقت إدخال البيانات ويمكن إجراؤه من خلال طرق يدوية أو آلية. يمكن تنفيذ التحقق الآلي من خلال لغات البرمجة، مثل Java أو Python أو SQL. من المهم التحقق من صحة البيانات لضمان جودة البيانات وتكامل المعلومات في النظام، مما سيساعد على اتخاذ قرارات أفضل وتحسين العمليات وتقليل الأخطاء.

### تاسعاً: تخزين البيانات

يشير تخزين البيانات [Data storage] الى الأساليب والتقنيات المستخدمة للاحتفاظ بالمعلومات الرقمية على جهاز حاسوب أو أي جهاز إلكتروني آخر. يمكن أن يشمل ذلك محركات الأقراص الثابتة ومحركات الأقراص ذات الحالة الصلبة ومحركات الأقراص المحمولة وخدمات التخزين السحابية. يمكن تخزين البيانات بتنسيقات متنوعة، بما في ذلك النصوص والصور والصوت والفيديو. يعتمد اختيار طريقة التخزين على نوع البيانات المخزنة، ومقدار مساحة التخزين المطلوبة، وإمكانية الوصول واحتياجات الأمان للبيانات.

هناك العديد من الأدوات المختلفة المتاحة لإدارة تخزين البيانات والعمل به، بما في ذلك:

1. برامج النسخ الاحتياطي والاسترداد: تساعد هذه الأدوات في إنشاء نسخ احتياطية للبيانات واستعادتها في حالة فقد البيانات .
2. برنامج إدارة الملفات: تساعد هذه الأدوات في تنظيم وإدارة كميات كبيرة من البيانات على جهاز حاسوب أو شبكة .
3. برامج ضغط البيانات وتشفيرها: تساعد هذه الأدوات في تقليل مقدار مساحة التخزين اللازمة للبيانات، وحماية البيانات من الوصول غير المصرح به .
4. خدمات التخزين السحابية: تتيح للمستخدمين تخزين البيانات والوصول إليها عبر الإنترنت، باستخدام موفري خدمات مثل Amazon S3 و Google Drive و Microsoft OneDrive.

هذه مجرد أمثلة قليلة على العديد من الأدوات المتاحة لتخزين البيانات. ستعتمد أفضل أداة لمهمة معينة على الاحتياجات المحددة للمستخدم أو المنظمة.

### مهارات محلي البيانات

محلل البيانات [Data analyst] هو الشخص المسؤول عن تحليل وتفسير البيانات المعقدة لمساعدة المؤسسات على اتخاذ قرارات أفضل. يستخدم التقنيات الإحصائية والحسابية لتحديد الأنماط والاتجاهات والرؤى في البيانات وإيصال نتائجهم إلى أصحاب المصلحة. يعمل محللو البيانات في مجموعة متنوعة من المجالات، بما في ذلك التمويل والرعاية الصحية وتجارة التجزئة والتصنيع، وقد

يكونون جزءًا من فريق تحليلات أو علوم بيانات أكبر. يمكن أن تختلف المسؤوليات المحددة لمحلل البيانات اعتمادًا على المنظمة والمشروع، ولكنها تشمل بشكل عام:

- جمع وتحليل مجموعات كبيرة من البيانات من مصادر مختلفة.
- تنظيف وتجهيز البيانات للتحليل.
- استخدام تقنيات التعلم الإحصائي والآلي لنمذجة البيانات وعمل التنبؤات.
- إنشاء تصورات البيانات لتوصيل النتائج إلى أصحاب المصلحة.
- تحديد الأنماط والاتجاهات في البيانات وتقديم التوصيات للعمل.
- إبلاغ النتائج والتوصيات لأصحاب المصلحة بطريقة واضحة وموجزة.
- التعاون مع فرق متعددة الوظائف لدمج تحليل البيانات في عمليات صنع القرار.
- مراقبة البيانات باستمرار وتحديث النماذج حسب الحاجة.
- مواكبة أدوات وتقنيات ومنهجيات تحليل البيانات الجديدة.
- التواصل مع مستخدمي الأعمال لفهم احتياجات البيانات الخاصة بهم.

غالبًا ما يعمل محللو البيانات عن كثب مع علماء البيانات ومحلي ذكاء الأعمال ومهندسي البيانات، الذين يشكلون معًا فريق تحليلات البيانات.

يحتاج محللو البيانات عادةً إلى مزيج من المهارات الفنية والتحليلية من أجل أن يكونوا فعالين. تتضمن بعض المهارات الأساسية ما يلي:

- البرمجة: يجب أن يتقن محللو البيانات لغة برمجة واحدة أو أكثر، مثل R أو Python أو SQL أو SAS، من أجل معالجة البيانات وتحليلها.
- تصور البيانات: يجب أن يكون محللو البيانات قادرين على إنشاء تصورات واضحة وفعالة يمكن استخدامها لإيصال نتائجهم إلى الآخرين. يتم استخدام أدوات مثل Tableau و Power BI و ggplot بشكل شائع لهذا الغرض.
- تنظيف البيانات: يحتاج محللو البيانات إلى أن يكونوا قادرين على العمل مع البيانات الفوضوية وغير المكتملة وتنظيفها بطريقة يمكن استخدامها للتحليل. يتضمن ذلك التعامل مع القيم المفقودة والقيم المتطرفة وأنواع البيانات.
- التحليل الإحصائي: يجب أن يكون محللو البيانات على دراية بالتقنيات الإحصائية مثل اختبار الفرضيات وتحليل الارتباط وتحليل الانحدار.

- التعلم الآلي: تعد معرفة تقنيات التعلم الآلي وكيفية تطبيقها على مشاكل العالم الحقيقي مهارة مرغوبة للغاية لمحلي البيانات .
- إدارة قواعد البيانات: تعد معرفة إدارة قواعد البيانات و SQL أمرًا مهمًا لمحلي البيانات لأنهم غالبًا ما يعملون مع مجموعات البيانات الضخمة المخزنة في قواعد البيانات .
- التواصل: يجب أن يكون محللو البيانات قادرين على توصيل نتائجهم إلى أصحاب المصلحة غير التقنيين بطريقة واضحة وموجزة .
- الفطنة التجارية: يعد فهم سياق العمل الذي يتم فيه تحليل البيانات وكيفية ارتباطها بالأهداف العامة للمؤسسة أمرًا مهمًا لمحلي البيانات .
- الفضول: يجب أن يكون محللو البيانات فضوليين وأن يستمتعوا باستكشاف البيانات وتجربتها من أجل الكشف عن رؤى قيمة .
- القدرة على التكيف: يجب أن يكون محللو البيانات قادرين على التكيف مع التقنيات والأدوات والمنهجيات الجديدة حيث يتطور مجال تحليلات البيانات باستمرار .
- إدارة المشروع: يجب أن يكون محللو البيانات قادرين على إدارة المشاريع من البداية إلى النهاية، بما في ذلك تحديد نطاق المشروع، وجمع البيانات، وتحليلها، وتقديم النتائج إلى أصحاب المصلحة.

## اسئلة الفصل الثاني

1. ما المقصود بعملية تحليل البيانات و ما هي فيئاتها؟
2. كيف يتم جمع البيانات . وما هي اهم أدوات جمعها؟
3. ما الفرق بين كل من الاستبيان والمقابلة والملاحظة ومتى يتم استخدام كل منها؟
4. ما المقصود بعملية تنظيف البيانات ؟ وما هي اهم أدواتها؟
5. ما هي المهارات المطلوب بمن يعمل في مجال تنظيف البيانات؟
6. ما المقصود بعملية استكشاف البيانات؟ وما هي المهارات المطلوب بمن يقوم بها؟
7. ما المقصود بعملية تحويل البيانات؟ وما هي اهم أدواتها وتقنياتها؟
8. ما المقصود بنمذجة البيانات؟ وما هي اهم انواع النماذج؟
9. ما المقصود بتصوير البيانات؟ وما هي اهم مبادئها؟
10. تكلم بنقاط عن المهارات الواجب توفرها بمن يقوم بعمليات تحليل البيانات.

## الفصل الثالث

### البيانات الضخمة

#### تمهيد

يهدف هذه الفصل إلى تقديم نظرة عامة شاملة عن مفهوم البيانات الضخمة وتطبيقاتها، والتي ستكون مفيدة للمهنيين والباحثين والطلاب المهتمين بفهم أساسيات البيانات الضخمة وتأثيرها في مختلف المجالات.

سوف نتعرف في هذا الفصل على تاريخ مفهوم البيانات الضخمة [Big Data]، والتحديات والفرص التي يقدمها، والتقنيات المستخدمة لإدارة وتحليل البيانات الضخمة. سنناقش أيضاً المصادر المختلفة للبيانات الضخمة، وأهمية حوكمة البيانات والخصوصية، وتأثير البيانات الضخمة على مختلف الصناعات والمجالات. سنستكشف أيضاً دور البيانات الضخمة في دفع عجلة الابتكار وأهمية البيانات الضخمة في عالم اليوم القائم على البيانات. سنختتم بمناقشة المهارات والمعرفة المطلوبة للمهنيين العاملين في مجال البيانات الضخمة وفوائد دراسة البيانات الضخمة لطلاب علوم المعلومات. يهدف هذه الفصل إلى تقديم نظرة عامة شاملة عن مفهوم البيانات الضخمة وتطبيقاتها في العالم الحالي، وستكون مفيدة للمهنيين والباحثين والطلاب المهتمين بفهم أساسيات البيانات الضخمة وتأثيرها في مختلف المجالات.

## مفهوم البيانات الضخمة

تعود جذور مفهوم "البيانات الضخمة" إلى أواخر التسعينيات وأوائل القرن الحادي والعشرين، عندما بدأت المؤسسات في جمع وتخزين كميات كبيرة من البيانات من مصادر مختلفة، مثل معاملات العملاء وسجلات مواقع الويب. في ذلك الوقت، كان تخزين ومعالجة مثل هذه المجموعات الكبيرة من البيانات يمثل تحديًا، ويتطلب تقنيات ومقاربات جديدة للتعامل مع حجم البيانات وتعقيدها.

في عام 2001، حدد دوج لاني، المحلل في شركة Gartner، للبيانات الضخمة ثلاث صفات اختصرها في: 3Vs (volume, variety and velocity) والحجم والسرعة والتنوع. وناقش بأن الحجم المتزايد للبيانات التي يتم إنشاؤها، والسرعة التي يتم إنشاؤها بها، وتنوع أنواع البيانات (مثل البيانات المهيكلة وغير المهيكلة) تجعل من الصعب على أدوات وتقنيات إدارة البيانات التقليدية مواكبة ذلك. في السنوات التالية، تم تطوير تقنيات وأطر عمل جديدة، مثل Hadoop وSpark، للتعامل مع تخزين ومعالجة البيانات الضخمة.

استندت هذه التقنيات إلى مفهوم الحوسبة الموزعة، والتي تسمح بتخزين البيانات ومعالجتها عبر أجهزة متعددة بالتوازي، مما يجعل من الممكن التعامل مع حجم وتعقيد البيانات الضخمة. فضلًا عن ذلك، تم تطوير قواعد بيانات NoSQL، مثل MongoDB وCassandra، للتعامل مع تخزين واسترجاع البيانات غير المهيكلة، وتم تطبيق خوارزميات التعلم الآلي على البيانات الضخمة للكشف عن الرؤى والتنبؤ. مع استمرار نمو حجم البيانات التي يتم إنشاؤها،

وانخفاض تكلفة التخزين والمعالجة، اكتسب مفهوم البيانات الضخمة قوة دفع في الصناعة، وبدأت المؤسسات في الاستثمار في مشاريع البيانات الضخمة لاكتساب رؤى وتحسين عملياتها. في السنوات الأخيرة، مع ظهور الحوسبة السحابية، أصبحت البيانات الضخمة أكثر سهولة وفعالية من حيث التكلفة للمؤسسات من جميع الأحجام. فضلاً عن ذلك، يتم استخدام البيانات الضخمة في مجالات مختلفة مثل الرعاية الصحية والتمويل وتجارة التجزئة وغيرها الكثير لاكتساب رؤى وتحسين عملية صنع القرار ودفع الابتكار.

تشير البيانات الضخمة إلى مجموعات كبيرة ومعقدة من البيانات التي يصعب معالجتها وتحليلها باستخدام أدوات معالجة البيانات التقليدية. إذ تتميز البيانات الضخمة بالحجم والسرعة والتنوع. يشير الحجم إلى الكمية الكبيرة من البيانات التي يتم إنشاؤها يوميًا، وتشير السرعة إلى السرعة التي يتم بها إنشاء البيانات ويشير التنوع إلى أنواع مختلفة من البيانات مثل البيانات المنظمة وشبه المنظمة وغير المنظمة. غالبًا ما تتطلب البيانات الضخمة تقنيات متخصصة، مثل المعالجة الموزعة والتعلم الآلي، للتعامل مع الحجم والتعقيد. يمكن استخدام تحليلات البيانات الضخمة للكشف عن رؤى جديدة، وتحديد الأنماط، وعمل تنبؤات يمكن أن تساعد في اتخاذ قرارات الأعمال وتحفيز النمو.

## مصادر البيانات الضخمة

هناك العديد من المصادر التي تنتج عنها البيانات الضخمة،

بما في ذلك :

1. وسائل التواصل الاجتماعي: تنشئ منصات الوسائط الاجتماعية مثل Facebook و Twitter و Instagram كميات كبيرة من البيانات من خلال تفاعلات المستخدم، مثل الإعجابات والمشاركة والتعليقات.
2. أجهزة إنترنت الأشياء (IoT): يمكن لأجهزة إنترنت الأشياء، مثل الأجهزة المنزلية الذكية والأجهزة القابلة للارتداء والمركبات، توليد كميات كبيرة من البيانات من خلال بيانات المستشعر والقياس عن بُعد.
3. ملفات السجل: يمكن أن توفر ملفات السجل التي تم إنشاؤها بواسطة الخوادم ومواقع الويب والتطبيقات معلومات قيمة حول سلوك المستخدم وأداء النظام.
4. التجارة الإلكترونية: تُنشئ منصات التجارة الإلكترونية مثل eBay و Amazon كميات كبيرة من البيانات حول سلوك العملاء ومبيعات المنتجات ومعلومات الشحن.
5. أجهزة الجوال: تقوم الأجهزة المحمولة مثل الهواتف الذكية والأجهزة اللوحية بإنشاء بيانات حول الموقع والاستخدام وسجل التصفح.

6. المالية والمصرفية: تقوم المؤسسات المالية والبنوك بإنشاء كميات كبيرة من البيانات حول المعاملات وسلوك العملاء واتجاهات السوق .

7. الرعاية الصحية: السجلات الصحية الإلكترونية (EHRs) ومستودعات البيانات السريرية (CDRs) وأنظمة البيانات الطبية الأخرى تولد كميات كبيرة من البيانات حول صحة المريض والعلاج والنتائج .

8. التعاملات الحكومية: تقوم الجهات الحكومية بإنتاج كميات كبيرة من البيانات حول التركيبة السكانية، والمؤشرات الاقتصادية، والبيانات البيئية.

9. الاستطلاعات والبحوث والبيانات العلمية: تُنشئ المسوحات والبحث العلمي والتجارب كمية كبيرة من البيانات التي يمكن أن تكون مفيدة للتحليل والرؤى .

هذه مجرد أمثلة قليلة، ومصادر البيانات الضخمة تتطور باستمرار مع تغير التكنولوجيا والمجتمع.

### تحديات تنظيم البيانات الضخمة

هناك العديد من المشكلات والتحديات التي قد تواجهها المؤسسات عند التعامل مع البيانات الضخمة، بما في ذلك :

1. جودة البيانات ودقتها: قد يكون ضمان دقة البيانات واتساقها وخلوها من الأخطاء أمرًا صعبًا عند التعامل مع كميات كبيرة من البيانات من مصادر مختلفة .

2. التخزين والمعالجة: يمكن أن يكون تخزين مجموعات البيانات الضخمة ومعالجتها كثيفًا للموارد ويتطلب تقنية وبنية تحتية متخصصة .
  3. خصوصية البيانات وأمنها: حماية البيانات الحساسة من الوصول غير المصرح به والخروقات أمر بالغ الأهمية، ويمكن أن يمثل تحديًا عند التعامل مع البيانات الضخمة .
  4. إدارة البيانات: قد يكون وضع السياسات والإجراءات والمعايير الخاصة بإدارة البيانات الضخمة والحفاظ عليها أمرًا صعبًا .
  5. تكامل البيانات: قد يكون دمج البيانات من مصادر متعددة أمرًا صعبًا، خاصة عند التعامل مع تنسيقات وهياكل بيانات مختلفة .
  6. قابلية التوسع: مع نمو البيانات، يجب أن يكون النظام قادرًا على التوسع أفقيًا وعموديًا للتعامل مع حجم البيانات المتزايد والتعقيد .
  7. تحليلات البيانات: يعد استخراج رؤى ذات مغزى من البيانات غير المنظمة وشبه المنظمة تحديًا كبيرًا .
  8. خط مسار البيانات وإدارة سير العمل: خط مسار البيانات ومهام سير العمل معقدة ويصعب إدارتها على نطاق واسع .
- تتطلب معالجة هذه المشكلات والتحديات مجموعة من الحلول التقنية، مثل المعالجة الموزعة والتعلم الآلي، والحلول التنظيمية، مثل اختيار أفضل ممارسات لإدارة البيانات.

## أدوات إدارة البيانات الضخمة

هناك العديد من الأدوات المتاحة لإدارة وتحليل البيانات الضخمة، بما في ذلك :

### Hadoop

هو إطار عمل مفتوح المصدر معد لأغراض تنفيذ عمليات التخزين الموزع ومعالجة مجموعات البيانات الضخمة . تم تطويره بواسطة Apache Software Foundation وهو مكتوب بلغة Java. جوهر Hadoop هو نظام الملفات الموزعة (HDFS) Hadoop ، وهو نظام ملفات موزع يسمح بتخزين مجموعات البيانات الضخمة عبر خوادم متعددة .

يتضمن Hadoop أيضًا معالج يسمى MapReduce ، وهو نموذج برمجة لمعالجة مجموعات البيانات الضخمة على التوازي. يستند MapReduce إلى فكرة تقسيم مجموعة كبيرة من البيانات إلى أجزاء أصغر، تسمى الكتل، ثم توزيع هذه الكتل عبر مجموعة من الأجهزة للمعالجة المتوازية. ثم يتم تجميع النتائج ودمجها لإنتاج الناتج النهائي. تم تصميم Hadoop للتوسيع أفقيًا، مما يعني أنه يمكنه إضافة المزيد من العقد إلى مجموعة للتعامل مع زيادة حجم البيانات وتعقيدها. كما أن لديه امكانية لتجاوز الخطأ، مما يعني أنه يمكن أن تستمر في العمل حتى إذا فشلت بعض العقد في الكتلة .

يحتوي Hadoop على العديد من مكونات النظام البيئي، مثل :  
الخلية: أداة تخزين بيانات وأداة استعلام شبيهة بـ SQL لبرنامج

Hadoop. Pig: منصة تدفق بيانات عالية المستوى لإنشاء برامج  
 Hbase: قاعدة بيانات NoSQL تعمل أعلى. HDFS.

## Apache Spark

هو نظام حوسبة موزع مفتوح المصدر تنفذ عمليات معالجة مجموعات البيانات الضخمة بسرعة فائقة. تم تطويره بواسطة Apache Software Foundation وهو مكتوب بلغة Scala ، وهي لغة برمجة تعمل على (Java Virtual Machine (JVM). تم تصميمه للعمل جنبًا إلى جنب مع Hadoop وتقنيات البيانات الضخمة الأخرى وغالبًا ما تستخدم كبديل لـ MapReduce Hadoop لمعالجة البيانات.

يمكن لـ Spark معالجة البيانات في الوقت الفعلي، مما يجعلها مناسبة تمامًا لحالات الاستخدام التي تتطلب معالجة بيانات منخفضة وعالية الإنتاجية. مثل Hadoop ، يمكن لـ Spark أيضًا التوسع أفقيًا ولديها تسامح داخلي مع الخطأ. يحتوي أيضًا على عدد من المكتبات المبنية فوقه، مثل Spark SQL لاستعلام SQL Streaming لمعالجة تدفق البيانات في الوقت الفعلي و MLlib للتعلم الآلي .

تتمثل إحدى المزايا الرئيسية لبرنامج Spark عبر MapReduce في إمكانات معالجة البيانات داخل الذاكرة. يقوم Spark بتخزين البيانات في الذاكرة بدلاً من قراءتها من القرص، مما قد يؤدي إلى تسريع معالجة البيانات بشكل كبير. يوفر Spark أيضًا نموذجًا أكثر

مرونة لمعالجة البيانات، مع دعم المعالجة الدفعية، و SQL التفاعلي، والتدفق، والتعلم الآلي .

يتم استخدام Spark من قبل العديد من المؤسسات لمجموعة واسعة من مهام معالجة البيانات الضخمة، مثل تخزين البيانات، وبحيرة البيانات، و التدفق، والتعلم الآلي. يستخدم Spark أيضًا على نطاق واسع في الأوساط الأكاديمية والبحثية لمعالجة البيانات والتحليلات .

باختصار، يعد Spark إطارًا قويًا لمعالجة البيانات الضخمة يمكنه معالجة مجموعات البيانات الضخمة بسرعة وهو مناسب تمامًا لحالات الاستخدام التي تتطلب معالجة بيانات ذات زمن انتقال منخفض وإنتاجية عالية. غالبًا ما يتم استخدامه كبديل ل MapReduce Hadoop ويتم اعتماده على نطاق واسع في الصناعة والأوساط الأكاديمية.

## NoSQL

اختصار [Not Only Sql]. وهي نوع من نظم إدارة قواعد البيانات المصممة للتعامل مع مجموعات البيانات الضخمة غير المنظمة. على عكس قواعد البيانات العلائقية التقليدية، التي تستخدم لغة الاستعلام المهيكلة (SQL) لتخزين البيانات واستردادها، تستخدم قواعد بيانات NoSQL مجموعة متنوعة من نماذج البيانات، مثل قيمة المفتاح والمستند والعمود والرسم البياني. تتضمن بعض الميزات الرئيسية لقواعد بيانات NoSQL ما يلي :

- مخطط أقل: لا تحتوي قواعد بيانات NoSQL على مخطط ثابت، مما يعني أنه يمكنها تخزين البيانات واستردادها بتنسيقات متنوعة، مثل JSON أو XML ، ولا تتطلب بنية محددة مسبقًا .
- القياس الأفقي: يمكن لقواعد بيانات NoSQL التوسع أفقيًا، مما يعني أنه يمكنها إضافة المزيد من الآلات إلى مجموعة للتعامل مع حجم البيانات المتزايد.
- التوافر العالي: قواعد بيانات NoSQL مصممة لتوفر عاليًا، مما يعني أنها يمكن أن تستمر في العمل حتى إذا فشلت بعض العقد في الكتلة.
- أداء عالٍ: تم تحسين قواعد بيانات NoSQL للحصول على أداء عالٍ، لا سيما لأعباء العمل المليئة بالقراءة .
- موجه نحو المستندات: بعض قواعد بيانات NoSQL مثل MongoDB و Couchbase و Cosmos DB موجهة نحو المستندات وتخزن البيانات بتنسيق شبه منظم، والتي يمكن الاستعلام عنها وفهرستها بسهولة باستخدام مجموعة متنوعة من نماذج البيانات .
- مخازن القيمة الرئيسية: بعض قواعد بيانات NoSQL مثل Redis و Riak عبارة عن مخازن Key-Value ، والتي تخزن البيانات بتنسيق القيمة الرئيسية ويتم تحسينها لتحقيق أداء عالٍ وزمن انتقال منخفض .

- المخازن العمودية: بعض قواعد بيانات NoSQL مثل Cassandra وHbase عبارة عن مخازن عمودية، والتي تخزن البيانات بتنسيق قائم على الأعمدة ويتم تحسينها لإنتاجية كتابة عالية.
- مخازن الرسم البياني: بعض قواعد بيانات NoSQL مثل Amazon Neptune وNeo4j هي مخازن رسوم بيانية، والتي تخزن البيانات في تنسيق قائم على الرسم البياني وتم تحسينها للعبور والاستعلام عن مجموعات بيانات الرسم البياني الكبيرة.

أصبحت قواعد بيانات NoSQL شائعة بشكل متزايد في السنوات الأخيرة حيث تسعى المؤسسات إلى تخزين ومعالجة كميات كبيرة من البيانات غير المهيكلة، مثل بيانات الوسائط الاجتماعية وبيانات أجهزة الاستشعار وملفات السجل. يتم استخدامها على نطاق واسع في العديد من الصناعات، بما في ذلك التجارة الإلكترونية والخدمات المالية والرعاية الصحية الخ....

### Amazon Redshift

هي خدمة لتخزين البيانات تقدمها Amazon Web Services (AWS). يسمح للمؤسسات بتخزين وتحليل كميات كبيرة من البيانات باستخدام مستودع بيانات بحجم بيتابايت. يعتمد Redshift على قاعدة بيانات PostgreSQL مفتوحة المصدر ويستخدم نهج التخزين العمودي، والذي تم تصميمه لتحسين أداء الاستعلام على مجموعات البيانات الضخمة. كما أنه يستخدم خوارزميات ضغط

متقدمة لتقليل تكاليف التخزين. تتمثل إحدى الميزات الرئيسية لـ Redshift في قدرتها على التوسيع والتصغير حسب الحاجة، مما يسمح للمؤسسات بإضافة السعة أو إزالتها بسهولة حيث يحتاج تخزين البيانات إلى التغيير. يوفر Redshift أيضًا عددًا من ميزات الأمان، بما في ذلك عزل الشبكة، والتشفير أثناء الراحة، والتكامل مع AWS Identity and Access Management (IAM). يسمح Redshift أيضًا بالتكامل مع خدمات AWS الأخرى، مثل Amazon S3 و Amazon EMR و Amazon Kinesis، مما يمكن المؤسسات من نقل البيانات ومعالجتها بسهولة داخل النظام الإيكولوجي لتخزين البيانات. فضلًا عن ذلك، فإنه يوفر عددًا من أدوات الإدارة والمراقبة، بما في ذلك Amazon Redshift Management Console و AWS SDKs و AWS CLI. يتم استخدام Redshift من قبل العديد من المؤسسات لمجموعة واسعة من تخزين البيانات ومهام التحليلات، مثل بحيرة البيانات وتخزين البيانات وذكاء الأعمال والتحليلات الخ... يسمح Redshift للمؤسسات بتخزين كميات كبيرة من البيانات وتحليلها بسهولة، ويوفر حلاً فعالاً من حيث التكلفة لتخزين البيانات على نطاق واسع.

### Google BigQuery

هي خدمة تخزين بيانات مُدارة بالكامل وقائمة على السحابة مقدمة من Google Cloud Platform (GCP). يسمح للمؤسسات بتخزين وتحليل كميات كبيرة من البيانات باستخدام مستودع بيانات بحجم بيتابايت. يعتمد BigQuery على ورقة Dremel من

Google، والتي تصف نظامًا لتشغيل استعلامات مخصصة على مجموعات بيانات كبيرة وموزعة. يستخدم نهج التخزين العمودي، والذي تم تصميمه لتحسين أداء الاستعلام على مجموعات البيانات الضخمة ويمكنه التعامل مع البيانات المنظمة وشبه المنظمة. تتمثل إحدى الميزات الرئيسية لـ BigQuery في قدرتها على التوسع والخفض حسب الحاجة، مما يسمح للمؤسسات بإضافة السعة أو إزالتها بسهولة حيث يحتاج تخزين البيانات إلى التغيير. يوفر BigQuery أيضًا عددًا من ميزات الأمان، بما في ذلك عزل الشبكة والتشفير أثناء عدم التشغيل والتكامل مع Google Cloud Identity and Access Management (IAM). يسمح BigQuery أيضًا بالتكامل مع خدمات GCP الأخرى، مثل Google Cloud Storage و Google Cloud Dataflow و Google Cloud Dataproc، مما يمكن المؤسسات من نقل البيانات ومعالجتها بسهولة داخل النظام الإيكولوجي لتخزين البيانات. فضلًا عن ذلك، فإنه يوفر عددًا من أدوات الإدارة والمراقبة، بما في ذلك BigQuery Web UI و BigQuery API وأداة سطر الأوامر.

يتم استخدام BigQuery بواسطة العديد من المؤسسات لمجموعة واسعة من تخزين البيانات ومهام التحليلات، مثل بحيرة البيانات وتخزين البيانات وذكاء الأعمال والتحليلات. يسمح للمؤسسات بتخزين كميات كبيرة من البيانات وتحليلها بسهولة ويقدم حلاً فعالاً من حيث التكلفة لتخزين البيانات على نطاق واسع.

## Google TensorFlow

هي مكتبة برامج مفتوحة المصدر للتعلم الآلي تم تطويرها بواسطة Google. يوفر نظامًا أساسيًا مرناً وعالي الأداء لتطوير نماذج التعلم الآلي والتدريب عليها ونشرها. تم تطوير TensorFlow في الأصل من قبل باحثين ومهندسين يعملون في فريق Google Brain ضمن منظمة بحث Google Machine Intelligence لأغراض إجراء التعلم الآلي وأبحاث الشبكات العصبية العميقة، ولكن النظام عام بما يكفي ليكون قابلاً للتطبيق في مجموعة متنوعة من المجالات الأخرى مثل نحن سوف. جوهر TensorFlow هو رسم بياني حسابي، رسم بياني دوري موجه (DAG) يمكن استخدامه لتمثيل أي حساب رياضي، حيث تمثل العقد عمليات رياضية وتمثل الجواف البيانات المتدفقة بينها. يسمح ذلك بحساب التدرجات اللونية بكفاءة باستخدام backpropagation، وهو أساس معظم خوارزميات التعلم الآلي. يوفر TensorFlow أيضًا مجموعة واسعة من الأدوات والمكتبات لتسهيل تطوير ونشر نماذج التعلم الآلي. وتشمل هذه:

TensorFlow Core: المركز المرن منخفض المستوى للمكتبة،

والذي يسمح ببناء وتنفيذ الرسوم البيانية الحسابية. طبقات

TensorFlow: مكتبة عالية المستوى لبناء الشبكات العصبية،

والتي توفر واجهة مبسطة لإنشاء النماذج والتدريب عليها.

TensorFlow Estimators مكتبة ذات مستوى أعلى لإنشاء نماذج التعلم الآلي ونشرها، والتي توفر نماذج وأدوات مُعدة مسبقًا للتدريب والتقييم والخدمة. خدمة

TensorFlow نظام خدمة مرّن وعالي الأداء لنشر نماذج التعلم الآلي في الإنتاج.

TensorFlow Hub مكتبة لاكتشاف واستخدام نماذج التعلم الآلي المدربة مسبقًا. يدعم TensorFlow مجموعة متنوعة من لغات البرمجة، بما في ذلك Python و C++ و Java وهو متاح على مجموعة متنوعة من الأنظمة الأساسية، بما في ذلك Windows و Linux و MacOS. بالإضافة إلى الأجهزة المحمولة ومتصفحات الويب. يستخدم TensorFlow على نطاق واسع في كل من الصناعة والأوساط الأكاديمية لمجموعة واسعة من مهام التعلم الآلي، بما في ذلك التعرف على الصور والكلام ومعالجة اللغة الطبيعية والتحليلات التنبؤية.

### Scikit-Learn

المعروفة أيضًا باسم sklearn، هي مكتبة Python مفتوحة المصدر للتعلم الآلي. إنه مبني على قمة مكتبات Python الشهيرة الأخرى مثل NumPy و SciPy وهو مصمم ليكون سهل الاستخدام وفعال لمجموعة واسعة من مهام التعلم الآلي. توفر Scikit-Learn مجموعة واسعة من الأدوات للتعلم الخاضع للإشراف وغير الخاضع للإشراف، بما في ذلك:

- التصنيف: خوارزميات لمشاكل التصنيف، مثل الانحدار اللوجستي، وأشجار القرار، وآلات المتجهات الداعمة.
- الانحدار: خوارزميات لمشاكل الانحدار، مثل الانحدار الخطي ودعم الانحدار المتجه.
- التجميع: خوارزميات للتجميع، مثل الوسائل k والتكتل الهرمي.
- تقليل الأبعاد: خوارزميات لتقليل عدد الميزات في مجموعة البيانات، مثل PCA وLLE.
- اختيار النموذج: أدوات لاختيار أفضل نموذج لمهمة معينة، مثل التحقق المتبادل والبحث في الشبكة.
- المعالجة المسبقة: أدوات للمعالجة المسبقة للبيانات، مثل القياس، والتضمين، واستخراج الميزات.

يوفر Scikit-Learn أيضًا عددًا من الميزات التي تجعله سهل الاستخدام، مثل واجهة برمجة تطبيقات متسقة لجميع الطرز، وميزة مسارنقل مدمجة تسمح بتسلسل خطوات معالجة متعددة مع نموذج.

### Apache Airflow

هو نظام أساسي مفتوح المصدر لتأليف وجدولة ومراقبة سير العمل برمجيًا. تم تطوير Airflow بواسطة Airbnb في عام 2014 وفتح مصدره في عام 2015. وهو مكتوب بلغة Python ويستخدم رسمًا بيانيًا لا دوريًا موجهًا (DAG) لتمثيل سير العمل، والذي يمكن

تصوره في واجهة ويب. Airflow يسمح تدفق الهواء بجدولة ومراقبة وإدارة تدفقات العمل وخطوط نقل البيانات، بما في ذلك :

- ديناميكي: يتم تعريف Airflow DAGs في كود Python ، مما يجعل من السهل إجراء التغييرات والتحكم في سير عملك .
- قابل للتوسيع: تم تصميم Airflow باستخدام بنية البرنامج المساعد، والتي تسمح بسهولة إضافة مشغلين جدد (على سبيل المثال، لتشغيل وظيفة Spark أو Pig script).
- قابلة للتطوير: تم تصميم Airflow للتعامل مع عدد كبير من المهام ويمكن تشغيله على مجموعة من الآلات .
- مرن: يمكن أن يعمل تدفق الهواء مع مجموعة متنوعة من محركات التنفيذ، بما في ذلك التنفيذ المحلي، والكرفس، و Mesos.
- المراقبة: يوفر Airflow واجهة مستخدم قائمة على الويب لمراقبة حالة سير العمل وتتبع حالات فشل المهام وعرض سجلات المهام.
- المشغلون: يوفر Airflow مجموعة واسعة من المشغلين المصممين مسبقًا للمهام الشائعة مثل تشغيل وظيفة Spark، ونقل البيانات من نظام إلى آخر وتشغيل استعلام SQL. يستخدم تدفق الهواء بشكل شائع لتدفقات عمل ETL (الاستخراج والتحويل والتحميل) وخطوط نقل البيانات، ولكن يمكن استخدامه أيضًا لأنواع أخرى من مهام سير العمل مثل سير عمل التعلم الآلي. يتم اعتماده على نطاق

واسع في الصناعة، لا سيما في النظام البيئي للبيانات الضخمة، ويستخدم أيضًا في الأوساط الأكاديمية والبحثية. مجتمع Airflow نشط ويوفر الدعم والبرامج التعليمية وأفضل الممارسات. إنه يتكامل مع أدوات البيانات الضخمة الأخرى مثل Hive و Pig و Presto و BigQuery مما يجعله أداة قوية لمهندسي البيانات وعلماء البيانات لإدارة مسار البيانات الخاص بهم.

### Apache Nifi

هي أداة لتكامل وإدارة البيانات مفتوحة المصدر تم تطويرها بواسطة Apache Software Foundation. إنه مصمم لأتمتة تدفق البيانات بين الأنظمة، مما يسهل نقل البيانات ومعالجتها وإدارتها. توفر NiFi واجهة قائمة على الويب لإنشاء تدفقات البيانات وإدارتها ومراقبتها.

يستخدم محرر تدفق البيانات المرئي لإنشاء خطوط نقل البيانات، وهو يدعم مجموعة واسعة من المعالجات لأداء مهام تكامل البيانات وإدارتها المختلفة مثل:

- استيعاب البيانات: يمكن ل NiFi جمع البيانات من مجموعة متنوعة من المصادر مثل الملفات وقواعد البيانات وخدمات الويب.

- معالجة البيانات: توفر NiFi مجموعة واسعة من المعالجات لأداء مهام معالجة البيانات المختلفة مثل تصفية البيانات وتحويلها وإثرائها.
- توجيه البيانات: يمكن لـ NiFi توجيه البيانات إلى وجهات مختلفة بناءً على القواعد والشروط.
- إدارة البيانات: توفر NiFi ميزات لحوكمة البيانات مثل نسب البيانات ومصدر البيانات وجودة البيانات.
- أمان البيانات: توفر NiFi مجموعة واسعة من ميزات الأمان مثل تشفير البيانات والتحكم في الوصول وتسجيل التدقيق. تم تصميم NiFi ليكون قابلاً للتطوير بدرجة عالية ويتحمل الأخطاء، ويمكنه التعامل مع كميات كبيرة من البيانات بزمن انتقال منخفض. يتم استخدامه على نطاق واسع في الصناعة لمجموعة واسعة من مهام تكامل البيانات وإدارتها، مثل بحيرة البيانات، وتخزين البيانات، و ETL، و IoT. كما أنها تستخدم في الأوساط الأكاديمية والبحثية لتكامل البيانات ومهام الإدارة. مجتمع NiFi نشط ويوفر الدعم والبرامج التعليمية وأفضل الممارسات.

### Apache Kafka

هو نظام مراسلة مفتوح المصدر وموزع ومتسامح مع الأخطاء. تم تطويره بواسطة LinkedIn وتم التبرع به لاحقًا لمؤسسة Apache Software Foundation. إنه مصمم للتعامل مع كميات كبيرة من

البيانات بزمن انتقال منخفض ولتوفير تدفق البيانات في الوقت الفعلي .

يعتمد على نموذج الاشتراك في النشر، حيث يرسل المنتجون رسائل إلى الموضوعات، ويشترك المستهلكون في الموضوعات لتلقي الرسائل. تتضمن بعض الميزات الرئيسية للنظام ما يلي :

- قابلية التوسع: يمكن أن يتعامل النظام مع كميات كبيرة من البيانات من خلال التوسع الأفقي للتعامل مع المزيد من حركة المرور.
- المتانة: يخزن النظام جميع الرسائل المنشورة لفترة زمنية قابلة للتهيئة، مما يسمح بالتسامح مع الخطأ في حالة الفشل .
- الوقت الفعلي: يسمح النظام بتدفق البيانات في الوقت الفعلي، مما يجعله مناسبًا تمامًا لحالات الاستخدام التي تتطلب معالجة بيانات بزمن انتقال منخفض .
- مصادر بيانات متعددة: يمكن أن يتعامل مع البيانات من مجموعة متنوعة من المصادر، بما في ذلك قواعد البيانات وملفات السجل وأجهزة إنترنت الأشياء .
- تعدد المستهلكين: يتيح النظام للعديد من المستهلكين القراءة من نفس الموضوع، مما يسهل تنفيذ خطوط بيانات متعددة .

- معالجة التدفق: يوفر النظام Streams API، والذي يسمح بتطوير تطبيقات معالجة الدفع التي يمكن أن تؤدي معالجة البيانات في الوقت الحقيقي على تدفقات البيانات الواردة.

تجدر الإشارة الى اهمية معرفة هذه الأدوات لأنها تُستخدم على نطاق واسع في الصناعة والأوساط الأكاديمية لمجموعة واسعة من مهام ذات الارتباط بالبيانات الضخمة. إنها توفر القدرة على تخزين ومعالجة وتحليل كميات كبيرة من البيانات، والتي أصبحت ذات أهمية متزايدة في عالم اليوم القائم على البيانات. من خلال فهم كيفية عمل هذه الأدوات وما هي قدرة عليه، يمكنك اتخاذ قرارات مستنيرة حول الأدوات التي يجب استخدامها في مهام محددة وكيفية استخدامها بفعالية. على سبيل المثال، يتم استخدام Hadoop و Spark على نطاق واسع لمعالجة البيانات الضخمة والتحليلات، ويمكن أن يساعدك فهم قدراتها وقيودها في تحديد أيهما أكثر ملاءمة لمهمة معينة. تُستخدم قواعد بيانات NoSQL على نطاق واسع لتخزين واسترداد مجموعات البيانات الضخمة غير المهيكلة، ويمكن أن يساعدك فهم ميزاتها وقيودها في تحديد أيها أكثر ملاءمة لحالة استخدام معينة.

تتيح لك معرفة هذه الأدوات أيضًا التواصل بشكل أفضل مع المهنيين الآخرين في هذا المجال، وفهم تقاريرهم والسياق وراءها، والمساهمة في عملية صنع القرار للفريق، خاصة عندما يتعلق الأمر باختيار الأداة المناسبة للوظيفة، وتقييم مزايا وعيوب كل منها. باختصار، يمكن أن يساعدك امتلاك فهم جيد لأدوات البيانات

الضخمة على إدارة كميات كبيرة من البيانات وتحليلها بشكل فعال، واتخاذ قرارات مستنيرة حول الأدوات التي يجب استخدامها في مهام محددة، والتواصل بشكل فعال مع محترفين آخرين في هذا المجال.

### تطبيقات البيانات الضخمة

تطبيقات البيانات الضخمة هي برامج وأنظمة تستخدم تقنيات البيانات الضخمة لحل المشكلات المعقدة واتخاذ قرارات أفضل. يمكن استخدامها في مجموعة واسعة من الصناعات والمجالات، بما في ذلك:

#### الرعاية الصحية.

في مجال الرعاية الصحية، يمكن استخدام تطبيقات البيانات الضخمة لتحسين نتائج المرضى وتقليل التكاليف وتسريع اكتشاف علاجات جديدة. تتضمن بعض الأمثلة المحددة لكيفية استخدام البيانات الضخمة في الرعاية الصحية ما يلي:

- السجلات الصحية الإلكترونية: السجلات الصحية الإلكترونية هي إصدارات رقمية من السجلات الورقية التي يستخدمها مقدمو الرعاية الصحية لتخزين معلومات المريض. باستخدام تقنيات البيانات الضخمة، يمكن استخدام السجلات الصحية الإلكترونية لتحليل كميات كبيرة من بيانات المرضى لتحديد الأنماط والاتجاهات، والتي يمكن أن تساعد مقدمي الرعاية الصحية على اتخاذ قرارات أفضل بشأن رعاية المرضى.

- التحليلات التنبؤية: يمكن استخدام التحليلات التنبؤية لتحليل كميات كبيرة من بيانات المريض لتحديد المرضى المعرضين لخطر الإصابة بحالات معينة، مثل مرض السكري أو أمراض القلب. يمكن أن يساعد هذا مقدمي الرعاية الصحية على التدخل المبكر ومنع تطور حالات أكثر خطورة.
- التجارب السريرية: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات من التجارب السريرية لتحديد العلاجات الجديدة وتحسين تصميم التجارب المستقبلية.
- التصوير الطبي: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من بيانات التصوير الطبي، مثل الأشعة السينية والتصوير بالرنين المغناطيسي، لتحديد الأنماط والاتجاهات التي يمكن أن تساعد مقدمي الرعاية الصحية في إجراء تشخيصات أكثر دقة.
- علم الجينات: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات الجينومية لتحديد العلامات الجينية المرتبطة بأمراض معينة ولتطوير خطط علاجية مخصصة.
- صحة السكان: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات من السكان لتحديد الأنماط والاتجاهات في النتائج الصحية، والتي يمكن أن تساعد

مقدمي الرعاية الصحية على تصميم وتنفيذ تدخلات الصحة العامة الأكثر فعالية.

بشكل عام، يمكن أن تساعد البيانات الضخمة في الرعاية الصحية في تحسين نتائج المرضى، وخفض التكاليف، وتسريع اكتشاف علاجات جديدة. يسمح لمقدمي الرعاية الصحية بتحليل كميات كبيرة من البيانات من مصادر متعددة، بما في ذلك السجلات الصحية الإلكترونية، والتجارب السريرية، والتصوير الطبي، وعلم الجينوم، لتحديد الأنماط والاتجاهات التي يمكن أن تساعد في اتخاذ القرارات السريرية وتحسين رعاية المرضى.

### التمويل

في مجال التمويل، يمكن استخدام تطبيقات البيانات الضخمة لتحسين إدارة المخاطر واكتشاف الاحتيال وتحسين قرارات الاستثمار. تتضمن بعض الأمثلة المحددة لكيفية استخدام البيانات الضخمة في التمويل ما يلي:

- إدارة المخاطر: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات المالية لتحديد الأنماط والاتجاهات التي يمكن أن توجه قرارات إدارة المخاطر. على سبيل المثال، من خلال تحليل كميات كبيرة من البيانات من الأسواق المالية، يمكن للمؤسسات المالية تحديد الأنماط التي تشير إلى زيادة المخاطر واتخاذ خطوات للتخفيف من تلك المخاطر.

- كشف الاحتيال: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات المالية لتحديد الأنماط والاتجاهات التي يمكن أن تشير إلى نشاط احتيالي. على سبيل المثال، من خلال تحليل كميات كبيرة من البيانات من معاملات بطاقات الائتمان، يمكن للمؤسسات المالية تحديد الأنماط التي تشير إلى نشاط احتيالي واتخاذ خطوات لمنعها.
- إدارة الاستثمار: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات المالية لتحديد الأنماط والاتجاهات التي يمكن أن توجه قرارات الاستثمار. على سبيل المثال، من خلال تحليل كميات كبيرة من البيانات من الأسواق المالية، يمكن لمديري الاستثمار تحديد الأنماط التي تشير إلى زيادة المخاطر أو زيادة العوائد المحتملة واتخاذ قرارات استثمارية أكثر استنارة.
- التداول الخوارزمي: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات المالية في الوقت الفعلي، مما يسمح للمؤسسات المالية باتخاذ قرارات استثمارية أسرع وأكثر دقة باستخدام الخوارزميات.
- تقسيم العملاء: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من بيانات العملاء لتحديد الأنماط والاتجاهات التي يمكن أن توجه قرارات تقسيم العملاء. على سبيل المثال، من خلال تحليل كميات كبيرة من البيانات من معاملات العملاء، يمكن للمؤسسات المالية تحديد الأنماط التي تشير

إلى زيادة المخاطر أو زيادة العوائد المحتملة واتخاذ قرارات أكثر استنارة بشأن العملاء الذين يجب استهدافهم لمنتجات أو خدمات معينة .

بشكل عام، يمكن أن تساعد البيانات الضخمة في التمويل المؤسسات المالية على تحسين إدارة المخاطر واكتشاف الاحتيال وتحسين قرارات الاستثمار. يسمح لهم بتحليل كميات كبيرة من البيانات من مصادر متعددة، مثل الأسواق المالية والمعاملات وبيانات العملاء، لتحديد الأنماط والاتجاهات التي يمكن أن تساعد في صنع القرار وتحسين الأداء

### البيع بالتجزئة

في البيع بالتجزئة، يمكن استخدام تطبيقات البيانات الضخمة لتحسين خدمة العملاء وتخصيص التسويق وتحسين إدارة سلسلة التوريد. تتضمن بعض الأمثلة المحددة لكيفية استخدام البيانات الضخمة في البيع بالتجزئة ما يلي:

- تحليلات العملاء: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من بيانات العملاء، مثل سجل الشراء، وسجل التصفح، والعوامل الديموغرافية، لتحديد الأنماط والاتجاهات التي يمكن أن تفيد قرارات خدمة العملاء والتسويق. على سبيل المثال، من خلال تحليل بيانات العملاء، يمكن لتجار التجزئة تحديد الأنماط التي تشير إلى زيادة المخاطر أو زيادة العوائد المحتملة واتخاذ قرارات أكثر

استنارة حول العملاء الذين يجب استهدافهم لمنتجات أو خدمات معينة .

- التخصيص: يمكن استخدام البيانات الضخمة لتخصيص تجربة التسوق للعملاء. على سبيل المثال، يمكن لبائعي التجزئة استخدام بيانات العملاء للتوصية بالمنتجات التي من المرجح أن تكون ذات فائدة لعميل معين.
- إدارة المخزون: يمكن استخدام البيانات الضخمة لتحسين إدارة المخزون من خلال تحليل كميات كبيرة من البيانات من المبيعات وبيانات العملاء وبيانات سلسلة التوريد. يمكن لبائعي التجزئة استخدام هذه البيانات لتحديد الأنماط والاتجاهات التي يمكن أن تحدد القرارات المتعلقة بالمنتجات التي يجب تخزينها ومتى تطلب المزيد .
- إدارة سلسلة التوريد: يمكن استخدام البيانات الضخمة لتحسين إدارة سلسلة التوريد من خلال تحليل كميات كبيرة من البيانات من الموردين واللوجستيات والإنتاج. يمكن لبائعي التجزئة استخدام هذه البيانات لتحديد الأنماط والاتجاهات التي يمكنها إبلاغ القرارات حول كيفية تحسين سلسلة التوريد وتقليل التكاليف .
- الصيانة التنبؤية: يمكن استخدام البيانات الضخمة للتنبؤ بالوقت الذي يحتمل فيه فشل المعدات أو المرافق. على سبيل المثال، يمكن لبائعي التجزئة استخدام بيانات المستشعر من

المعدات للتنبؤ بوقت الحاجة إلى الصيانة، مما قد يساعد في تقليل وقت التوقف عن العمل وتحسين الكفاءة .

بشكل عام، يمكن أن تساعد البيانات الضخمة في البيع بالتجزئة تجار التجزئة على تحسين خدمة العملاء وتخصيص التسويق وتحسين إدارة سلسلة التوريد. يسمح لهم بتحليل كميات كبيرة من البيانات من مصادر متعددة، مثل بيانات المبيعات وبيانات العملاء وبيانات سلسلة التوريد، لتحديد الأنماط والاتجاهات التي يمكن أن تساعد في صنع القرار وتحسين الأداء.

### النقل

في مجال النقل، يمكن استخدام تطبيقات البيانات الضخمة لتحسين تدفق حركة المرور وتقليل الحوادث وتحسين الخدمات اللوجستية. تتضمن بعض الأمثلة المحددة لكيفية استخدام البيانات الضخمة في النقل ما يلي:

- إدارة حركة المرور: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من بيانات حركة المرور، مثل بيانات GPS من المركبات، لتحديد الأنماط والاتجاهات التي يمكن أن توجه قرارات إدارة حركة المرور. على سبيل المثال، من خلال تحليل بيانات حركة المرور، يمكن لوكالات النقل تحديد الأنماط التي تشير إلى زيادة الازدحام المروري واتخاذ خطوات للتخفيف من حدته.

- أنظمة النقل الذكية: يمكن استخدام البيانات الضخمة لتطوير أنظمة نقل ذكية، مثل السيارات المتصلة، والتي يمكنها تحسين تدفق حركة المرور من خلال مشاركة المعلومات حول ظروف حركة المرور في الوقت الفعلي.
- الصيانة التنبؤية: يمكن استخدام البيانات الضخمة للتنبؤ بالوقت الذي يحتمل فيه فشل المعدات أو المرافق. على سبيل المثال، يمكن لوكالات النقل استخدام بيانات الاستشعار من المركبات والبنية التحتية للتنبؤ بوقت الحاجة إلى الصيانة، مما قد يساعد في تقليل وقت التوقف عن العمل وتحسين الكفاءة.
- النقل العام: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات من أنظمة النقل العام، مثل جداول الحافلات والقطارات، لتحديد الأنماط والاتجاهات التي يمكن أن توجه القرارات حول كيفية تحسين النظام وتقليل التكاليف.
- الخدمات اللوجستية: يمكن استخدام البيانات الضخمة لتحسين الخدمات اللوجستية عن طريق تحليل كميات كبيرة من البيانات من شركات النقل والخدمات اللوجستية، مثل بيانات الشحن والتتبع، لتحديد الأنماط والاتجاهات التي يمكن أن توجه القرارات حول كيفية تحسين سلسلة التوريد وتقليل التكاليف.

- المركبات الذاتية: يمكن استخدام البيانات الضخمة لتدريب وتحسين أداء المركبات ذاتية القيادة من خلال تحليل كميات كبيرة من البيانات من أجهزة الاستشعار والكاميرات ومصادر أخرى.

بشكل عام، يمكن أن تساعد البيانات الضخمة في مجال النقل وكالات وشركات النقل على تحسين تدفق حركة المرور وتقليل الحوادث وتحسين الخدمات اللوجستية. يسمح لهم بتحليل كميات كبيرة من البيانات من مصادر متعددة، مثل بيانات GPS ، وبيانات الاستشعار، وبيانات الخدمات اللوجستية،

### الطاقة

في صناعة الطاقة، يمكن استخدام تطبيقات البيانات الضخمة لتحسين الموثوقية وخفض التكاليف وتحسين تكامل مصادر الطاقة المتجددة. تتضمن بعض الأمثلة المحددة لكيفية استخدام البيانات الضخمة في صناعة الطاقة ما يلي:

- الشبكة الذكية: يمكن استخدام البيانات الضخمة لتطوير الشبكات الذكية وتحسينها، والتي يمكنها تحسين موثوقية شبكة الطاقة من خلال توفير معلومات في الوقت الفعلي حول استخدام الطاقة وظروف الشبكة.
- الصيانة التنبؤية: يمكن استخدام البيانات الضخمة للتنبؤ بالوقت الذي يحتمل فيه فشل المعدات أو المرافق. على سبيل المثال، يمكن لشركات الطاقة استخدام بيانات أجهزة

الاستشعار من محطات الطاقة وشبكات التوزيع للتنبؤ بوقت الحاجة إلى الصيانة، الأمر الذي يمكن أن يساعد في تقليل وقت التوقف عن العمل وتحسين الكفاءة.

- إدارة الطاقة: يمكن استخدام البيانات الضخمة لتحسين إدارة الطاقة من خلال تحليل كميات كبيرة من البيانات من المباني والمرافق الأخرى المستهلكة للطاقة لتحديد الأنماط والاتجاهات التي يمكن أن توجه القرارات حول كيفية تقليل استهلاك الطاقة والتكاليف.

- تكامل الطاقة المتجددة: يمكن استخدام البيانات الضخمة لتحسين تكامل مصادر الطاقة المتجددة، مثل طاقة الرياح والطاقة الشمسية، في شبكة الطاقة من خلال تحليل كميات كبيرة من البيانات من هذه المصادر لتحديد الأنماط والاتجاهات التي يمكن أن تثرى القرارات حول كيفية تحسين عملياتها وتقليل التكاليف.

- إدارة انبعاثات الكربون: يمكن استخدام البيانات الضخمة لمراقبة انبعاثات الكربون وإدارتها من خلال تحليل كميات كبيرة من البيانات من محطات الطاقة والمرافق الأخرى المستهلكة للطاقة لتحديد الأنماط والاتجاهات التي يمكن أن توجه القرارات حول كيفية تقليل الانبعاثات والامتثال للوائح.

- محطات شحن المركبات الكهربائية: يمكن استخدام البيانات الضخمة لتحسين عمليات وإدارة محطات شحن المركبات

الكهربائية من خلال تحليل كميات كبيرة من البيانات، مثل بيانات الشحن والتفريغ، لتحديد الأنماط والاتجاهات التي يمكن أن تحدد القرارات حول كيفية تحسين المحطات وخفض التكاليف .

بشكل عام، يمكن أن تساعد البيانات الضخمة في صناعة الطاقة شركات ووكالات الطاقة على تحسين الموثوقية وخفض التكاليف وتحسين تكامل الطاقة المتجددة

### التصنيع

في التصنيع، يمكن استخدام تطبيقات البيانات الضخمة لتحسين الكفاءة وتقليل وقت التوقف عن العمل وتحسين الصيانة. تتضمن بعض الأمثلة المحددة لكيفية استخدام البيانات الضخمة في التصنيع ما يلي:

- الصيانة التنبؤية: يمكن استخدام البيانات الضخمة للتنبؤ بالوقت الذي يحتمل فيه فشل المعدات أو المرافق. على سبيل المثال، يمكن للمصنعين استخدام بيانات أجهزة الاستشعار من المعدات الصناعية للتنبؤ بوقت الحاجة للصيانة، الأمر الذي يمكن أن يساعد في تقليل وقت التوقف عن العمل وتحسين الكفاءة .
- مراقبة الجودة: يمكن استخدام البيانات الضخمة لتحسين مراقبة الجودة من خلال تحليل كميات كبيرة من البيانات من عمليات الإنتاج لتحديد الأنماط والاتجاهات التي يمكن

أن توجه القرارات حول كيفية تحسين جودة المنتج وتقليل العيوب .

- إدارة سلسلة التوريد: يمكن استخدام البيانات الضخمة لتحسين إدارة سلسلة التوريد من خلال تحليل كميات كبيرة من البيانات من الموردين والخدمات اللوجستية لتحديد الأنماط والاتجاهات التي يمكن أن توجه القرارات حول كيفية تحسين سلسلة التوريد وتقليل التكاليف .

- التعلم الآلي: يمكن استخدام البيانات الضخمة لتدريب نماذج التعلم الآلي التي يمكنها تحسين أداء المعدات الصناعية وأتمتة عمليات صنع القرار في الصناعة التحويلية . النمذجة التنبؤية: يمكن استخدام البيانات الضخمة للنمذجة التنبؤية لتحسين تخطيط الإنتاج، من خلال تحديد الأنماط في البيانات التاريخية وتوقع الطلب المستقبلي، مما يسمح للمصنعين باتخاذ قرارات أفضل بشأن جداول المخزون والإنتاج .

- تحسين العمليات: يمكن استخدام البيانات الضخمة لتحسين عمليات التصنيع من خلال تحليل كميات كبيرة من البيانات من عمليات الإنتاج لتحديد الأنماط والاتجاهات التي يمكن أن توجه القرارات حول كيفية تحسين الكفاءة وخفض التكاليف .

بشكل عام، يمكن أن تساعد البيانات الضخمة في التصنيع الشركات المصنعة على تحسين الكفاءة وتقليل وقت التوقف عن

العمل وتحسين الصيانة. يسمح لهم بتحليل كميات كبيرة من البيانات من مصادر متعددة، مثل بيانات الاستشعار وبيانات الإنتاج وبيانات سلسلة التوريد، لتحديد الأنماط والاتجاهات التي يمكن أن تساعد في صنع القرار وتحسين الأداء.

### الاعلام

في مجال الاعلام، يمكن استخدام تطبيقات البيانات الضخمة لتحسين استهداف الإعلانات وتحسين توصيات المحتوى. تتضمن بعض الأمثلة المحددة لكيفية استخدام البيانات الضخمة في صناعة الإعلام ما يلي:

- تقسيم الجمهور: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات من المشاهدين والمستمعين والقراء لتحديد الأنماط والاتجاهات التي يمكن أن تحدد قرارات الجمهور التي يجب استهدافها لأنواع معينة من المحتوى أو الإعلان.
- الإعلان: يمكن استخدام البيانات الضخمة لتحسين الإعلان عن طريق تحليل كميات كبيرة من البيانات من المشاهدين والمستمعين والقراء لتحديد الأنماط والاتجاهات التي يمكن أن تحدد القرارات بشأن أنواع الإعلانات التي يجب استخدامها والجمهور التي يجب استهدافها.
- تحسين المحتوى: يمكن استخدام البيانات الضخمة لتحسين المحتوى من خلال تحليل كميات كبيرة من البيانات من

المشاهدين والمستمعين والقراء لتحديد الأنماط والاتجاهات التي يمكن أن تحدد القرارات حول أنواع المحتوى التي يجب إنتاجها والجمهور التي يجب استهدافها .

- وسائل التواصل الاجتماعي: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات من منصات وسائل التواصل الاجتماعي لتحديد الأنماط والاتجاهات التي يمكن أن تحدد القرارات المتعلقة بأنواع المحتوى التي يجب إنتاجها والجمهور التي يجب استهدافها.

- التخصيص: يمكن استخدام البيانات الضخمة لتخصيص تجربة الاعلام للمشاهدين والمستمعين والقراء من خلال تحليل كميات كبيرة من البيانات للتوصية بالمحتوى الذي من المرجح أن يثير اهتمامهم.

- تحليل المشاعر: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات من وسائل التواصل الاجتماعي والمصادر الأخرى لتحديد الأنماط والاتجاهات في الرأي العام والمشاعر حول موضوعات معينة، والتي يمكن أن تحدد القرارات حول أنواع المحتوى التي يجب إنتاجها والجمهور التي يجب استهدافها.

بشكل عام، يمكن للبيانات الضخمة في صناعة الإعلام أن تساعد الشركات الإعلامية على تحسين استهداف الإعلانات وتحسين توصيات المحتوى. يسمح لهم بتحليل كميات كبيرة من البيانات من مصادر متعددة، مثل بيانات المشاهد وبيانات الوسائط الاجتماعية وبيانات

الجمهور، لتحديد الأنماط والاتجاهات التي يمكن أن تساعد في صنع القرار وتحسين الأداء.

### الأمن

في مجال الأمن، يمكن استخدام تطبيقات البيانات الضخمة لتحديد ومنع التهديدات السيبرانية وتحسين الاستجابة للحوادث. تتضمن بعض الأمثلة المحددة لكيفية استخدام البيانات الضخمة في الأمان ما يلي:

- كشف التسلسل: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات من حركة مرور الشبكة وسجلات النظام والمصادر الأخرى لتحديد الأنماط والاتجاهات التي يمكن أن تشير إلى هجوم إلكتروني.
- مصادر التهديدات: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات من مصادر متعددة، مثل وسائل التواصل الاجتماعي، والويب المظلم، ومصادر أخرى، لتحديد الأنماط والاتجاهات في التهديدات السيبرانية ولتطوير ذكاء التهديدات الذي يمكن أن يسترشد بالقرارات الأمنية.
- معلومات الأمان وإدارة الأحداث (SIEM): يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات من الأجهزة الأمنية، مثل جدران الحماية وأنظمة كشف التسلسل وبرامج مكافحة الفيروسات، لتحديد الأنماط والاتجاهات

التي يمكن أن تشير إلى هجوم إلكتروني. تحليلات سلوك المستخدم: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات من المستخدمين، مثل محاولات تسجيل الدخول والوصول إلى الملفات والإجراءات الأخرى، لتحديد الأنماط والاتجاهات التي يمكن أن تشير إلى هجوم إلكتروني .

- إدارة الثغرات الأمنية: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات من عمليات مسح الثغرات الأمنية وتقييمات الأمان لتحديد الأنماط والاتجاهات التي يمكن أن تشير إلى ثغرة أمنية يجب معالجتها .
- الاستجابة للحوادث: يمكن استخدام البيانات الضخمة لتحليل كميات كبيرة من البيانات من إجراءات الاستجابة للحوادث، مثل بيانات الطب الشرعي من الأنظمة المخترقة، لتحديد الأنماط والاتجاهات التي يمكن أن توجه القرارات حول كيفية الرد على هجوم إلكتروني .

بشكل عام، يمكن أن تساعد البيانات الضخمة في الأمان المؤسسات على تحديد التهديدات السيبرانية ومنعها، وتحسين الاستجابة للحوادث. يسمح لهم بتحليل كميات كبيرة من البيانات من مصادر متعددة، مثل حركة مرور الشبكة، وسجلات النظام، وذكاء التهديدات، لتحديد الأنماط والاتجاهات التي يمكن أن توجه قرارات الأمان وتحسين الوضع الأمني العام للمؤسسة.

## اهمية دراسة البيانات الضخمة

تعد دراسة البيانات الضخمة مهمة لطلاب علوم المعلومات لعدد من الأسباب :

1. فرص العمل: مع تزايد الطلب على المهنيين الذين يمكنهم إدارة وتحليل البيانات الضخمة، يمكن أن تفتح دراسة البيانات الضخمة مجموعة واسعة من الفرص الوظيفية في مجالات مثل علم البيانات وذكاء الأعمال وتحليلات البيانات.
2. المهارات التقنية: من خلال دراسة البيانات الضخمة، سيكتسب الطلاب فهمًا قويًا للتقنيات والأدوات المستخدمة لإدارة وتحليل البيانات الضخمة، مثل قواعد بيانات Hadoop و Spark و NoSQL. هذه المهارات مطلوبة بشدة في سوق العمل.
3. صنع القرار المستند إلى البيانات: تؤدي البيانات الضخمة دورًا حاسمًا في الأعمال والمؤسسات الحديثة، ويمكن أن تمنح دراسة البيانات الضخمة الطلاب المهارات التي يحتاجون إليها لاتخاذ قرارات تعتمد على البيانات واستخراج الرؤى من مجموعات البيانات الضخمة، والتي يمكن تطبيقها في مختلف المجالات مثل الرعاية الصحية، والتمويل، والبيع بالتجزئة، وأكثر من ذلك.
4. نهج متعدد التخصصات: البيانات الضخمة هي مجال متعدد التخصصات يعتمد على العديد من المجالات المختلفة، مثل علوم الحاسوب والإحصاءات والأعمال و المعلومات. من

خلال دراسة البيانات الضخمة، سيكتسب الطلاب فهماً واسعاً لهذه المجالات وسيكونون قادرين على العمل بفعالية مع محترفين من خلفيات مختلفة .

5. القضايا الاجتماعية والأخلاقية: تجلب البيانات الضخمة العديد من القضايا الأخلاقية والاجتماعية التي تحتاج إلى معالجة، مثل الخصوصية والتحيز والشفافية. من خلال دراسة البيانات الضخمة، سيتعرض الطلاب لهذه المشكلات ويتعلمون كيفية معالجتها .

6. الابتكار: تمتلك البيانات الضخمة القدرة على دفع عجلة الابتكار في العديد من المجالات، ويمكن أن تمنح دراسة البيانات الضخمة الطلاب المهارات التي يحتاجونها لتحديد فرص عمل جديدة وتطوير منتجات وخدمات جديدة .

باختصار، تعد دراسة البيانات الضخمة مهمة لطلاب علوم المعلومات لأنها يمكن أن تزودهم بالمهارات التقنية والمعرفة والفهم التي يزداد الطلب عليها في سوق العمل، وتمنحهم القدرة على اتخاذ قرارات تعتمد على البيانات في مختلف المجالات و معالجة القضايا الأخلاقية والاجتماعية التي تنشأ من استخدام البيانات الضخمة.

### اسئلة الفصل الثالث

1. ما المقصود بالبيانات الضخمة وما هي اهم مصادرها؟
2. ما هي التحديات التي يمكن ان تواجه المؤسسات والمنظمات في تعاملها مع البيانات الضخمة؟
3. Hadoop . هو إطار عمل مفتوح المصدر معد لاغراض تنفيذ عمليات التخزين الموزع ومعالجة مجموعات البيانات الضخمة . تكلم عنه بالتفصيل.
4. الرعاية الصحية احدى مجالات تطبيق البيانات الضخمة. بين بالشرح مجالاتها ومصادرها ضمن هذا القطاع.
5. تعد دراسة البيانات الضخمة مهمة لطلاب علوم المعلومات لعدد من الأسباب. اذكرها بالتفصيل.
6. تكلم بنقاط عن الأمثلة لتطبيق البيانات الضخمة في مجال الطاقة.
7. استهداف الإعلانات وتحسين توصيات المحتوى. نموذج لتطبيق البيانات الضخمة في الاعلام. بين ذلك بالامثلة.
8. تكلم بنقاط عن مميزات نظام ادارة قواعد البيانات NoSql.
9. ما هي مجالات استخدام البيانات الضخمة في مجال الأمن.
10. تكلم بالتفصيل عن Google TensorFlow.

## الفصل الرابع

### إدارة البيانات وحوكمتها

#### تمهيد

إدارة البيانات هي عملية تنظيم وتخزين وحماية وصيانة سلامة بيانات المنظمة. إنها عنصر أساسي في عمليات أي مؤسسة، حيث تعد البيانات من الأصول القيمة التي يمكن أن تساعد في صنع القرار، وتحفيز نمو الأعمال، ودعم التخطيط الاستراتيجي.

تسمح الإدارة الفعالة للبيانات للمؤسسات بجمع كميات كبيرة من البيانات وتخزينها وتحليلها بطريقة فعالة وآمنة ومتوافقة مع المتطلبات القانونية والتنظيمية. تتضمن عمليات إدارة البيانات مهام مثل جمع البيانات وأمن البيانات وخصوصية البيانات وامتثال البيانات.

ومع زيادة كمية البيانات التي يتم إنشاؤها، أصبحت إدارة البيانات مهارة بالغة الأهمية للمؤسسات لتظل قادرة على المنافسة. متخصصو إدارة البيانات مسؤولون عن تصميم أنظمة إدارة البيانات وتنفيذها وصيانتها، والتأكد من دقة البيانات واكتمالها وأمانها.

سوف تستكشف في هذا الفصل المفاهيم وأفضل الممارسات لإدارة البيانات وكيف يمكن للمؤسسات الاستفادة من إدارة البيانات الفعالة. كما سيناقدش المهارات والمعرفة المطلوبة للمهنيين العاملين في مجال إدارة البيانات.

## إدارة قواعد البيانات

إدارة قواعد البيانات [DataBase Management] هي عملية تصميم قواعد البيانات وتنفيذها وصيانتها وتحسينها. قاعدة البيانات هي مجموعة منظمة من البيانات المخزنة في جهاز الحاسوب، وتسمح بتخزين واسترجاع وإدارة البيانات بكفاءة. هناك عدة أنواع من قواعد البيانات، بما في ذلك :

1. قواعد البيانات العلائقية: هي نوع من قواعد البيانات التي تنظم البيانات في جدول واحد أو أكثر مع صفوف وأعمدة ، حيث يحتوي كل جدول على مجموعة من البيانات ذات الصلة. ترتبط الجداول بعلاقات محددة بين البيانات في الجداول ، مثل علاقات رأس برأس أو رأس بأطراف أو علاقات أطراف بأطراف . تتضمن أمثلة قواعد البيانات العلائقية MySQL و PostgreSQL و Microsoft SQL Server. تُستخدم قواعد البيانات هذه بشكل شائع في تطبيقات الأعمال والمؤسسات لتخزين واسترداد كميات كبيرة من البيانات المنظمة. وتعد من أكثر أنواع قواعد البيانات شيوعًا وتستخدم لغة استعلام منظمة (SQL) لتخزين واسترداد البيانات .

2. قواعد بيانات NoSQL: تم تصميم قواعد البيانات هذه للتعامل مع كميات كبيرة من البيانات غير المهيكلة أو شبه المنظمة. إنهم لا يستخدمون SQL ، وغالبًا ما يتم استخدامها

للبيانات الضخمة وتطبيقات الويب في الوقت الفعلي.

تتضمن الأمثلة MongoDB و Cassandra و Hbase.

3. قواعد بيانات الرسم البياني: تم تصميم قواعد البيانات هذه

لتخزين والاستعلام عن البيانات المنظمة كرسم بياني، حيث

يتم تمثيل البيانات كعقد وحواف. إنها مناسبة تمامًا لحالات

الاستخدام مثل الشبكات الاجتماعية وأنظمة التوصية.

تتضمن الأمثلة Neo4j و JanusGraph.

4. قواعد بيانات السلاسل الزمنية: تم تصميم قواعد البيانات

هذه للتعامل مع بيانات السلاسل الزمنية، وهي عبارة عن

سلسلة من نقاط البيانات التي يتم جمعها على فترات زمنية

منتظمة. إنها مناسبة تمامًا لحالات الاستخدام مثل إنترنت

الأشياء والمراقبة والتحليلات. تتضمن الأمثلة InfluxDB و

OpenTSDB.

5. قواعد البيانات الذاكرة: تقوم قواعد البيانات هذه بتخزين

البيانات في ذاكرة الوصول العشوائي بدلاً من القرص. إنها

مناسبة تمامًا لحالات الاستخدام مثل التحليلات في الوقت

الفعلي والتداول عالي التردد والتجارة الإلكترونية. تتضمن

الأمثلة Redis و MemSQL و VoltDB.

تتطلب إدارة قواعد البيانات مجموعة من المهارات الفنية

والتحليلية، بما في ذلك الكفاءة في SQL ونمذجة البيانات وإدارة

قواعد البيانات. كما يتطلب أيضًا فهمًا لأمن البيانات وخصوصية

البيانات، ومعرفة باللوائح الخاصة بالصناعة ومعايير الامتثال.

تعد إدارة قواعد البيانات جزءًا أساسيًا من إدارة البيانات، حيث تسمح للمؤسسات بتخزين واسترداد وإدارة كميات كبيرة من البيانات بطريقة فعالة وأمنة. مع زيادة كمية البيانات التي يتم إنشاؤها، أصبحت إدارة قواعد البيانات مهارة بالغة الأهمية للمؤسسات لتظل قادرة على المنافسة.

### حوكمة البيانات

حوكمة البيانات [Data Governance] هي الإدارة العامة لتوافر البيانات في المؤسسة وإمكانية استخدامها وسلامتها وأمانها. وهي تشمل السياسات والمعايير والإجراءات والأدوار والمسؤوليات التي تحكم جمع البيانات وتخزينها واستخدامها وصيانتها واتلافها.

تعد حوكمة البيانات مهمة لأنها تساعد المؤسسات على ضمان دقة البيانات واكتمالها وأمانها، واستخدامها بشكل أخلاقي ومتوافق مع المتطلبات القانونية والتنظيمية. كما أنه يساعد على ضمان تناسق البيانات عبر المنظمة ودعمها لأهداف المنظمة وغاياتها. تتضمن حوكمة البيانات عدة مكونات رئيسية:

1. سياسة حوكمة البيانات: هذه هي السياسة العامة التي تحدد نهج المؤسسة في حوكمة البيانات. يحدد أهداف وغايات ومسؤوليات إدارة البيانات.
2. إطار حوكمة البيانات: هذا هو الإطار العام الذي يحدد الأدوار والمسؤوليات والعمليات والإجراءات الخاصة بحوكمة البيانات.

3. مجلس حوكمة البيانات: مجموعة من الأفراد المسؤولين عن الإشراف وإدارة البيانات داخل المنظمة .
4. مقاييس حوكمة البيانات: هي المقاييس التي يمكن استخدامها لرصد وقياس فعالية حوكمة البيانات. يمكن أن تشمل هذه المقاييس جودة البيانات وأمن البيانات وخصوصية البيانات وامتثال البيانات .
5. أدوات حوكمة البيانات: وهي الأدوات التي يمكن استخدامها لأتمتة ودعم عمليات حوكمة البيانات. يمكن أن تتضمن هذه الأدوات أدوات جودة البيانات وأدوات أمان البيانات وأدوات خصوصية البيانات وأدوات امتثال البيانات .
6. التدريب على حوكمة البيانات: يقصد به التدريب الذي يتم توفيره للأفراد داخل المنظمة للتأكد من أنهم يفهمون سياسات وإجراءات حوكمة البيانات الخاصة بالمنظمة وقدرتهم على الامتثال لها .

أصبحت حوكمة البيانات ذات أهمية متزايدة مع استمرار نمو كمية البيانات التي يتم إنشاؤها، ويجب على المؤسسات التأكد من أنها يمكن أن تثق في البيانات التي تستخدمها لاتخاذ القرارات. تعد حوكمة البيانات جزءًا أساسيًا من إدارة البيانات وتتطلب مزيجًا من المعرفة الفنية والقانونية والتجارية.

## اخلاقيات البيانات

تشير أخلاقيات البيانات [Data Ethics] إلى المبادئ والقيم التي توجه جمع البيانات واستخدامها وإدارتها. إنه ينطوي على النظر في الآثار الأخلاقية للقرارات المتعلقة بالبيانات، مثل قضايا الخصوصية والأمن والتحيز. أصبحت أخلاقيات البيانات ذات أهمية متزايدة مع استمرار نمو كمية البيانات التي يتم جمعها واستخدامها، ويجب على المؤسسات التأكد من أنها تستخدم البيانات بطريقة مسؤولة وأخلاقية. تتضمن أخلاقيات البيانات عدة مفاهيم أساسية مثل:

1. الخصوصية: يشير هذا المفهوم إلى حماية البيانات الشخصية والحق في التحكم في كيفية جمع هذه البيانات واستخدامها ومشاركتها.
2. الأمان: يشير هذا المفهوم إلى حماية البيانات من الوصول غير المصرح به أو الاستخدام أو الكشف أو التعطيل أو التعديل أو الإتلاف.
3. الشفافية: يشير هذا المفهوم إلى توافر البيانات وإمكانية الوصول إليها، وقدرة الأفراد على فهم كيفية استخدام بياناتهم.
4. التحيز: يشير هذا المفهوم إلى احتمال تحيز البيانات أو إدامة التحيزات الموجودة، والحاجة إلى اتخاذ خطوات للتخفيف من التحيز في البيانات.

5. الإنصاف: يشير هذا المفهوم إلى الحاجة إلى ضمان استخدام البيانات بطريقة عادلة ومنصفة، وتجنب أي عواقب غير مقصودة أو آثار سلبية .

6. المسؤولية: يشير هذا المفهوم إلى حاجة المؤسسات لتحمل المسؤولية عن البيانات التي تجمعها وتستخدمها وتديرها، ولضمان أنها تستخدم البيانات بطريقة مسؤولة وأخلاقية .

تعتبر أخلاقيات البيانات مجالاً معقداً ومتطوراً، ويتطلب نهجاً متعدد التخصصات، يشارك فيه متخصصون من مجالات مثل علوم الحاسوب والقانون والفلسفة وعلم الاجتماع. من المهم للمؤسسات أن تتبنى وتنفذ سياسات وإرشادات وإجراءات أخلاقيات البيانات للتأكد من أنها تستخدم البيانات بطريقة مسؤولة وأخلاقية. وأيضاً، قم بتوفير التدريب والتعليم للموظفين حول موضوعات أخلاقيات البيانات لمساعدتهم على فهم مبادئ أخلاقيات البيانات وتطبيقها في عملهم اليومي.

## هندسة البيانات

هندسة البيانات [Data Engineering] هي عملية تصميم وبناء وصيانة الأنظمة والبنية التحتية المستخدمة في جمع البيانات وتخزينها ومعالجتها وتحليلها. يتضمن مجموعة واسعة من المهام مثل نمذجة البيانات وتخزين البيانات وتكامل البيانات وجودة البيانات وأمن البيانات .

تعد هندسة البيانات مكونًا مهمًا لإدارة البيانات، حيث تتيح للمؤسسات جمع كميات كبيرة من البيانات وتخزينها ومعالجتها بشكل فعال. كما أنه يمكن المنظمات من إتاحة البيانات للتحليل واتخاذ القرار. تتضمن هندسة البيانات عدة مهام رئيسية:

1. نمذجة البيانات: يتضمن ذلك تصميم هياكل البيانات والعلاقات التي سيتم استخدامها لتخزين البيانات في قاعدة بيانات أو مستودع بيانات .
2. تخزين البيانات: يتضمن ذلك تصميم وبناء نظام لتخزين وإدارة والاستعلام عن كميات كبيرة من البيانات .
3. تكامل البيانات: يتضمن ذلك جمع البيانات ودمجها من مصادر متعددة، مثل قواعد البيانات والملفات وواجهات برمجة التطبيقات، وإتاحتها للتحليل .
4. جودة البيانات: يتضمن ذلك التأكد من أن البيانات دقيقة وكاملة ومتسقة، وأنها تلي معايير الجودة الخاصة بالمنظمة .
5. أمان البيانات: يتضمن ذلك حماية البيانات من الوصول غير المصرح به أو الاستخدام أو الكشف أو التعطيل أو التعديل أو الإتلاف .
6. مسار نقل البيانات: يتضمن هذا تصميم وبناء وصيانة العمليات والأنظمة التي تنقل البيانات من المصدر إلى الوجهة .

تتطلب هندسة البيانات مجموعة من المهارات الفنية والتحليلية، بما في ذلك إتقان لغات البرمجة وتصميم قواعد البيانات وتخزين

البيانات. كما يتطلب أيضًا فهمًا لأمن البيانات وخصوصية البيانات، ومعرفة باللوائح الخاصة بالصناعة ومعايير الامتثال. تعد هندسة البيانات جزءًا أساسيًا من إدارة البيانات، وتتطلب مجموعة من المهارات الفنية والتحليلية، بما في ذلك الكفاءة في مسار نقل البيانات، وتخزين البيانات، وتكامل البيانات، وجودة البيانات. مع زيادة كمية البيانات التي يتم إنشاؤها، أصبحت هندسة البيانات مهارة بالغة الأهمية للمؤسسات للحفاظ على قدرتها التنافسية وإتاحة البيانات للتحليل واتخاذ القرار.

### نَسَبُ الْبِیَانَاتِ

نَسَبُ الْبِیَانَاتِ [Data Rates] هِيَ عَمَلِیة تَتَّبَعُ وَتُوَثِّقُ رَحْلَةَ الْبِیَانَاتِ مِنْ أَصْلِهَا إِلَى حَالَتِهَا الْحَالِیةِ وَاسْتِخْدَامِهَا، بِمَا فِي ذَلِكَ جَمِیعَ عَمَلِیَاتِ التَّحْوِيلِ وَالحَرَكَاتِ وَتَخْزِینِ الْبِیَانَاتِ. إِنَّهَا طَرِیْقَةُ لِفَهْمِ تَارِیخِ الْبِیَانَاتِ وَسِیَاقِهَا، وَتَسَاعِدُ عَلَى ضَمَانِ جُودَةِ الْبِیَانَاتِ وَالْأَمَانِ وَالْإِمْتِثَالِ. یَعْدُ نَسَبُ الْبِیَانَاتِ مَهْمًا لِأَنَّهُ یَسْمَحُ لِلْمَوْسَّسَاتِ بِمَا یَلِی:

1. فِهْمُ أَصْلِ الْبِیَانَاتِ وَسِیَاقِهَا: یُوفِّرُ نَسَبُ الْبِیَانَاتِ مَعْلُومَاتَ حَوْلَ مَصْدَرِ الْبِیَانَاتِ وَكِیْفِ تَمَّ تَحْوِيلِهَا، مِمَّا یَسَاعِدُ عَلَى ضَمَانِ أَنَّ الْبِیَانَاتِ دَقِیْقَةٌ وَكَامِلَةٌ وَمُتَسَقَةٌ.
2. تَحْسِینُ جُودَةِ الْبِیَانَاتِ: مِنْ خِلَالِ فِهْمِ نَسَبِ الْبِیَانَاتِ، یُمْكِنُ لِلْمَوْسَّسَاتِ تَحْدِیدَ وَتَصْحِیحَ الْأَخْطَاءِ أَوْ التَّنَاقُضَاتِ فِي الْبِیَانَاتِ، مِمَّا یُؤَدِّي إِلَى تَحْسِینِ جُودَةِ الْبِیَانَاتِ.

3. تعزيز أمان البيانات والامتثال: يوفر نَسب البيانات معلومات حول من يمكنه الوصول إلى البيانات وكيفية استخدامها، مما يساعد على ضمان حماية البيانات من الوصول أو الاستخدام غير المصرح به. كما أنه يساعد المؤسسات على الامتثال للمتطلبات التنظيمية مثل القانون العام لحماية البيانات، وقانون التأمين الصحي وحماية الملكية الفكرية، وقانون حماية خصوصية المستهلك.

4. تسهيل إدارة البيانات: يوفر نَسب البيانات المعلومات التي يمكن استخدامها لدعم عمليات إدارة البيانات، مثل جودة البيانات وأمن البيانات وامتثال البيانات.

5. بيانات الدعم: يمكن أن يساعد نَسب البيانات المؤسسات على فهم وإدارة البيانات التي تجمعها بشكل أفضل، وجعلها أكثر فائدة للتحليل واتخاذ القرار. يمكن التقاط نَسب البيانات وتسجيلها باستخدام تقنيات وتقنيات مختلفة، بما في ذلك أدوات نَسب البيانات، ورسم خرائط نَسب البيانات، وبيانات تعريف نَسب البيانات، وتصور نَسب البيانات. يمكن استخدام هذه الأدوات والتقنيات لالتقاط معلومات نَسب البيانات تلقائيًا، وتزويد المؤسسات بطريقة لعرض نَسب البيانات وتحليلها وإعداد تقارير عنها.

يعد نَسب البيانات جانبًا مهمًا من إدارة البيانات، فهو يسمح للمؤسسات بفهم أصل البيانات وسياقها وتاريخها، مما يساعد على ضمان جودة البيانات والأمان والامتثال. لقد أصبحت مهمة بشكل

متزايد حيث تواجه المنظمات المزيد من المتطلبات التنظيمية وتحتاج إلى أن تكون أكثر مسؤولية عن كيفية جمع البيانات واستخدامها وتخزينها.

## مقاييس البيانات

مقاييس إدارة البيانات [Data metrics] هي قياسات تستخدم لرصد وتقييم فعالية إدارة البيانات داخل المؤسسة. فهي تساعد المؤسسات على تحديد المجالات التي تحتاج إلى تحسينات، وتتبع التقدم بمرور الوقت. يمكن استخدام مقاييس إدارة البيانات لقياس الجوانب المختلفة للبيانات، بما في ذلك :

1. جودة البيانات: يتضمن ذلك مقاييس مثل النسبة المئوية للبيانات الدقيقة والكاملة والمتسقة وعدد الأخطاء أو التناقضات في البيانات .
2. أمان البيانات: يتضمن هذا مقاييس مثل عدد عمليات اختراق البيانات، وعدد محاولات الوصول غير المصرح بها، والنسبة المئوية للبيانات المشفرة .
3. خصوصية البيانات: يتضمن ذلك مقاييس مثل عدد حوادث الخصوصية وعدد انتهاكات البيانات والنسبة المئوية للبيانات المجهولة المصدر .
4. امتثال البيانات: يتضمن هذا مقاييس مثل النسبة المئوية للبيانات المتوافقة مع المتطلبات التنظيمية، وعدد حوادث الامتثال .

5. عمليات إدارة البيانات: تتضمن مقاييس مثل عدد سياسات وإجراءات حوكمة البيانات المعمول بها، والنسبة المئوية لمالكي البيانات الذين أكملوا التدريب على حوكمة البيانات، وعدد اجتماعات مجلس إدارة البيانات التي عقدت .
6. أداء إدارة البيانات: يتضمن هذا مقاييس مثل النسبة المئوية لمشكلات جودة البيانات التي يتم حلها ضمن إطار زمني محدد، والنسبة المئوية لحوادث أمان البيانات التي يتم حلها خلال إطار زمني محدد، والنسبة المئوية لحوادث خصوصية البيانات التي يتم حلها خلال إطار زمني محدد .

تعد مقاييس إدارة البيانات أداة مهمة للمؤسسات لمراقبة وإدارة البيانات التي تجمعها وتستخدمها وتخزنها بشكل فعال. أنها توفر للمنظمات طريقة لتحديد المجالات التي تحتاج إلى تحسينات، وتتبع التقدم مع مرور الوقت. من المهم أن تكون هذه المقاييس واضحة وقابلة للقياس وذات مغزى وأن تتماشى مع الأهداف والغايات العامة للمؤسسة.

### أدوات إدارة البيانات

أدوات إدارة البيانات هي تطبيقات ومنصات برمجية تُستخدم لأتمتة ودعم عمليات حوكمة البيانات. يمكنهم مساعدة المؤسسات على إدارة وتنفيذ سياسات وإجراءات ومعايير حوكمة البيانات، وتحسين جودة البيانات والأمان والامتثال. تتضمن أمثلة أدوات إدارة البيانات ما يلي:

## أولاً: منصات إدارة البيانات

منصات إدارة البيانات [Data Platforms] هي تطبيقات ومنصات برمجية توفر مستودعًا مركزيًا لسياسات وإجراءات ومعايير حوكمة البيانات. يمكنها أيضًا توفير أدوات لأتمتة عمليات إدارة البيانات، مثل تصنيف البيانات، والإشراف على البيانات، ونسب البيانات. يمكن أن تساعد منصات إدارة البيانات المؤسسات على:

- مركزية وأتمتة عمليات إدارة البيانات: توفر الأنظمة الأساسية لإدارة البيانات موقعًا مركزيًا لسياسات وإجراءات ومعايير حوكمة البيانات، ويمكنها أتمتة عملية إنشاء السياسات وتحديثها والموافقة عليها.
- تحسين جودة البيانات: يمكن أن توفر الأنظمة الأساسية لإدارة البيانات أدوات لتوصيف البيانات وتنقية البيانات وتحويل البيانات، والتي يمكن استخدامها لتحسين جودة البيانات.
- تعزيز أمان البيانات والامتثال: يمكن أن توفر الأنظمة الأساسية الأساسية لإدارة البيانات أدوات لتشفير البيانات وإخفاء البيانات ومنع فقدان البيانات، والتي يمكن استخدامها لحماية البيانات من الوصول أو الاستخدام غير المصرح به. يمكنهم أيضًا مساعدة المؤسسات على الامتثال للمتطلبات التنظيمية مثل GDPR و HIPAA و CCPA.
- تسهيل إدارة البيانات: يمكن أن توفر الأنظمة الأساسية لإدارة البيانات أدوات لأتمتة عملية إدارة البيانات، مثل

تصنيف البيانات، والإشراف على البيانات، ونسب البيانات، والتي يمكن استخدامها لدعم عمليات حوكمة، مثل جودة البيانات، وأمن البيانات، وامتثال البيانات .

- دعم إدارة البيانات: يمكن أن توفر الأنظمة الأساسية لإدارة البيانات أدوات لأتمتة عملية إدارة البيانات، مثل تصنيف البيانات، والإشراف على البيانات، ونسب البيانات، والتي يمكن استخدامها لدعم جهود إدارة البيانات، مثل جودة البيانات، وأمن البيانات، وامتثال البيانات .

يمكن استخدام منصات إدارة البيانات لأتمتة عمليات حوكمة البيانات وإدارتها، مما يضمن جودة البيانات والأمان والامتثال. كما أنها توفر موقعًا مركزيًا لسياسات وإجراءات ومعايير حوكمة البيانات. مع زيادة كمية البيانات التي يتم إنشاؤها، من المهم أن تقوم المؤسسات بتنفيذ منصات إدارة البيانات للحفاظ على قدرتها التنافسية وحماية البيانات التي تجمعها وتستخدمها وتخزنها.

### ثانياً: أدوات جودة البيانات

أدوات جودة البيانات هي تطبيقات برمجية تُستخدم لتقييم ومراقبة وتحسين جودة البيانات. يمكن استخدامها لتحديد أخطاء البيانات وعدم الاتساق، ولإجراء تنقية البيانات وتحويل البيانات. تتضمن أمثلة أدوات جودة البيانات ما يلي :

1. تنميط البيانات: تُستخدم هذه الأدوات لتحليل وفهم بنية البيانات ومحتواها وعلاقاتها. يمكن استخدامها لتحديد

أخطاء البيانات وعدم الاتساق، وإنشاء مقاييس جودة البيانات.

2. تنقية البيانات: تُستخدم هذه الأدوات لتحديد أخطاء البيانات وعدم تناسقها وتصحيحها، مثل السجلات المكررة والقيم المفقودة والبيانات غير الصالحة. يمكن استخدامها أيضًا لتوحيد البيانات، مثل تحويل البيانات إلى تنسيق أو حالة متسقة.

3. تحويل البيانات: تُستخدم هذه الأدوات لتحويل البيانات من تنسيق أو بنية إلى أخرى، مثل تحويل البيانات من ملف ثابت إلى قاعدة بيانات علائقية. يمكن استخدامها أيضًا لإجراء حسابات البيانات وتجميعات البيانات.

4. التحقق من صحة البيانات: تُستخدم هذه الأدوات للتحقق من صحة البيانات مقابل القواعد والقيود المحددة، مثل نوع البيانات والطول والنطاق. يمكن استخدامها أيضًا للتحقق من صحة البيانات مقابل مصادر خارجية، مثل البيانات المرجعية أو البيانات الرئيسية.

5. إثراء البيانات: تُستخدم هذه الأدوات لإضافة معلومات إضافية إلى البيانات، مثل الترميز الجغرافي لعنوان أو إضافة معلومات ديموغرافية. لوحة معلومات جودة البيانات: توفر هذه الأدوات تمثيلًا مرئيًا لمقاييس جودة البيانات، مثل اكتمال البيانات ودقة البيانات واتساق البيانات.

تعد أدوات جودة البيانات جانبًا مهمًا من إدارة البيانات، فهي تسمح للمؤسسات بضمان أن البيانات التي تجمعها وتستخدمها وتخزنها دقيقة وكاملة ومتسقة. مع زيادة كمية البيانات التي يتم إنشاؤها، توفر أدوات جودة البيانات وسيلة للمؤسسات لفهم بياناتها، وإتاحة البيانات للتحليل واتخاذ القرار.

### إدارة مشاريع البيانات

تتضمن إدارة مشروع البيانات التخطيط والتنفيذ والتحكم في المشاريع التي تتضمن المهام المتعلقة بالبيانات، مثل جمع البيانات والمعالجة والتحليل والتصوير. مديرو مشروع البيانات مسؤولون عن ضمان اكتمال مشاريع البيانات في الوقت المحدد وفي حدود الميزانية وبما يرضي أصحاب المصلحة. تشمل المسؤوليات الرئيسية لإدارة مشروع البيانات ما يلي:

- تحديد نطاق المشروع وأهدافه.
- تحديد متطلبات البيانات.
- تطوير خطة المشروع.
- إدارة الجداول الزمنية للمشروع والميزانيات.
- تنسيق المهام المتعلقة بالبيانات بين أعضاء الفريق.
- تقرير تقدم المشروع لأصحاب المصلحة.
- ضمان جودة البيانات ودقتها.
- إدارة مخاطر المشروع.
- إدارة وثائق المشروع.

قد يستخدم مدير مشروع البيانات مجموعة متنوعة من منهجيات إدارة المشروع، مثل Agile أو Waterfall ، اعتماداً على المتطلبات المحددة للمشروع. يمكنهم أيضاً استخدام برنامج إدارة المشاريع لتتبع التقدم والتعاون مع أعضاء الفريق. تُستخدم إدارة مشروع البيانات في مجموعة واسعة من المجالات، بما في ذلك الرعاية الصحية والتمويل والتسويق، وقد أصبح مجالاً مهماً بشكل متزايد مع استمرار نمو كمية البيانات التي يتم إنشاؤها. وفيما يأتي عرض مفصل للمسؤوليات الرئيسية لإدارة مشروعات البيانات:

### أولاً: تحديد نطاق المشروع وأهدافه

يعد تحديد نطاق المشروع وأهدافه خطوة حاسمة في عملية إدارة المشروع. إنه يتضمن تحديداً واضحاً ومحددًا لما سينجزه المشروع، والمخرجات التي سيتم إنتاجها، والقيود التي يجب أخذها في الاعتبار. يشير نطاق المشروع إلى الحدود المحددة للمشروع، بما في ذلك ما يتم تضمينه واستبعاده. من المهم تحديد نطاق المشروع في وقت مبكر للتأكد من أن كل شخص مشارك لديه فهم واضح لما سينجزه المشروع، وللمساعدة في منع زحف النطاق، حيث يأخذ المشروع مهامًا أو أهدافًا أكثر مما كان مخططاً له في الأصل .

أهداف المشروع هي الأهداف أو النتائج المحددة التي يهدف المشروع إلى تحقيقها. يجب أن تكون محددة وقابلة للقياس وقابلة للتحقيق وذات صلة ومحددة زمنياً. تساعد أهداف المشروع في توجيه المشروع وتوفير توجيه واضح للفريق لاتباعه. يمكن تحديد نطاق المشروع وأهدافه من خلال إنشاء ميثاق المشروع، وهو مستند يحدد

غرض المشروع وأهدافه وأصحاب المصلحة الرئيسيين. تُستخدم هذه الوثيقة كنقطة مرجعية في جميع مراحل المشروع وتساعد على ضمان أن يكون لدى كل شخص مشارك فهم واضح لأهداف المشروع وقيوده. كما أنه يساعد على إبقاء المشروع على المسار الصحيح ومنع زحف النطاق.

### ثانياً: تحديد متطلبات البيانات

يعد تحديد متطلبات البيانات خطوة مهمة في إدارة مشروع البيانات. يتضمن تحديد البيانات المطلوبة للمشروع، ومن أين ستأتي، وكيف سيتم استخدامها. لتحديد متطلبات البيانات، يجب على مدير مشروع البيانات:

- افهم مشكلة أو فرصة العمل التي يحاول المشروع معالجتها.
- حدد أصحاب المصلحة الرئيسيين وصناع القرار الذين سيستخدمون البيانات.
- حدد البيانات المحددة اللازمة لدعم المشروع.
- فهم مصادر البيانات وكيف سيتم جمع البيانات ومعالجتها.
- تحديد تنسيق البيانات وهيكلها ومتطلبات الجودة.
- حدد أي متطلبات تنظيمية أو امتثال يجب الوفاء بها.

من المهم أن يتم تحديد متطلبات البيانات في وقت مبكر من المشروع بحيث يمكن إدراجها في خطة المشروع والجدول الزمني. يساعد تحديد متطلبات البيانات أيضاً على ضمان توفر البيانات الضرورية عند الحاجة إليها، وأنها ذات جودة كافية لدعم أهداف

المشروع. يمكن إجراء هذه العملية عن طريق إنشاء مستند متطلبات البيانات، وهو وصف تفصيلي لاحتياجات البيانات الخاصة بالمشروع. يجب أن تتضمن هذه الوثيقة معلومات مثل مصادر البيانات، ومتطلبات جودة البيانات، وتنسيق البيانات وهيكلها، وأي متطلبات تنظيمية أو متطلبات الامتثال التي يجب الوفاء بها. سيتم استخدامه كنقطة مرجعية طوال المشروع وسيساعد على ضمان أن يكون لدى فريق مشروع البيانات فهم واضح لمتطلبات البيانات.

### ثالثاً: تطوير خطة المشروع

يعد تطوير خطة المشروع خطوة حاسمة في إدارة المشروع. يتضمن إنشاء خارطة طريق مفصلة لإكمال المشروع، بما في ذلك المهام التي يجب القيام بها، والموارد المطلوبة، والجدول الزمني للإنجاز. لتطوير خطة المشروع، يجب على مدير المشروع:

- تقسيم المشروع إلى مهام أصغر يمكن إدارتها.
- تحديد التبعيات بين المهام.
- تقدير الوقت والموارد المطلوبة لكل مهمة.
- وضع جدولاً زمنياً لاستكمال المشروع.
- تحديد المخاطر المحتملة ووضع خطة لإدارة المخاطر.
- تعيين الأدوار والمسؤوليات لأعضاء الفريق.

تعمل خطة المشروع كدليل لفريق المشروع، ويجب مراجعتها وتحديثها بانتظام لضمان بقاء المشروع على المسار الصحيح. إنها أيضاً أداة رئيسية لإبلاغ تقدم المشروع لأصحاب المصلحة والتأكد من أن

الجميع على دراية بحالة المشروع. يمكن تطوير خطة المشروع باستخدام مجموعة متنوعة من الأدوات والتقنيات، مثل مخططات جانث وتحليل المسار الحرج ومخططات بيرت. يمكن أيضًا إدارتها باستخدام برنامج إدارة المشروع الذي يسمح لأعضاء الفريق بالتعاون وتتبع التقدم والإبلاغ عن حالة المشروع. من المهم ملاحظة أن خطة المشروع هي وثيقة حية، ويجب تحديثها مع تقدم المشروع لتعكس أي تغييرات في النطاق أو الجدول الزمني أو الموارد.

#### رابعاً: إدارة الميزانية والجدول الزمنية للمشروع

تعد إدارة الجدول الزمنية للمشروع والميزانيات جانبًا حاسمًا في إدارة المشروع. يتضمن ضمان اكتمال المشروع في الوقت المحدد وضمن قيود الميزانية. لإدارة الجدول الزمنية للمشروع والميزانيات، يجب على مدير المشروع:

- تطوير جدول زمني مفصل للمشروع، بما في ذلك المعالم والمواعيد النهائية المحددة.
- تتبع تقدم المشروع بانتظام ومقارنته بالجدول الزمني.
- تحديد ومعالجة أي تأخير أو مشاكل قد تنشأ.
- اضبط الجدول الزمني حسب الحاجة لضمان بقاء المشروع على المسار الصحيح.
- اعلام مراحل تقدم المشروع لأصحاب المصلحة لإبقائهم على اطلاع.

من حيث الميزانية، يجب على مدير المشروع:

- وضع ميزانية مفصلة للمشروع، بما في ذلك جميع التكاليف المرتبطة بالمشروع.
- تتبع نفقات المشروع ومقارنتها بالميزانية.
- تحديد ومعالجة أي تجاوزات في الميزانية.
- ضبط الميزانية حسب الحاجة لضمان بقاء المشروع ضمن قيود الميزانية.
- تقديم تقرير عن حالة الميزانية لأصحاب المصلحة لإبقائهم على اطلاع.

تعد إدارة الجداول الزمنية والميزانيات بشكل فعال أمرًا بالغ الأهمية لنجاح المشروع. فهي تساعد على ضمان بقاء المشروع على المسار الصحيح واستخدام الموارد بفعالية. من المهم ملاحظة أن الجداول الزمنية للمشروع والميزانيات مرتبطة ارتباطًا وثيقًا، ويمكن للتغيرات في أحدهما أن تؤثر على الآخر. لذلك من المهم النظر في التأثير المحتمل لأي تغييرات على كل من الجدول الزمني والميزانية وإجراء التعديلات حسب الحاجة.

#### خامساً: تنسيق واجبات الفريق

يعد تنسيق المهام المتعلقة بالبيانات بين أعضاء الفريق جانبًا مهمًا من إدارة مشروع البيانات. إنه ينطوي على التأكد من أن جميع أعضاء الفريق على دراية بمسؤولياتهم، وأن لديهم الموارد التي يحتاجونها لإكمال مهامهم. لتنسيق المهام المتعلقة بالبيانات بين أعضاء الفريق، يجب على مدير مشروع البيانات:

- تحديد الأدوار والمسؤوليات لكل عضو في الفريق .
- التأكد من أن أعضاء الفريق لديهم المهارات والمعرفة والموارد اللازمة لإكمال مهامهم .
- القيام بجدولة اجتماعات منتظمة لمناقشة التقدم الحاصل ومعالجة أي قضايا .
- استخدام برنامج إدارة المشروع للتعاون ومشاركة المعلومات مع أعضاء الفريق .
- التواصل مع أعضاء الفريق للتأكد من أنهم على دراية بأي تغييرات أو تحديثات على خطة المشروع .

من المهم أيضًا إنشاء خطوط اتصال واضحة ونظام للإبلاغ عن التقدم، بحيث يمكن لأعضاء الفريق مشاركة المعلومات بسهولة والبقاء على اطلاع على حالة المشروع. يعد تنسيق المهام المتعلقة بالبيانات بين أعضاء الفريق أمرًا مهمًا لضمان بقاء المشروع على المسار الصحيح وأن الجميع على دراية بمسؤولياتهم. كما أنه يساعد على تقليل مخاطر التأخير أو الأخطاء، ولضمان تلبية المشروع لاحتياجات أصحاب المصلحة. من المهم ملاحظة أن إدارة مشروع البيانات تتطلب مستوى عالٍ من التعاون والتنسيق، وينبغي تشجيع أعضاء الفريق على التواصل وتبادل المعلومات مع بعضهم البعض طوال فترة المشروع.

### سادساً: مراحل تقدم المشروع

يعتبر الإبلاغ عن تقدم المشروع لأصحاب المصلحة جانباً مهماً من إدارة المشروع. وهو ينطوي على إبقاء أصحاب المصلحة على علم بحالة المشروع، ومعالجة أي مخاوف أو قضايا قد تنشأ. للإبلاغ عن تقدم المشروع لأصحاب المصلحة، يجب على مدير المشروع:

- اعلام أصحاب المصلحة بانتظام بشأن تقدم المشروع، بما في ذلك أي معالم تم تحقيقها والمهام القادمة والمواعيد النهائية.
- استخدام مجموعة متنوعة من طرق الاتصال، مثل البريد الإلكتروني والهاتف والاجتماعات الشخصية للوصول إلى أصحاب المصلحة.
- تزويد أصحاب المصلحة بتقرير مفصل عن تقدم المشروع، بما في ذلك أي معوقات وكيفية معالجتها.
- معالجة أي مخاوف أو أسئلة قد تكون لدى أصحاب المصلحة في الوقت المناسب وبطريقة مهنية.
- قم بإبلاغ عن أي تغييرات على خطة المشروع أو الجدول الزمني لأصحاب المصلحة بمجرد حدوثها.

يعد التواصل الفعال مع أصحاب المصلحة أمراً بالغ الأهمية لضمان بقاء المشروع على المسار الصحيح وأن أصحاب المصلحة على دراية بحالة المشروع. يساعد على بناء الثقة والحفاظ على العلاقات، كما أنه يسمح لأصحاب المصلحة بتقديم الملاحظات وتقديم الاقتراحات التي من شأنها تحسين المشروع. من المهم ملاحظة أن إبلاغ

تقدم المشروع لأصحاب المصلحة يجب أن يتم بطريقة واضحة وموجزة ومتسقة. يجب أن يتم تصميمه وفقاً للاحتياجات المحددة لكل صاحب مصلحة ويجب أن يتم بشكل منتظم، وليس فقط في نهاية المشروع.

### سابعاً: ضمان جودة البيانات ودقتها

يعد ضمان جودة البيانات ودقتها جانباً مهماً في إدارة مشروع البيانات. يتضمن التأكد من أن البيانات المستخدمة في المشروع دقيقة وكاملة وذات صلة. لضمان جودة البيانات ودقتها، يجب على مدير مشروع البيانات :

- تطوير وتنفيذ معايير وإجراءات جودة البيانات .
  - التحقق من صحة البيانات والتحقق منها بانتظام لضمان دقتها واكتمالها .
  - تحديد وتصحيح أي أخطاء أو تناقضات في البيانات .
  - استخدام أدوات وتقنيات جودة البيانات لأتمتة التحقق من صحة البيانات وتصحيحها .
  - إجراء تحديد البيانات وتحليلها لتحديد مشاكل جودة البيانات وتصحيحها.
  - تثقيف أعضاء الفريق بشأن أفضل ممارسات جودة البيانات، وتشجيعهم على الإبلاغ عن أي مشكلات يواجهونها.
- يعد ضمان جودة البيانات ودقتها أمراً مهماً لنجاح المشروع، حيث يساعد على ضمان أن البيانات المستخدمة موثوقة ويمكن

الوثوق بها. كما أنه يساعد في تقليل مخاطر الأخطاء وعدم الاتساق في البيانات، مما قد يؤدي إلى استنتاجات غير دقيقة واتخاذ قرارات خاطئة. من المهم ملاحظة أن جودة البيانات ودقتها هي عمليات مستمرة تحتاج إلى مراقبتها وإدارتها طوال فترة المشروع. هذا يعني أن مدير مشروع البيانات يجب أن ينشئ عملية حوكمة البيانات، والتي تضمن الحفاظ على جودة البيانات ودقتها طوال دورة حياة البيانات بأكملها.

### ثامناً: إدارة مخاطر المشروع

تعد إدارة مخاطر وقضايا المشروع جانباً مهماً من جوانب إدارة المشروع. يتضمن تحديد المخاطر والمشاكل المحتملة التي قد تنشأ أثناء المشروع، ووضع استراتيجيات للتخفيف منها أو معالجتها. لإدارة مخاطر المشروع، يجب على مدير المشروع:

- تحديد المخاطر والقضايا المحتملة من خلال مراجعة خطة المشروع وتقييم البيئة والتشاور مع أصحاب المصلحة.
- قم بتقييم احتمالية وتأثير كل خطر أو مشكلة.
- تحديد أولويات المخاطر والقضايا بناءً على احتمالية حدوثها وتأثيرها.
- تطوير استراتيجيات لتخفيف أو معالجة كل خطر أو مشكلة.
- مراقبة ومراجعة المخاطر والمشكلات بانتظام، وتحديث خطة إدارة المخاطر حسب الحاجة.

إدارة المخاطر هي عملية مستمرة يجب مراقبتها ومراجعتها بشكل مستمر طوال فترة المشروع. يساعد مدير المشروع على تحديد ومنع المشاكل المحتملة، ويساعد على تقليل تأثير أي مشاكل قد تظهر. من المهم ملاحظة أن الإدارة الفعالة للمخاطر تتطلب نهجًا استباقيًا، حيث يتم تحديد المخاطر والقضايا ومعالجتها قبل أن تصبح مشاكل رئيسية. كما يتطلب اتصالًا جيدًا، حيث يحتاج مدير المشروع إلى إبقاء أصحاب المصلحة على اطلاع بالمخاطر والقضايا وكيفية إدارتها.

### تاسعاً" إدارة وثائق المشروع

تعد إدارة وثائق المشروع جانبًا مهمًا من جوانب إدارة المشروع. يتضمن إنشاء المستندات المتعلقة بالمشروع وصيانتها وتنظيمها، مثل خطط المشروع والجدول الزمني والميزانيات ومحاضر الاجتماعات وتقارير الحالة. لإدارة وثائق المشروع، يجب على مدير المشروع:

- إنشاء نظام مركزي لتخزين وإدارة مستندات المشروع .
- إنشاء مسميات وهيكل ملفات واضحة لوثائق المشروع .
- تأكد من أن جميع المستندات المتعلقة بالمشروع دقيقة وحديثة ويسهل الوصول إليها من قبل أعضاء الفريق .
- قم بمراجعة وتحديث مستندات المشروع بانتظام حسب الحاجة .
- أرشفة وثائق المشروع في نهاية المشروع للرجوع إليها في المستقبل.

تعد إدارة وثائق المشروع بشكل صحيح أمرًا مهمًا لضمان سهولة الوصول إلى جميع المعلومات المتعلقة بالمشروع وتنظيمها جيدًا. كما أنه يساعد على التأكد من أن جميع أعضاء الفريق على دراية بحالة المشروع ولديهم المعلومات التي يحتاجونها لإكمال مهامهم. كما أنه يساعد على تتبع تقدم المشروع وإنجازاته. من المهم ملاحظة أنه يجب إنشاء وثائق المشروع وتحديثها ومشاركتها بطريقة متسقة وفي الوقت المناسب. يجب أيضًا أن تكون مصممة وفقًا للاحتياجات المحددة لكل صاحب مصلحة وأن تكون سهلة الفهم. سيضمن ذلك أن فريق المشروع يمكنه الوصول إلى المعلومات التي يحتاجون إليها لاتخاذ القرارات وإكمال عملهم، وأن أصحاب المصلحة لديهم إمكانية الوصول إلى المعلومات التي يحتاجونها لفهم تقدم المشروع ونتائجه.

### عوامل نجاح مشروع البيانات

هناك العديد من العوامل الرئيسية التي يمكن أن تسهم في

نجاح مشروع إدارة البيانات :

- أهداف المشروع الواضحة: يجب أن يكون للمشروع أهداف وغايات محددة بوضوح تتماشى مع الإستراتيجية العامة للمنظمة .
- إدارة قوية للمشروع: يجب تخطيط المشروع وتنفيذه والتحكم فيه بشكل فعال، مع خطة مشروع وجدول زمني وميزانية محددة جيدًا .

- جودة البيانات ودقتها: يجب أن يضمن المشروع أن البيانات المستخدمة دقيقة وكاملة وذات صلة، وأن معايير جودة البيانات المناسبة موجودة
- إدارة البيانات: يجب أن يلتزم المشروع بأفضل الممارسات واللوائح لإدارة وحماية والتحكم في الوصول إلى البيانات .
- إشراك أصحاب المصلحة: يجب أن يشارك المشروع أصحاب المصلحة في تخطيط المشروع وتنفيذه، والتأكد من أخذ احتياجاتهم في الاعتبار .
- التواصل الجيد: يجب أن يكون للمشروع خطة اتصال واضحة وفعالة، مع تحديثات منتظمة حول تقدم المشروع .
- المرونة: يجب أن يكون المشروع مرناً بما يكفي للتكيف مع التغييرات في المتطلبات أو المشكلات غير المتوقعة .
- العمل الجماعي: يجب أن يعزز المشروع بيئة تعاونية حيث يمكن لأعضاء الفريق مشاركة المعلومات والعمل معاً بفعالية .
- الخبرة الفنية: يجب أن يتمتع المشروع بإمكانية الوصول إلى الخبرة الفنية اللازمة لتصميم وتنفيذ وصيانة أنظمة إدارة البيانات .
- التحسين المستمر: يجب أن يكون للمشروع عملية تحسين مستمرة

## عوامل فشل مشروع البيانات

هناك العديد من العوامل الرئيسية التي يمكن أن تسهم في

فشل مشروع إدارة البيانات :

- عدم وجود أهداف واضحة للمشروع: قد يفتقر المشروع إلى أهداف أو أهداف محددة بوضوح، مما قد يؤدي إلى الارتباك والافتقار إلى الاتجاه .
- ضعف إدارة المشروع: قد لا يتم التخطيط للمشروع أو تنفيذه أو التحكم فيه بشكل فعال، مما يؤدي إلى تأخيرات وتجاوزات في الميزانية ونتائج سيئة الجودة .
- قضايا جودة البيانات: قد لا يكون للمشروع معايير جودة بيانات مناسبة، أو قد لا يتحقق بشكل فعال من دقة البيانات واكتمالها والتحقق منها .
- الافتقار إلى إدارة البيانات: قد لا يلتزم المشروع بأفضل الممارسات واللوائح لإدارة وحماية والتحكم في الوصول إلى البيانات، مما قد يؤدي إلى انتهاكات للبيانات أو مشكلات الامتثال .
- مشاركة غير كافية لأصحاب المصلحة: قد لا يشرك المشروع أصحاب المصلحة في تخطيط المشروع وتنفيذه، مما قد يؤدي إلى نقص في التأييد والدعم .
- ضعف التواصل: قد يفتقر المشروع إلى خطة اتصال واضحة وفعالة، مما قد يؤدي إلى سوء الفهم والتأخير .

- نقص المرونة: قد لا يكون المشروع مرناً بما يكفي للتكيف مع التغييرات في المتطلبات أو المشكلات غير المتوقعة، مما قد يؤدي إلى التأخير أو الفشل في تحقيق الأهداف .
- الافتقار إلى العمل الجماعي: قد يفتقر المشروع إلى بيئة تعاونية حيث يمكن لأعضاء الفريق مشاركة المعلومات والعمل معاً بشكل فعال، مما قد يؤدي إلى تأخيرات أو نتائج سيئة الجودة .
- الخبرة الفنية غير الكافية: قد لا يتمكن المشروع من الوصول إلى الخبرة الفنية اللازمة للتصميم،

### مهارات مدير مشروع البيانات

يجب أن يمتلك مدير مشروع البيانات مجموعة من المهارات الفنية والتحليلية والإدارية. فيما يلي بعض المهارات الأساسية المطلوبة عادةً لهذا الدور :

- المهارات الفنية: الإلمام بتقنيات وأدوات إدارة البيانات مثل قواعد البيانات وتخزين البيانات واستخراج البيانات وتصوير البيانات .
- المهارات التحليلية: القدرة على تحليل وتفسير البيانات، وتحديد الأنماط والاتجاهات، واستخلاص النتائج .
- نمذجة البيانات: معرفة تقنيات وأدوات نمذجة البيانات لتصميم وتنفيذ وصيانة بنية بيانات متسقة ودقيقة .

- مهارات إدارة المشروع: القدرة على تخطيط وتنفيذ ومراقبة مشاريع إدارة البيانات، ضمن قيود الوقت والميزانية .
- مهارات الاتصال: القدرة على التواصل بشكل فعال مع مختلف أصحاب المصلحة، بما في ذلك الجماهير التقنية وغير الفنية .
- مهارات حل المشكلات: القدرة على تحديد المشكلات والقضايا المتعلقة بالبيانات وتحليلها وحلها .
- القيادة: القدرة على قيادة وإدارة فريق من المتخصصين في إدارة البيانات .
- إدارة البيانات: معرفة أفضل الممارسات واللوائح لإدارة وحماية والتحكم في الوصول إلى البيانات .
- إدارة المخاطر: القدرة على تحديد وإدارة مخاطر المشروع، ووضع استراتيجيات للتخفيف منها أو معالجتها .
- القدرة على التكيف: القدرة على التكيف مع التقنيات الجديدة والتغيرات في ممارسات إدارة البيانات .

من المهم ملاحظة أن المهارات المحددة المطلوبة لدور مدير مشروع البيانات يمكن أن تختلف اعتمادًا على حجم وتعقيد المنظمة، وكذلك الاحتياجات المحددة للمشروع.

### أهمية دراسة إدارة البيانات

تعد إدارة البيانات موضوعًا مهمًا لطلاب علوم المعلومات لدراسته لعدة أسباب :

1. الفرص الوظيفية: تعد إدارة البيانات مجالاً متنامياً مع ارتفاع الطلب على المهنيين ذوي المهارات اللازمة لإدارة وتحليل كميات كبيرة من البيانات. يمكن أن تفتح دراسة إدارة البيانات فرصاً وظيفية في مجالات مثل علم البيانات وذكاء الأعمال وإدارة المعلومات .
2. الصلة بالعالم الحقيقي: مهارات إدارة البيانات وثيقة الصلة بالعالم الحقيقي. يمكن أن يساعد فهم كيفية إدارة البيانات بفعالية المؤسسات على اتخاذ قرارات أفضل وتحسين العمليات واكتساب ميزة تنافسية .
3. حل المشكلات: تتطلب إدارة البيانات القدرة على تحليل وحل المشكلات المتعلقة بالبيانات. هذه المهارة قابلة للتحويل بشكل كبير إلى مجالات أخرى .
4. إدارة البيانات: أصبحت إدارة البيانات موضوعاً ساخناً في الصناعة وفهم مفاهيم حوكمة البيانات مهم لطلاب علوم المعلومات لتطوير المهارات المهنية في إدارة البيانات وحوكمة البيانات. الامتثال: يتيح فهم مفاهيم إدارة البيانات، وخاصة حوكمة البيانات، للطلاب فهم المتطلبات القانونية والتنظيمية التي يجب على المؤسسات اتباعها عند إدارة البيانات، بما في ذلك خصوصية البيانات وأمن البيانات .
5. تعدد التخصصات: إدارة البيانات هي مجال متعدد التخصصات، يتطلب المعرفة والمهارات من مختلف المجالات مثل علوم الحاسوب والإحصاء والأعمال .

من خلال دراسة إدارة البيانات، سيكتسب طلاب علوم المعلومات مجموعة مهارات قيّمة يمكن تطبيقها في مجموعة واسعة من المجالات والمؤسسات، ومساعدتهم على أن يصبحوا محترفين ومجهزين للتميز في العالم القائم على البيانات.

## اسئلة الفصل الرابع

1. ما المقصود بقواعد البيانات .؟ وما هي اهم انواعها.؟
2. ما اهمية حوكمة البيانات .؟ وما هي اهم مكوناتها.؟
3. تكلم بنقاط عن المفاهيم الاساسية لاخلاقيات البيانات.؟
4. ما المقصود بهندسة البيانات.؟ وما هي مهامها الرئيسية.؟
5. كيف يمكن لمنصات البيانات مساعدة مؤسسات المعلومات.؟  
تكلّم بنقاط.
6. تشمل المسؤوليات الرئيسية لإدارة مشروع البيانات تسعة مراحل. عددها وشرح واحدة منها.؟
7. ما هي الجوانب التي يمكن توظيف مقاييس البيانات فيها. ضمن اطار عمل ادارة مشاريع البيانات.
8. هناك العديد من العوامل الرئيسية التي يمكن أن تسهم في نجاح مشروع إدارة البيانات. تكلم عنها بالتفصيل.
9. هناك العديد من العوامل الرئيسية التي يمكن أن تسهم في فشل مشروع إدارة البيانات. تكلم عنها بالتفصيل.
10. ما هي مجموعة المهارات الفنية والتقنية والادارية المطلوبة لمدير مشروع البيانات. ؟

## الفصل الخامس

### علم البيانات: الأدوار والمسؤوليات

#### تمهيد

في هذا الفصل، سوف نلقي نظرة فاحصة على الأدوار والمسؤوليات المختلفة في مجال علم البيانات. علم البيانات هو مجال متعدد التخصصات يتضمن استخدام الأساليب الإحصائية والحسابية لاستخراج الأفكار والمعرفة من البيانات. هناك العديد من الأدوار المختلفة في علم البيانات، ولكل منها مجموعة فريدة من المسؤوليات والمهارات.

سنقوم بفحص أدوار محلل البيانات، ومهندس البيانات، ومهندس التعلم الآلي، ومحلل ذكاء الأعمال، ومهندس تصور البيانات، وعالم البيانات. سنناقش مسؤوليات كل دور، والمهارات المطلوبة للنجاح في كل دور، وأنواع المشاريع التي يعمل عليها كل دور عادة.

بنهاية هذا الفصل، سيكون لديك فهم أفضل للأدوار المختلفة في علم البيانات والمهارات المطلوبة للنجاح في كل دور. ستكون هذه المعلومات مفيدة لأي شخص مهتم بمهنة في علم البيانات، وكذلك للمؤسسات التي تتطلع إلى بناء فريق علم البيانات.

هذا الفصل مهم لطلاب علم المعلومات لأنه يقدم نظرة عامة شاملة للأدوار والمسؤوليات المختلفة في مجال علم البيانات. يمكن

أن يساعد فهم الأدوار المختلفة في علم البيانات الطلاب على اتخاذ قرارات مستنيرة حول المسار الوظيفي الذي يجب متابعته، وأنواع المهارات والمعرفة التي سيحتاجون إلى تطويرها من أجل تحقيق النجاح في هذا الدور .

فضلا عن ذلك، يمكن أن يساعد هذا الفصل الطلاب على فهم الأنواع المختلفة من المشاريع التي يعمل عليها متخصصو علوم البيانات وأنواع المؤسسات التي توظف متخصصين في علم البيانات . يمكن أن تكون هذه المعلومات مفيدة للطلاب عند التفكير في التدريب أو فرص العمل، ويمكن أن تساعد في تحديد مجالات علم البيانات التي تتوافق مع اهتماماتهم وأهدافهم المهنية .علاوة على ذلك، أصبح علم البيانات أكثر أهمية في مجال علم المعلومات، حيث تستمر كمية البيانات التي يتم إنشاؤها في النمو .مع وضع ذلك في الاعتبار، يمكن أن تساعد معرفة الأدوار والمسؤوليات في علم البيانات الطلاب على فهم كيفية تطبيق علم البيانات على مجالات مختلفة من علم المعلومات، مثل استرجاع المعلومات ومعالجة اللغة الطبيعية وإدارة المعلومات.

## محلل البيانات

محلل البيانات هو المسؤول عن جمع البيانات وتنظيفها وتحليلها باستخدام الأساليب الإحصائية والحسابية .يستخدم أدوات وتقنيات مختلفة لاستخراج الرؤى من البيانات وإيصال نتائجها إلى أصحاب المصلحة .تتمثل المهمة الرئيسية لمحلل البيانات في دعم

اتخاذ القرارات التي تعتمد على البيانات داخل المؤسسة من خلال تقديم رؤى حول مؤشرات واتجاهات الأداء الرئيسية. تشمل مسؤوليات محلل البيانات ما يلي :

1. جمع البيانات من مصادر مختلفة: محللو البيانات مسؤولون عن جمع البيانات من مجموعة متنوعة من المصادر، بما في ذلك قواعد البيانات وجداول البيانات وواجهات برمجة التطبيقات. يجب أن يكونوا بارعين في SQL وتقنيات الوصول إلى البيانات الأخرى حتى يتمكنوا من استخراج البيانات التي يحتاجون إليها .
2. تنظيف البيانات وتحويلها: يجب على محلي البيانات تنظيف وتحويل البيانات التي يجمعونها للتأكد من أنها دقيقة ومتسقة. قد يشمل ذلك إزالة القيم المتطرفة وملء القيم المفقودة وتحويل البيانات إلى تنسيق يمكن تحليله بسهولة .
3. تحليل البيانات: يستخدم محللو البيانات الأساليب الإحصائية والحسابية لتحليل البيانات التي جمعوها. يستخدمون أدوات مثل R و Python و Excel لإجراء تحليل إحصائي وإنشاء تصورات وبناء نماذج .
4. توصيل النتائج: يجب أن يكون محللو البيانات قادرين على توصيل نتائجهم إلى أصحاب المصلحة بطريقة واضحة وموجزة. قد يستخدمون أدوات تصور البيانات لإنشاء تقارير ولوحات معلومات توفر رؤى حول مؤشرات الأداء الرئيسية،

ويجب أن يكونوا قادرين على شرح النتائج التي توصلوا إليها بلغة واضحة .

5. دعم القرارات المستندة إلى البيانات: محللو البيانات مسؤولون عن توفير رؤى للمؤسسة يمكنها دعم اتخاذ القرارات المستندة إلى البيانات. إنهم يعملون عن كثب مع فرق أخرى في المؤسسة لتحديد المجالات التي يمكن استخدام البيانات فيها لتحسين الأداء ولتطوير منتجات وحلول تعتمد على البيانات .

للنجاح في دور محلل البيانات، يجب أن يمتلك الفرد المهارات

التالية :

1. مهارات تحليلية قوية: يجب أن يكون محللو البيانات قادرين على تحليل البيانات واستخراج الأفكار التي يمكن استخدامها لاتخاذ قرارات تعتمد على البيانات .
2. مهارات تقنية قوية: يجب أن يكون محللو البيانات بارعين في SQL وتقنيات الوصول إلى البيانات الأخرى، بالإضافة إلى برامج التحليل الإحصائي مثل R و Python و Excel.
3. مهارات اتصال قوية: يجب أن يكون محللو البيانات قادرين على توصيل نتائجهم إلى أصحاب المصلحة بطريقة واضحة وموجزة .
4. مهارات حل المشكلات القوية: يجب أن يكون محللو البيانات قادرين على تحديد وحل المشكلات التي قد تنشأ في عملية جمع البيانات وتحليلها .

5. اهتمام قوي بالتفاصيل: يجب أن يكون محللو البيانات موجّهين نحو التفاصيل للتأكد من أن البيانات التي يجمعونها دقيقة ومتسقة .

6. الإلمام بالتعلم الآلي: قد يُطلب من بعض محلي البيانات أيضًا العمل مع نماذج التعلم الآلي لتطوير التحليلات التنبؤية أو أتمتة عمليات صنع القرار .

يعمل محللو البيانات عادةً على مجموعة متنوعة من المشاريع، مثل :

1. تحليل بيانات العملاء لتحديد الأنماط والاتجاهات التي يمكن أن توجه استراتيجيات التسويق.
2. تحليل بيانات المبيعات لتحديد فرص نمو الإيرادات.
3. تحليل بيانات موقع الويب لفهم سلوك المستخدم وتحديد مجالات التحسين.
4. تحليل البيانات التشغيلية لتحديد مجالات عدم الكفاءة وفرص توفير التكاليف.

بشكل عام، يؤدي محلل البيانات دورًا مهمًا في دعم اتخاذ القرار المستند إلى البيانات داخل المؤسسة. فهم مسؤولون عن جمع مجموعات كبيرة من البيانات وتنظيفها وتحليلها وتفسيرها باستخدام الأساليب الإحصائية والحسابية، كما يتمتعون بالمهارة في استخلاص الرؤى من البيانات التي يمكن أن تساعد في اتخاذ قرارات العمل.

## مهندس البيانات

مهندس البيانات هو المسؤول عن تصميم وبناء وصيانة البنية التحتية والأنظمة التي تدعم عملية علم البيانات. إنهم مسؤولون عن بناء خطوط نقل البيانات والبنى التي تسمح لعلماء البيانات والمحللين بالوصول إلى مجموعات البيانات الضخمة ومعالجتها وتحليلها وصيانتها. تشمل مسؤوليات مهندس البيانات ما يلي:

1. تصميم وتنفيذ أنظمة تخزين ومعالجة البيانات: يجب على مهندسي البيانات تصميم وتنفيذ أنظمة تخزين ومعالجة البيانات التي يمكنها التعامل مع كميات كبيرة من البيانات ودعم احتياجات علماء البيانات والمحللين.
2. بناء خطوط نقل البيانات وصيانتها: مهندسو البيانات مسؤولون عن بناء وصيانة خطوط نقل البيانات التي تنقل البيانات من مصادر مختلفة إلى أنظمة تخزين البيانات ومعالجتها. يجب عليهم التأكد من أن خطوط النقل موثوقة وفعالة وقابلة للتطوير.
3. ضمان جودة البيانات: مهندسو البيانات مسؤولون عن ضمان أن البيانات المخزنة في الأنظمة عالية الجودة ودقيقة ومتسقة. يجب عليهم تنفيذ عمليات التحقق من صحة البيانات وتنظيفها لضمان جودة البيانات.
4. تحسين معالجة البيانات: يجب على مهندسي البيانات تحسين أنظمة معالجة البيانات للأداء وقابلية التوسع. يجب

أن يكونوا قادرين على ضبط وتحسين الأنظمة لضمان معالجة البيانات بسرعة وكفاءة .

5. التعاون مع فرق أخرى: يجب أن يتعاون مهندسو البيانات مع فرق أخرى، مثل علماء ومحليي البيانات، لفهم احتياجات البيانات الخاصة بهم.

ولضمان النجاح في دور مهندس البيانات، يجب أن يمتلك الفرد المهارات التالية :

1. مهارات تقنية قوية: يجب أن يتقن مهندسو البيانات لغات البرمجة مثل Python و Java و SQL ، بالإضافة إلى تقنيات البيانات الضخمة مثل Hadoop و Spark و Kafka. يجب أن يكونوا على دراية أيضًا بأنظمة تخزين البيانات السحابية ومعالجتها مثل AWS و Azure و Google Cloud.
2. مهارات حل المشكلات القوية: يجب أن يكون مهندسو البيانات قادرين على تحديد وحل المشكلات التي قد تنشأ في مسار نقل البيانات وعملية المعالجة .
3. يجب أن يكونوا قادرين على التفكير بشكل نقدي وإبداعي لتطوير الحلول التي تعالج التحديات التقنية المعقدة .
4. الاهتمام الشديد بالتفاصيل: يجب أن يكون مهندسو البيانات موجّهين نحو التفاصيل للتأكد من أن خطوط نقل البيانات والأنظمة التي يصممونها وبنونها موثوقة وفعالة .
5. مهارات الاتصال القوية: يجب أن يكون مهندسو البيانات قادرين على التواصل بشكل فعال مع الفرق الأخرى، مثل

علماء البيانات والمحللين، لفهم احتياجات البيانات الخاصة بهم والتأكد من أن الأنظمة التي يصممونها وبينونها تلي تلك الاحتياجات .

6. الإلمام بإدارة البيانات وأمنها: يجب أن يكون مهندسو البيانات على دراية بأفضل ممارسات إدارة البيانات والأمان وأن يكونوا قادرين على تطبيقها في عملهم لضمان حماية البيانات واستخدامها بطريقة أخلاقية ومتوافقة .

يعمل مهندسو البيانات عادةً على مجموعة متنوعة من المشاريع،  
مثل :

1. بناء خطوط نقل البيانات لجمع ومعالجة البيانات من مصادر مختلفة.
2. تصميم وتنفيذ أنظمة تخزين ومعالجة البيانات لدعم احتياجات علماء ومحليي البيانات.
3. تحسين أنظمة معالجة البيانات للأداء وقابلية التوسع.
4. بناء وصيانة بحيرات البيانات ومخازن البيانات.
5. تطوير وتنفيذ حوكمة وأمن البيانات.

## عالم البيانات

عالم البيانات هو محترف يستخدم الأساليب الإحصائية والحسابية لاستخراج الأفكار والمعرفة من البيانات. يستخدمون أدوات وتقنيات مختلفة لتحليل البيانات وبناء النماذج وتطوير المنتجات والحلول القائمة على البيانات. تتمثل المهمة الرئيسية لعالم البيانات في

استخدام البيانات لدعم اتخاذ القرار وزيادة قيمة الأعمال داخل المؤسسة. تشمل مسؤوليات عالم البيانات ما يلي:

1. جمع البيانات وتنظيفها: علماء البيانات مسؤولون عن جمع البيانات وتنظيفها من مصادر مختلفة، بما في ذلك قواعد البيانات وجداول البيانات وواجهات برمجة التطبيقات. يجب أن يكونوا بارعين في SQL وتقنيات الوصول إلى البيانات الأخرى حتى يتمكنوا من استخراج البيانات التي يحتاجون إليها.
2. تحليل البيانات: يستخدم علماء البيانات الأساليب الإحصائية والحسابية لتحليل البيانات التي جمعوها. يستخدمون أدوات مثل R و Python و Excel لإجراء تحليل إحصائي وإنشاء تصورات وبناء نماذج.
3. توصيل النتائج: يجب أن يكون علماء البيانات قادرين على توصيل نتائجهم إلى أصحاب المصلحة بطريقة واضحة وموجزة. قد يستخدمون أدوات تصور البيانات لإنشاء تقارير ولوحات معلومات توفر رؤى حول مؤشرات الأداء الرئيسية، ويجب أن يكونوا قادرين على شرح النتائج التي توصلوا إليها بلغة واضحة.
4. دعم القرارات المستندة إلى البيانات: علماء البيانات مسؤولون عن تقديم رؤى للمؤسسة يمكنها دعم اتخاذ القرارات المستندة إلى البيانات. إنهم يعملون عن كثب مع فرق أخرى في المؤسسة لتحديد المجالات التي يمكن استخدام

البيانات فيها لتحسين الأداء ولتطوير منتجات وحلول تعتمد على البيانات .

5. تطوير ونشر النماذج التنبؤية: يستخدم علماء البيانات تقنيات التعلم الآلي لتطوير نماذج تنبؤية يمكن استخدامها لعمل تنبؤات أو أتمتة عمليات صنع القرار .

و للنجاح في دور عالم البيانات، يجب أن يمتلك الفرد المهارات التالية :

1. مهارات تحليلية قوية: يجب أن يكون علماء البيانات قادرين على تحليل البيانات واستخراج الأفكار التي يمكن استخدامها لاتخاذ قرارات تعتمد على البيانات .

2. مهارات تقنية قوية: يجب أن يكون علماء البيانات بارعين في لغات البرمجة مثل Python و R ، بالإضافة إلى مكتبات التعلم الآلي وأطر العمل مثل TensorFlow و scikit-Learn و Keras. يجب أن يكونوا أيضًا على دراية بتقنيات البيانات الضخمة مثل Hadoop و Spark.

3. مهارات حل المشكلات القوية: يجب أن يكون علماء البيانات قادرين على تحديد وحل المشكلات التي قد تنشأ في عملية تحليل البيانات. يجب أن يكونوا قادرين على التفكير بشكل نقدي وإبداعي لتطوير الحلول التي تعالج التحديات التقنية المعقدة .

4. مهارات اتصال قوية: يجب أن يكون علماء البيانات قادرين على التواصل بشكل فعال مع الفرق الأخرى، مثل مهندسي

ومحللي البيانات، لفهم احتياجات البيانات الخاصة بهم والتأكد من أن النماذج والحلول التي يطورونها تلبي تلك الاحتياجات .

5. مهارات تصور البيانات القوية: يجب أن يكون علماء البيانات قادرين على إنشاء تصورات فعالة يمكن استخدامها لإيصال نتائجهم إلى أصحاب المصلحة.

يعمل علماء البيانات عادةً على مجموعة متنوعة من المشاريع،  
مثل :

1. بناء نماذج تنبؤية لدعم اتخاذ القرار وأتمتة العمليات.
2. تحليل بيانات العملاء لتحديد الأنماط والاتجاهات التي يمكن أن توجه استراتيجيات التسويق.
3. تحليل بيانات موقع الويب لفهم سلوك المستخدم وتحديد مجالات التحسين.
4. تحليل البيانات التشغيلية لتحديد مجالات عدم الكفاءة وفرص توفير التكاليف تطوير وتنفيذ منتجات وحلول تعتمد على البيانات لدعم احتياجات العمل .

بشكل عام، يؤدي علماء البيانات دورًا مهمًا في استخدام البيانات لدعم اتخاذ القرار وتعزيز قيمة الأعمال داخل المؤسسة. فهم مسؤولون عن جمع مجموعات كبيرة من البيانات وتنظيفها وتحليلها وتفسيرها باستخدام الأساليب الإحصائية والحسابية، كما يتمتعون بالمهارة في استخلاص الرؤى من البيانات التي يمكن أن تساعد في

اتخاذ قرارات العمل. يستخدمون أيضًا نماذج التعلم الآلي لتطوير التحليلات التنبؤية وأتمتة عمليات صنع القرار.

## الخبير الإحصائي

الإحصائي هو محترف يستخدم الأساليب الإحصائية لاستخراج الأفكار والمعرفة من البيانات. يستخدمون أدوات وتقنيات مختلفة لتحليل البيانات وبناء النماذج وإجراء التنبؤات. تتمثل المهمة الرئيسية للإحصائي في استخدام الأساليب الإحصائية لدعم اتخاذ القرار ودفع قيمة الأعمال داخل المؤسسة. تشمل مسؤوليات الإحصائي ما يلي:

1. جمع البيانات وتنظيفها: الإحصائيون مسؤولون عن جمع البيانات وتنظيفها من مصادر مختلفة، بما في ذلك قواعد البيانات وجدول البيانات والاستطلاعات. يجب أن يكونوا بارعين في البرامج الإحصائية ولغات البرمجة مثل R و SAS حتى يتمكنوا من استخراج البيانات التي يحتاجون إليها ومعالجتها.
2. تحليل البيانات: يستخدم الإحصائيون الأساليب الإحصائية لتحليل البيانات التي جمعوها. يستخدمون أدوات مثل R و SAS و Excel لإجراء تحليل إحصائي وإنشاء تصورات وبناء نماذج.
3. توصيل النتائج: يجب أن يكون الإحصائيون قادرين على توصيل نتائجهم إلى أصحاب المصلحة بطريقة واضحة

وموجزة. قد يستخدمون أدوات تصور البيانات لإنشاء تقارير ولوحات معلومات توفر رؤى حول مؤشرات الأداء الرئيسية، ويجب أن يكونوا قادرين على شرح النتائج التي توصلوا إليها بلغة واضحة .

4. دعم القرارات المستندة إلى البيانات: الإحصائيون مسؤولون عن توفير رؤى للمنظمة يمكنها دعم اتخاذ القرارات التي تعتمد على البيانات. إنهم يعملون عن كثب مع فرق أخرى في المؤسسة لتحديد المجالات التي يمكن استخدام البيانات فيها لتحسين الأداء ولتطوير منتجات وحلول تعتمد على البيانات .
5. تطوير ونشر النماذج الإحصائية: يستخدم الإحصائيون النماذج الإحصائية لعمل التنبؤات ودعم عمليات صنع القرار. يحتاجون أيضاً إلى أن يكونوا قادرين على التحقق من صحة هذه النماذج للتأكد من أنها دقيقة وموثوقة.

و للنجاح في دور الإحصائي، يجب أن يمتلك الفرد المهارات التالية :

1. مهارات تحليلية قوية: يجب أن يكون الإحصائيون قادرين على تحليل البيانات واستخراج الأفكار التي يمكن استخدامها لاتخاذ قرارات تعتمد على البيانات .
2. مهارات إحصائية قوية: يجب أن يكون الإحصائيون بارعين في الأساليب والتقنيات الإحصائية، مثل اختبار الفرضيات وتحليل الانحدار والتصميم التجريبي .

3. مهارات حل المشكلات القوية: يجب أن يكون الإحصائيون قادرين على تحديد وحل المشكلات التي قد تنشأ في عملية تحليل البيانات .
4. يجب أن يكونوا قادرين على التفكير بشكل نقدي وإبداعي لتطوير الحلول التي تعالج التحديات التقنية المعقدة .
5. مهارات اتصال قوية: يجب أن يكون الإحصائيون قادرين على التواصل بشكل فعال مع الفرق الأخرى، مثل مهندسي ومحلي البيانات، لفهم احتياجاتهم من البيانات والتأكد من أن النماذج والحلول التي يطورونها تلي تلك الاحتياجات .
6. مهارات تصور البيانات القوية: يجب أن يكون الإحصائيون قادرين على إنشاء تصورات فعالة يمكن استخدامها لإيصال نتائجهم إلى أصحاب المصلحة .

يعمل الإحصائيون عادةً في مجموعة متنوعة من المشاريع، مثل :

1. إجراء التحليلات الإحصائية لدعم الدراسات والتجارب البحثية.
2. تحليل بيانات المسح لفهم سلوك المستهلك وتفضيلاته.
3. تحليل البيانات المالية لتحديد الأنماط وعمل التنبؤات.
4. تطوير وتنفيذ النماذج الإحصائية لدعم عمليات اتخاذ القرار.
5. تقديم الاستشارات الإحصائية للمنظمات .

بشكل عام، يؤدي الإحصائيون دورًا مهمًا في استخدام الأساليب الإحصائية لدعم اتخاذ القرار ودفعة قيمة الأعمال داخل المؤسسة .

فهم مسؤولون عن جمع مجموعات كبيرة من البيانات وتنظيفها وتحليلها وتفسيرها باستخدام الأساليب الإحصائية، كما يتمتعون بالمهارة في استخلاص الرؤى من البيانات التي يمكن أن تساعد في اتخاذ قرارات العمل. كما أنهم يستخدمون النماذج الإحصائية لعمل التنبؤات ودعم عمليات صنع القرار.

### معماري البيانات

معماري البيانات هو محترف مسؤول عن تصميم وبناء وصيانة بنية البيانات للمؤسسة. إنهم مسؤولون عن إنشاء وتنفيذ نماذج البيانات وأنظمة تخزين البيانات ومعالجتها وسياسات إدارة البيانات التي تدعم احتياجات المنظمة. تشمل مسؤوليات معماري البيانات ما يلي:

1. تصميم نماذج البيانات: معماريو البيانات مسؤولون عن تصميم نماذج البيانات التي تمثل البيانات في المؤسسة. يستخدمون تقنيات النمذجة المختلفة لإنشاء نماذج بيانات منطقية ومادية يمكن استخدامها لتخزين البيانات ومعالجتها.
2. بناء أنظمة تخزين البيانات ومعالجتها: معماريو البيانات مسؤولون عن بناء وتنفيذ أنظمة تخزين ومعالجة البيانات التي يمكنها التعامل مع كميات كبيرة من البيانات ودعم احتياجات المنظمة.
3. تنفيذ سياسات إدارة البيانات: معماريو البيانات مسؤولون عن تنفيذ سياسات إدارة البيانات التي تضمن أن البيانات

في المؤسسة دقيقة ومتسقة ومحمية. يجب عليهم التأكد من استخدام البيانات بطريقة أخلاقية ومتوافقة .

4. تحسين معالجة البيانات: يجب على معماري البيانات تحسين أنظمة معالجة البيانات للأداء وقابلية التوسع. يجب أن يكونوا قادرين على ضبط وتحسين الأنظمة لضمان معالجة البيانات بسرعة وكفاءة.

5. التعاون مع فرق أخرى: يجب أن يتعاون معماريو البيانات مع فرق أخرى، مثل علماء البيانات والمعماريين والمحليلين، لفهم احتياجات البيانات الخاصة بهم والتأكد من أن بنية البيانات تلبى هذه الاحتياجات .

وللنجاح في دور معماري البيانات، يجب أن يمتلك الفرد المهارات التالية :

1. مهارات تقنية قوية: يجب أن يتقن معماريو البيانات لغات البرمجة مثل Python و Java و SQL ، بالإضافة إلى تقنيات البيانات الضخمة مثل Hadoop و Spark و Kafka. يجب أن يكونوا على دراية أيضًا بأنظمة تخزين البيانات السحابية ومعالجتها مثل AWS و Azure و Google Cloud.

2. مهارات تحليلية قوية: يجب أن يكون معماريو البيانات قادرين على تحليل البيانات واستخراج الأفكار التي يمكن استخدامها لاتخاذ قرارات تعتمد على البيانات .

3. مهارات حل المشكلات القوية: يجب أن يكون معماريو البيانات قادرين على تحديد وحل المشكلات التي قد تنشأ في عملية تخزين البيانات ومعالجتها .
4. يجب أن يكونوا قادرين على التفكير بشكل نقدي وإبداعي لتطوير الحلول التي تعالج التحديات التقنية المعقدة .
5. اهتمام قوي بالتفاصيل: يجب أن يكون معماريو البيانات موجّهين نحو التفاصيل لضمان أن تكون أنظمة تخزين البيانات والمعالجة التي يصممونها وبينونها موثوقة وفعالة.
6. مهارات الاتصال القوية: يجب أن يكون معماريو البيانات قادرين على التواصل بشكل فعال مع الفرق الأخرى، مثل علماء البيانات والمحللين، لفهم احتياجات البيانات الخاصة بهم والتأكد من أن الأنظمة التي يصممونها وبينونها تلبي هذه الاحتياجات .

يعمل معماريو البيانات عادةً على مجموعة متنوعة من

المشاريع، مثل :

1. تصميم وتنفيذ معماريات البيانات لدعم احتياجات المنظمة.
2. بناء مستودعات البيانات وبحيرات البيانات لتخزين ومعالجة كميات كبيرة من البيانات.
3. تطوير وتنفيذ سياسات حوكمة البيانات لضمان دقة البيانات واتساقها وأمنها.
4. تحسين أنظمة معالجة البيانات للأداء وقابلية التوسع.

5. تصميم وتنفيذ نماذج البيانات لدعم تحليل البيانات وإعداد التقارير التعاون مع فرق أخرى، مثل علماء ومحللي البيانات، لفهم احتياجات البيانات الخاصة بهم والتأكد من أن بنية البيانات تلي تلك الاحتياجات .

باختصار، يؤدي معماريو البيانات دورًا مهمًا في تصميم وبناء بنية البيانات للمؤسسة. إنهم مسؤولون عن إنشاء وتنفيذ نماذج البيانات وأنظمة تخزين البيانات ومعالجتها وسياسات إدارة البيانات التي تدعم احتياجات المنظمة. يجب أن يكونوا بارعين في لغات البرمجة وتقنيات البيانات الضخمة وأنظمة تخزين البيانات السحابية ومعالجتها حتى يتمكنوا من إنشاء بنية بيانات قوية والحفاظ عليها. كما أنهم يعملون بشكل وثيق مع فرق أخرى لفهم احتياجاتهم من البيانات والتأكد من أن بنية البيانات تلي تلك الاحتياجات.

### مسؤول البيانات

مسؤول البيانات هو الشخص المسؤول عن إدارة وصيانة أنظمة البيانات للمؤسسة. يعمل بشكل وثيق مع فرق أخرى للتأكد من أن البيانات دقيقة ومتسقة وأمنة. إنهم مسؤولون عن مراقبة أنظمة البيانات واستكشاف الأخطاء وإصلاحها وتنفيذ التغييرات والتحديثات حسب الحاجة. تشمل مسؤوليات مسؤول البيانات ما يلي :

1. إدارة أنظمة البيانات: مسؤولو البيانات مسؤولون عن إدارة وصيانة أنظمة البيانات في المؤسسة. يجب عليهم التأكد من أن الأنظمة محدثة وآمنة وتعمل على النحو الأمثل .
2. استكشاف الأخطاء وإصلاحها وحل المشكلات: مسؤولو البيانات مسؤولون عن استكشاف الأخطاء وإصلاحها وحل أي مشكلات قد تنشأ في أنظمة البيانات. يجب أن يكونوا قادرين على تحديد المشكلات وحلها بسرعة وفعالية لتقليل الاضطرابات التي تتعرض لها المنظمة .
3. تنفيذ التغييرات والتحديثات: مسؤولو البيانات مسؤولون عن تنفيذ التغييرات والتحديثات على أنظمة البيانات حسب الحاجة. يجب عليهم التأكد من أن أي تغييرات أو تحديثات يتم اختبارها ونشرها بشكل صحيح لتقليل الاضطرابات التي تتعرض لها المؤسسة .
4. المراقبة والإبلاغ: مسؤولو البيانات مسؤولون عن مراقبة أنظمة البيانات والإبلاغ عن أي مشكلات أو مخاوف إلى الأطراف المناسبة. يجب أن يكونوا قادرين على تحديد مؤشرات الأداء الرئيسية والإبلاغ عنها لضمان أداء أنظمة البيانات على النحو الأمثل .
5. التواصل مع الفرق الأخرى: يجب على مسؤولي البيانات التواصل بشكل فعال مع الفرق الأخرى، مثل مهندسي ومحلي البيانات، لفهم احتياجات البيانات الخاصة بهم والتأكد من أن أنظمة البيانات تلبى هذه الاحتياجات.

و للنجاح في دور مسؤول البيانات، يجب أن يمتلك الفرد المهارات التالية :

1. مهارات تقنية قوية: يجب أن يكون مسؤولو البيانات بارعين في لغات البرمجة مثل SQL ، بالإضافة إلى أنظمة إدارة قواعد البيانات مثل MySQL و Oracle و SQL Server. كما يجب أن يكونوا على دراية بتخزين البيانات ونمذجة البيانات وعمليات ETL (Extract-Transform-Load).
2. مهارات تحليلية قوية: يجب أن يكون مسؤولو البيانات قادرين على تحليل البيانات واستخراج الأفكار التي يمكن استخدامها لاتخاذ قرارات تعتمد على البيانات .
3. مهارات حل المشكلات القوية: يجب أن يكون مسؤولو البيانات قادرين على تحديد وحل المشكلات التي قد تنشأ في أنظمة البيانات. يجب أن يكونوا قادرين على التفكير بشكل نقدي وإبداعي لتطوير الحلول التي تعالج التحديات التقنية المعقدة .
4. الاهتمام الشديد بالتفاصيل: يجب أن يكون مسؤولو البيانات موجّهين نحو التفاصيل للتأكد من أن أنظمة البيانات التي يديرونها وصيانتها موثوقة وفعالة .
5. مهارات اتصال قوية: يجب أن يكون مسؤولو البيانات قادرين على التواصل بشكل فعال مع الفرق الأخرى، مثل مهندسي ومحلي البيانات، لفهم احتياجات البيانات الخاصة بهم والتأكد من أن أنظمة البيانات تلبّي هذه الاحتياجات .

يعمل مسؤولو البيانات عادةً على مجموعة متنوعة من المشاريع، مثل: إدارة وصيانة ومراقبة أنظمة البيانات للمؤسسة لضمان دقة البيانات واتساقها وأمانها.

## محلل الأعمال

محلل الأعمال هو الشخص المسؤول عن تحديد وتحليل احتياجات العمل للمؤسسة. إنهم يعملون بشكل وثيق مع فرق أخرى لفهم متطلبات العمل ولتطوير الحلول التي تدعم أهداف المنظمة. يعمل كجسر بين فرق العمل والفرق الفنية، مما يساعد على مواءمة أهداف العمل مع القدرات الفنية. تشمل مسؤوليات محلل الأعمال ما يلي:

1. تحديد وتحليل احتياجات العمل: محللو الأعمال مسؤولون عن تحديد وتحليل احتياجات العمل للمؤسسة. يستخدمون تقنيات مختلفة، مثل المقابلات والاستطلاعات وورش العمل، لجمع المعلومات حول متطلبات العمل.
2. تطوير الحلول: محللو الأعمال مسؤولون عن تطوير الحلول التي تلبى احتياجات العمل للمؤسسة. يستخدمون أدوات وتقنيات مختلفة، مثل نمذجة العملية وتطوير حالة الاستخدام، لإنشاء حلول تدعم أهداف المنظمة.
3. التواصل مع أصحاب المصلحة: يجب على محللي الأعمال التواصل بشكل فعال مع أصحاب المصلحة لفهم احتياجاتهم والتأكد من أن الحلول التي يطوروها تلبى تلك الاحتياجات.

يجب أن يكونوا قادرين على تقديم نتائجهم وتوصياتهم بطريقة واضحة وموجزة .

4. إدارة نطاق المشروع: محللو الأعمال مسؤولون عن إدارة نطاق المشاريع لضمان تسليم الحلول التي يطورونها في الوقت المحدد وفي حدود الميزانية. يجب أن يكونوا قادرين على تحديد وإدارة مخاطر المشروع والتأكد من بقاء المشروع على المسار الصحيح .

5. دعم تنفيذ الحلول: محللو الأعمال مسؤولون عن دعم تنفيذ الحلول، والعمل عن كثب مع فرق أخرى مثل محللي البيانات ومهندسي البيانات وعلماء البيانات لضمان تنفيذ الحلول بشكل صحيح.

و للنجاح في دور محلل الأعمال، يجب أن يمتلك الفرد المهارات التالية :

1. مهارات تحليلية قوية: يجب أن يكون محللو الأعمال قادرين على تحليل البيانات واستخراج الأفكار التي يمكن استخدامها لاتخاذ قرارات تعتمد على البيانات .
2. مهارات حل المشكلات القوية: يجب أن يكون محللو الأعمال قادرين على تحديد وحل المشكلات التي قد تنشأ في المشروع. يجب أن يكونوا قادرين على التفكير بشكل نقدي وإبداعي لتطوير الحلول التي تعالج تحديات الأعمال المعقدة .
3. مهارات اتصال قوية: يجب أن يكون محللو الأعمال قادرين على التواصل بشكل فعال مع أصحاب المصلحة، التقنيين

وغير التقنيين، لفهم احتياجاتهم ولضمان أن الحلول التي يطورونها تلبي تلك الاحتياجات .

4. مهارات قوية في إدارة المشاريع: يجب أن يكون محللو الأعمال قادرين على إدارة نطاق المشروع والجداول الزمنية والميزانية.
5. معرفة قوية بالمجال: يجب أن يكون لدى محلي الأعمال فهم جيد للصناعة أو القطاع الذي يعملون فيه، مما يمكن أن يساعد على فهم احتياجات العمل للمؤسسة وتطوير حلول مصممة خصيصًا لتلك الاحتياجات .

يعمل محللو الأعمال عادةً على مجموعة متنوعة من المشاريع، مثل: تحديد وتحليل احتياجات العمل للمؤسسة تطوير الحلول التي تدعم أهداف المنظمة إدارة المشاريع لضمان تسليم الحلول في الوقت المحدد وفي حدود الميزانية دعم تنفيذ الحلول، والعمل بشكل وثيق مع فرق أخرى مثل محلي البيانات ومهندسي البيانات وعلماء البيانات .

بشكل عام، يؤدي محللو الأعمال دورًا مهمًا في تحديد وتحليل البيانات لتحقيق جودة في إدارة الأعمال.

### مدير البيانات

مدير البيانات هو محترف مسؤول عن إدارة بيانات المؤسسة وصيانتها. إنهم يعملون بشكل وثيق مع فرق أخرى للتأكد من أن البيانات دقيقة ومتسقة وآمنة. إنهم مسؤولون عن مراقبة أنظمة

البيانات واستكشاف الأخطاء وإصلاحها وتنفيذ التغييرات والتحديثات حسب الحاجة. تشمل مسؤوليات مدير البيانات ما يلي:

1. إدارة أنظمة البيانات: مديرو البيانات مسؤولون عن إدارة وصيانة أنظمة البيانات في المنظمة. يجب عليهم التأكد من أن الأنظمة محدثة وآمنة وتعمل على النحو الأمثل.
2. تطوير سياسات حوكمة البيانات: مديرو البيانات مسؤولون عن تطوير وتنفيذ سياسات حوكمة البيانات التي تضمن أن البيانات في المؤسسة دقيقة ومتسقة ومحمية. يجب عليهم التأكد من استخدام البيانات بطريقة أخلاقية ومتوافقة.
3. نمذجة البيانات: مديرو البيانات مسؤولون عن تصميم وتنفيذ نماذج البيانات المنطقية والمادية التي يمكن استخدامها لتخزين البيانات ومعالجتها.
4. إدارة جودة البيانات: مديرو البيانات مسؤولون عن مراقبة جودة البيانات وتنفيذ العمليات لضمان دقة البيانات واكتمالها واتساقها.
5. التواصل مع الفرق الأخرى: يجب على مديري البيانات التواصل بشكل فعال مع الفرق الأخرى، مثل علماء ومحللي البيانات، لفهم احتياجاتهم من البيانات والتأكد من أن أنظمة البيانات تلبى هذه الاحتياجات.

و للنجاح في دور مدير البيانات، يجب أن يمتلك الفرد المهارات التالية:

1. مهارات تقنية قوية: يجب أن يتقن مديرو البيانات لغات البرمجة مثل SQL ، بالإضافة إلى أنظمة إدارة قواعد البيانات مثل MySQL و Oracle و SQL Server. كما يجب أن يكونوا على دراية بتخزين البيانات ونمذجة البيانات وعمليات ETL (Extract-Transform-Load).
2. مهارات تحليلية قوية: يجب أن يكون مديرو البيانات قادرين على تحليل البيانات واستخراج الرؤى التي يمكن استخدامها لاتخاذ قرارات تعتمد على البيانات .
3. مهارات حل المشكلات القوية: يجب أن يكون مديرو البيانات قادرين على تحديد وحل المشكلات التي قد تنشأ في أنظمة البيانات. يجب أن يكونوا قادرين على التفكير بشكل نقدي وإبداعي لتطوير الحلول التي تعالج التحديات التقنية المعقدة .
4. الاهتمام الشديد بالتفاصيل: يجب أن يكون مديرو البيانات موجهين نحو التفاصيل لضمان أن أنظمة البيانات التي يديرونها وصيانتها موثوقة وفعالة .
5. مهارات الاتصال القوية: يجب أن يكون مديرو البيانات قادرين على التواصل بشكل فعال مع الفرق الأخرى، مثل مهندسي ومحلي البيانات، لفهم احتياجات البيانات الخاصة بهم والتأكد من أن أنظمة البيانات تلبى هذه الاحتياجات .
6. مهارات قوية في إدارة المشاريع: يجب أن يكون مديرو البيانات قادرين على إدارة المشاريع والجداول الزمنية والميزانية.

7. معرفة قوية بالمجال: يجب أن يكون لدى مديري البيانات فهم جيد للصناعة أو القطاع الذي يعملون فيه، مما يمكن أن يساعدهم على فهم احتياجات العمل للمؤسسة وتطوير الحلول المصممة خصيصًا لتلك الاحتياجات .

يعمل مديرو البيانات عادةً على مجموعة متنوعة من المشاريع،  
مثل :

1. إدارة وصيانة أنظمة البيانات للمؤسسة.
2. تطوير وتنفيذ سياسات حوكمة البيانات لضمان دقة البيانات واتساقها وأمنها.
3. إدارة جودة البيانات.
4. تصميم وتنفيذ نماذج البيانات.
5. إدارة المشاريع لضمان تسليم أنظمة البيانات في الوقت المحدد وفي حدود الميزانية.
6. دعم تنفيذ أنظمة البيانات، والعمل بشكل وثيق مع فرق أخرى مثل محلي البيانات ومهندسي البيانات وعلماء البيانات .

باختصار، يؤدي مديرو البيانات دورًا مهمًا في إدارة أنظمة البيانات الخاصة بالمؤسسة وصيانتها. إنهم مسؤولون عن ضمان أن البيانات دقيقة ومتسقة وآمنة. يجب أن يكونوا بارعين في لغات البرمجة وأنظمة إدارة قواعد البيانات حتى يتمكنوا من إنشاء وصيانة أنظمة بيانات قوية. كما أنهم يعملون بشكل وثيق مع فرق أخرى لفهم

احتياجاتهم من البيانات والتأكد من أن أنظمة البيانات تلبى تلك الاحتياجات.

## لغات برمجة البيانات

هناك العديد من لغات البرمجة التي يشيع استخدامها في علم البيانات، بما في ذلك

1. Python - Python: هي لغة برمجة عالية المستوى مستخدمة على نطاق واسع وهي مناسبة بشكل خاص لعلوم البيانات بسبب العديد من المكتبات والأطر القوية. تشمل المكتبات الشهيرة NumPy و Pandas و sci-kit-Learn لمعالجة البيانات والتنظيف والتعلم الآلي.
2. R - R: هي لغة برمجة شائعة بين الإحصائيين، وهي مناسبة تمامًا لتصوير البيانات والنمذجة الإحصائية. تتضمن المكتبات الشائعة ggplot2 و dplyr.
3. SQL - SQL: هي لغة برمجة خاصة بالمجال تستخدم لإدارة قواعد البيانات العلائقية ومعالجتها. يستخدم SQL لاستخراج البيانات والمعالجة المسبقة للبيانات.
4. Julia - Julia: هي لغة برمجة جديدة نسبيًا مصممة للحوسبة الرقمية والعلمية عالية الأداء.
5. SAS - SAS: عبارة عن مجموعة برامج قوية لإدارة البيانات والتحليل الإحصائي. يستخدم على نطاق واسع في الأعمال.

6. MATLAB - MATLAB. هي بيئة حوسبة رقمية ولغة برمجة مع مجموعة واسعة من التطبيقات في تحليل البيانات والتصوير.

تجدر الإشارة أيضًا إلى أن العديد من علماء البيانات يستخدمون لغات برمجة متعددة لأداء مهام مختلفة في نفس المشروع. وفيما يأتي نستعرض بتفصيل أكثر هذه اللغات والأدوات.

### Python

هي لغة برمجة عالية المستوى مستخدمة على نطاق واسع وهي مناسبة بشكل خاص لعلوم البيانات. ترجع شعبية Python في مجتمع علوم البيانات إلى حد كبير إلى بساطتها وقابليتها للقراءة، مما يجعل من السهل على المبتدئين التعلم وعلى المبرمجين ذوي الخبرة فهمها والعمل معها بسرعة.

فضلا عن ذلك، تمتلك Python نظامًا بيئيًا واسعًا من المكتبات والأطر القوية المصممة خصيصًا لمهام علوم البيانات، مثل NumPy وPandas و scikit-Learn. تعد NumPy واحدة من المكتبات الرئيسية في Python المستخدمة على نطاق واسع في علم البيانات.

NumPy هي مكتبة للغة برمجة Python توفر دعمًا لمصفوفات ومصفوفات كبيرة ومتعددة الأبعاد من البيانات الرقمية، بالإضافة إلى مجموعة كبيرة من الوظائف الرياضية للعمل على هذه المصفوفات. تستخدم على نطاق واسع في الحساب العددي ومعالجة البيانات وتنظيفها والجبر الخطي.

Pandas هي مكتبة شائعة أخرى في Python تستخدم على نطاق واسع في علم البيانات Pandas. هي مكتبة برامج مكتوبة للغة برمجة Python لمعالجة البيانات وتحليلها. توفر هياكل البيانات والعمليات لمعالجة الجداول العددية وبيانات السلاسل الزمنية. يستخدم على نطاق واسع لتنظيف البيانات وتحويل البيانات وتحليل البيانات.

Scikit-Learn هي مكتبة أخرى مستخدمة على نطاق واسع في Python للتعلم الآلي. إنها مبنية على NumPy و Pandas وتوفر واجهة متسقة لمختلف خوارزميات التعلم الآلي. إنها أدوات بسيطة وفعالة لاستخراج البيانات وتحليل البيانات. تم تصميمه للتعامل مع مكتبات Python العددية والعلمية NumPy و SciPy. تمتلك Python أيضًا مجموعة واسعة من مكتبات التصور مثل Matplotlib و Seaborn و Plotly و Bokeh. توفر هذه المكتبات أدوات سهلة الاستخدام وفعالة لإنشاء تصورات ثابتة و متحركة و تفاعلية للبيانات. بالإضافة إلى هذه المكتبات، تمتلك Python أيضًا مجموعة واسعة من أطر التعلم العميق مثل TensorFlow و Keras و PyTorch و Caffe التي تُستخدم على نطاق واسع لمهام التعلم الآلي القائمة على الشبكة العصبية.

في الختام، تعد Python لغة برمجة شائعة في مجتمع علوم البيانات بسبب بساطتها وقابليتها للقراءة ومكتباتها وأطر عملها القوية. يستخدم على نطاق واسع لتنظيف البيانات وتحويل البيانات وتحليل البيانات والتعلم الآلي وتصوير البيانات.

هي لغة برمجة شائعة بين الإحصائيين وعلماء البيانات، وهي مناسبة تمامًا لتصور البيانات والنمذجة الإحصائية. ترجع شعبية R في مجتمع علوم البيانات إلى قدراتها القوية في معالجة البيانات والتصور، بالإضافة إلى مجموعتها الكبيرة من المكتبات والأطر المصممة خصيصًا لمهام علوم البيانات.

Tidyverse هي إحدى المكتبات الرئيسية في R المستخدمة على نطاق واسع في علم البيانات Tidyverse. عبارة عن مجموعة من حزم R لمعالجة البيانات والتصور، بما في ذلك ggplot2 و dplyr و readr. و tidyrr إنه مصمم للعمل معًا بسلاسة ولجعل معالجة البيانات والتصور أمرًا سهلًا ومتسقًا.

ggplot2 هي إحدى المكتبات في Tidyverse ، وهي تستخدم على نطاق واسع لتصور البيانات. إنه نظام تخطيط لـ R يعتمد على قواعد الرسومات، والذي يوفر واجهة متسقة لإنشاء مجموعة واسعة من الرسومات الإحصائية. وهي مناسبة بشكل خاص لإنشاء تصورات معقدة ومتعددة الطبقات للبيانات.

dplyr هي مكتبة أخرى في Tidyverse ، وهي تستخدم على نطاق واسع لمعالجة البيانات. توفر مجموعة من الأدوات البسيطة والمتسقة للعمل مع إطارات البيانات. إنها مصممة لتسهيل التعامل مع البيانات، وتوفر مجموعة من العمليات لمعالجة البيانات، مثل التصفية والتحديد والمجموعة. مكتبة أخرى شائعة في R هي علامة

الإقحام، وهي عبارة عن مجموعة من الوظائف التي تحاول تبسيط عملية إنشاء نماذج تنبؤية. توفر واجهة متسقة لمختلف خوارزميات التعلم الآلي وتقنيات المعالجة المسبقة. تستخدم على نطاق واسع لمهام التعلم الآلي مثل التصنيف والانحدار والتكتل .

لدى R أيضًا مجموعة واسعة من المكتبات للنمذجة الإحصائية المتقدمة مثل lme4 و glmmTMB و MCMCpack. توفر هذه المكتبات أدوات قوية لتلائم النماذج الإحصائية المعقدة للبيانات، وتستخدم على نطاق واسع في مجالات مثل علم الأحياء وعلم الاجتماع والاقتصاد .

بالإضافة إلى هذه المكتبات، تمتلك R أيضًا مجموعة واسعة من مكتبات التصور مثل الشبكة، والرسم. توفر هذه المكتبات أدوات سهلة الاستخدام وفعالة لإنشاء تصورات ثابتة ومتحركة وتفاعلية للبيانات .

في الختام، تعد R لغة برمجة شائعة بين الإحصائيين وعلماء البيانات بسبب قدراتها القوية في معالجة البيانات والتصوير، بالإضافة إلى مجموعتها الكبيرة من المكتبات والأطر المصممة خصيصًا لمهام علوم البيانات. يستخدم على نطاق واسع لتصوير البيانات ومعالجة البيانات والنمذجة الإحصائية ومهام التعلم الآلي.

## SQL

لغة الاستعلام الهيكلية، هي لغة برمجة خاصة بالمجال تستخدم لإدارة قواعد البيانات العلائقية ومعالجتها. تستخدم على

نطاق واسع في علم البيانات لأنها أداة قوية لاستخراج البيانات والمعالجة المسبقة للبيانات واستكشاف البيانات. يعد استخراج البيانات أحد الاستخدامات الرئيسية لـ SQL في علم البيانات. يمكن استخدام SQL للاستعلام عن قاعدة بيانات واستخراج البيانات اللازمة لتحليل أو مشروع معين. يمكن استخدام هذه البيانات لتدريب نماذج التعلم الآلي، وبناء التصورات، أو لمهام علوم البيانات الأخرى. تسمح SQL لعلماء البيانات باستخراج كميات كبيرة من البيانات من قواعد البيانات ببضعة أسطر من التعليمات البرمجية، مما يوفر الوقت والجهد. يستخدم SQL أيضًا في المعالجة المسبقة للبيانات. المعالجة المسبقة للبيانات هي عملية تنظيف وتحويل البيانات لجعلها أكثر ملاءمة للتحليل. يمكن استخدام SQL لتصفية البيانات وفرزها وتجميعها لإعدادها للتحليل. يمكن أن يشمل ذلك إزالة القيم الخالية، ومعالجة البيانات المفقودة، وإنشاء أعمدة جديدة مع العمليات الحسابية. هذه الخطوة حاسمة قبل أي نمذجة أو تحليل لأنها تضمن أن البيانات دقيقة وموثوقة. يمكن أيضًا استخدام SQL لاستكشاف البيانات. استكشاف البيانات هو عملية تحليل البيانات لاكتساب فهم أفضل لها. يمكن استخدام SQL لإنشاء جداول محورية وجداول تبادلية لاستكشاف العلاقات بين المتغيرات المختلفة في البيانات. يتيح ذلك لعلماء البيانات تحديد الأنماط والاتجاهات في البيانات، والتي يمكن أن تفيد في الخطوات التالية في التحليل. يستخدم SQL أيضًا في علم البيانات لتخزين البيانات وإدارتها. تُستخدم قواعد بيانات SQL على نطاق واسع

لتخزين كميات كبيرة من البيانات وتسهيل الوصول إليها لعلماء البيانات. تعد قواعد بيانات SQL أيضًا خيارًا رائعًا لتخزين البيانات لأنها قابلة للتطوير بدرجة كبيرة ويمكنها التعامل مع كميات كبيرة من البيانات. فضلًا عن ذلك، يتم استخدام SQL أيضًا لتخزين البيانات وتكامل البيانات وحوكمة البيانات. تخزين البيانات هو عملية جمع وتخزين البيانات من مصادر مختلفة في موقع مركزي. يمكن استخدام SQL لاستخراج البيانات من مصادر مختلفة وتخزينها في مستودع بيانات لمزيد من التحليل. تكامل البيانات هو عملية دمج البيانات من مصادر مختلفة في عرض واحد موحد. يمكن استخدام SQL لاستخراج البيانات من مصادر مختلفة ودمجها في جدول واحد أو عرض للتحليل. حوكمة البيانات هي عملية إدارة البيانات كأصل تنظيمي. يمكن استخدام SQL لتنفيذ سياسات وإجراءات حوكمة البيانات مثل التحكم في الوصول إلى البيانات وإدارة جودة البيانات وتتبع نسب البيانات. في الختام، تعد SQL أداة قوية لمهام علوم البيانات مثل استخراج البيانات والمعالجة المسبقة للبيانات واستكشاف البيانات وإدارة البيانات. يتم استخدامه على نطاق واسع في علم البيانات لأنه أداة متعددة الاستخدامات ومتاحة على نطاق واسع للاستعلام عن البيانات ومعالجتها وهي مدعومة على نطاق واسع من قبل معظم قواعد البيانات العلائقية. يتم استخدامه أيضًا في تخزين البيانات وتكامل البيانات ومهام إدارة البيانات، مما يجعلها أداة متعددة الاستخدامات لإدارة البيانات والتحكم فيها في المؤسسة.

## MatLab

[Matrix Laboratory] بيئة حوسبة رقمية ولغة برمجة مع مجموعة واسعة من التطبيقات في تصور وتحليل البيانات تستخدم على نطاق واسع في علم البيانات بسبب قدراتها الرياضية والرسومية القوية. أحد الاستخدامات الرئيسية لـ MATLAB في علم البيانات هو تصور البيانات. يحتوي MATLAB على مجموعة متنوعة من الوظائف والأدوات المضمنة لإنشاء مخططات ثنائية وثلاثية الأبعاد وخرائط حرارية وأنواع أخرى من التصورات. يتيح ذلك لعلماء البيانات تصور البيانات واستكشافها بسهولة، مما قد يكون مفيداً في تحديد الأنماط والاتجاهات. يستخدم MATLAB أيضاً على نطاق واسع لتحليل البيانات. ويحتوي أيضاً على مجموعة متنوعة من الوظائف المضمنة لأداء العمليات الرياضية، مثل معالجة المصفوفة والجبر الخطي والتحسين. يمكن استخدام هذه الوظائف لمهام مثل المعالجة المسبقة للبيانات واستخراج الميزات والتحليل الإحصائي. يحتوي MATLAB أيضاً على مجموعة كبيرة من صناديق الأدوات المصممة خصيصاً لمهام علوم البيانات، مثل مجموعة أدوات الإحصاء والتعلم الآلي، وصندوق أدوات نظام رؤية الحاسوب، وصندوق أدوات معالجة الإشارات. توفر صناديق الأدوات هذه مجموعة واسعة من الوظائف والخوارزميات المبنية مسبقاً والتي يمكن استخدامها لتحليل البيانات ومهام النمذجة مثل التصنيف والانحدار والتكتل. يستخدم MATLAB أيضاً على نطاق واسع في وضع النماذج الأولية وتطوير نماذج التعلم الآلي. يحتوي على مجموعة متنوعة من المكتبات

وصناديق الأدوات التي توفر وظائف وخوارزميات مسبقة الصنع لمهام مثل التعلم الخاضع للإشراف والتعلم غير الخاضع للإشراف والتعلم العميق ومعالجة اللغة الطبيعية. فضلاً عن ذلك، توفر MATLAB أيضاً بيئة لاختبار ونشر نماذج التعلم الآلي. يسمح لعلماء البيانات باختبار النماذج على مجموعة واسعة من البيانات وضبط المعلمات لتحقيق أفضل أداء. كما يسمح لعلماء البيانات بنشر النماذج في بيئات مختلفة مثل السحابة و سطح المكتب والأنظمة المضمنة. في الختام، MATLAB هي أداة مستخدمة على نطاق واسع في علم البيانات بسبب قدراتها الرياضية والرسومية القوية. لديها مجموعة واسعة من الوظائف والأدوات المضمنة لتصوير البيانات وتحليلها ومهام النمذجة. يحتوي أيضاً على مجموعة كبيرة من صناديق الأدوات المصممة خصيصاً لمهام علوم البيانات. كما يوفر بيئة لاختبار ونشر نماذج التعلم الآلي، مما يجعلها أداة متعددة الاستخدامات لمهام علوم البيانات.

## SAS

[Statistical Analysis System]: SAS هو مجموعة برامج قوية لإدارة البيانات والتحليل الإحصائي. يستخدم على نطاق واسع في الأعمال والصناعة لمهام مثل إدارة البيانات وتصوير البيانات والنمذجة الإحصائية. تستخدم SAS على نطاق واسع في علم البيانات بسبب إدارتها القوية للبيانات وتصوير البيانات وقدرات النمذجة الإحصائية. أحد الاستخدامات الرئيسية لـ SAS في علم البيانات هو إدارة البيانات. توفر SAS مجموعة واسعة من أدوات إدارة البيانات

التي تسمح لعلماء البيانات باستيراد البيانات وتنظيفها وتحويلها بسهولة. يتضمن ذلك أدوات لتكامل البيانات وإدارة جودة البيانات وحوكمة البيانات. يستخدم SAS أيضًا على نطاق واسع لتصوير البيانات. يوفر مجموعة واسعة من أدوات تصور البيانات، مثل SAS GRAPH، التي تسمح لعلماء البيانات بإنشاء وتخصيص تصورات البيانات بسهولة. يتضمن ذلك أدوات لإنشاء مخططات ثنائية وثلاثية الأبعاد وخرائط حرارية وأنواع أخرى من التصورات. يستخدم SAS أيضًا على نطاق واسع للنمذجة الإحصائية. يوفر مجموعة واسعة من أدوات النمذجة الإحصائية، مثل SAS / STAT، التي تسمح لعلماء البيانات بإنشاء النماذج الإحصائية وتخصيصها بسهولة. يتضمن ذلك أدوات لمهام مثل الانحدار والتصنيف وتحليل السلاسل الزمنية. توفر SAS أيضًا مجموعة متنوعة من إمكانيات التعلم الآلي من خلال برنامج SAS / EM (Enterprise Miner)، المصمم لاستخراج البيانات والتعلم الآلي. ساس

## JAVA

هي لغة برمجة شائعة للأغراض العامة تُستخدم على نطاق واسع في علم البيانات لمجموعة متنوعة من المهام. وهي معروفة بفحص النوع القوي والأداء وقابلية التوسع. أحد الاستخدامات الرئيسية لـ Java في علم البيانات هو بناء أنظمة موزعة على نطاق واسع. تعد Java لغة شائعة لبناء أنظمة البيانات الضخمة مثل Apache Hadoop و Apache Spark و Apache Kafka. تم تصميم هذه الأنظمة لمعالجة وتحليل كميات كبيرة من البيانات بشكل متوازٍ

وجافا مناسبة تمامًا لهذه المهمة بسبب أدائها وقابليتها للتوسع. تُستخدم Java أيضًا على نطاق واسع للتعلم الآلي. مكتبة Java Machine Learning Library (Java-ML) هي مكتبة شائعة للتعلم الآلي في Java. يوفر مجموعة واسعة من الخوارزميات لمهام مثل التصنيف والتجميع والانحدار. فضلًا عن ذلك، فإنها توفر أيضًا واجهة سهلة الاستخدام لتحميل البيانات ومعالجتها. تُستخدم Java أيضًا على نطاق واسع لمعالجة اللغة الطبيعية (NLP) مجموعة أدوات اللغة الطبيعية (NLTK) هي مكتبة مشهورة للغة البرمجة اللغوية العصبية في جافا. توفر مجموعة واسعة من الأدوات للمهام مثل الترميز، والاشتقاق، وعلامات جزء من الكلام. تُستخدم Java أيضًا على نطاق واسع لتصوير البيانات. توفر مكتبات مثل JFreeChart و Jz3d و JavaFX أدوات قوية لإنشاء تصورات ثابتة و متحركة وتفاعلية للبيانات. توفر هذه المكتبات أدوات فعالة وسهلة الاستخدام لإنشاء المخططات والرسوم البيانية وأنواع أخرى من المرئيات. فضلًا عن ذلك، توفر Java أيضًا مجموعة واسعة من المكتبات والأطر لتطوير الويب، مثل Spring and Struts، والتي يمكن استخدامها لبناء تطبيقات الويب لمهام علوم البيانات. في الختام، تعد Java لغة برمجة شائعة للأغراض العامة تُستخدم على نطاق واسع في علم البيانات لمجموعة متنوعة من المهام مثل أنظمة البيانات الضخمة والتعلم الآلي ومعالجة اللغة الطبيعية وتصوير البيانات وتطوير الويب. إن فحص النوع القوي والأداء وقابلية التوسع يجعله مناسبًا تمامًا لهذه المهام، ولديه مجموعة واسعة من المكتبات والأطر المصممة خصيصًا.

### اسئلة الفصل الخامس

1. تكلم بنقاط عن مهام ومسؤوليات محلل البيانات.؟
2. تكلم بنقاط عن مهام ومسؤوليات مهندس البيانات.؟
3. تكلم بنقاط عن مهام ومسؤوليات عالم البيانات.؟
4. تكلم بنقاط عن مهام ومسؤوليات مدير البيانات.؟
5. تكلم بنقاط عن مهام ومسؤوليات معماري البيانات.؟
6. تكلم بنقاط عن مهام ومسؤوليات مسؤول البيانات.؟
7. عرف لغة البرمجة بايثون. وعرف باهم مكتباتها.
8. عرف بلغة R. موضحاً اهم مكتباتها.؟
9. هناك العديد من لغات البرمجة التي يشيع استخدامها في علم البيانات. عرف بكل منها باختصار.
10. قارن بين مسؤول البيانات و مدير البيانات على اساس المهام والمسؤوليات.

## مشاريع بحوث ودراسات علمية مقترحة :

1. تقييم فاعلية أنظمة استرجاع المعلومات - مشروع يقارن أداء أنظمة استرجاع المعلومات المختلفة باستخدام مقاييس التقييم القياسية .
2. نمذجة موضوعات المقالات العلمية - مشروع يستخدم تقنيات نمذجة الموضوعات لتحديد الموضوعات الرئيسية التي تمت مناقشتها في مجموعة من المقالات العلمية .
3. أنظمة التوصية لموارد المكتبة - مشروع يطور خوارزمية التخصيص لاقتراح موارد المكتبة للمستفيدين بناءً على استخدامهم وتفضيلاتهم السابقة .
4. الحفظ الرقمي لمواد التراث الثقافي - مشروع يستكشف استخدام تقنيات علم البيانات للحفاظ على مواد التراث الثقافي الرقمي مثل الصور والمخطوطات والأفلام .
5. تصور بيانات إحصائيات استخدام المكتبات - مشروع يستخدم تقنيات تصور البيانات لتحليل وتقديم إحصائيات استخدام المكتبة للإبلاغ عن قرارات تطوير المجموعات ووضع الميزانية .
6. أتمتة وضع العلامات وفهرسة موارد المكتبة - مشروع يستخدم خوارزميات التعلم الآلي لأتمتة عملية وضع علامات على موارد المكتبة وفهرستها .
7. تصنيف الأوراق البحثية - مشروع يستخدم تقنيات التعلم الآلي لتصنيف الأوراق البحثية بناءً على محتواها ومجال موضوعها" .
8. اكتشاف الانحرافات في المقاييس العلمية - مشروع يستخدم تقنيات إحصائية لتحديد الأنماط غير العادية أو الانحرافات في المقاييس العلمية مثل عدد الاقتباسات ومؤشر h .
9. تحليل بيانات الوسائط الاجتماعية لتسويق المكتبات" - مشروع يستخدم معالجة اللغة الطبيعية وتحليل المشاعر لتحليل بيانات الوسائط الاجتماعية لإبلاغ استراتيجيات تسويق المكتبات .

10. تحليل سلوك المستخدم" - مشروع يستخدم تقنيات تحليل البيانات لفهم سلوك المستخدمين وتفضيلاتهم وأنماط استخدامهم.
11. توقع المورد التالي الذي سيتم استعارته - مشروع يستخدم تقنيات التعلم الآلي للتنبؤ بالمورد التالي الذي من المحتمل أن يستعيه المستخدم بناءً على سلوكه السابق.
12. إدارة البيانات في المكتبات - مشروع يدرس استخدام ممارسات إدارة البيانات في المكتبات لضمان أمن وسلامة وتوافر بيانات المكتبة .
13. الفهرسة الآلية وإنشاء البيانات الوصفية - مشروع يستخدم تقنيات معالجة اللغة الطبيعية والتعلم الآلي لأتمتة عملية إنشاء الفهرسة والبيانات الوصفية لموارد المكتبة .
14. تحليل بيانات استخدام الكتاب الإلكتروني - مشروع يستخدم تقنيات تحليل البيانات لفهم نمط استخدام الكتب الإلكترونية في المكتبة ولإبلاغ قرارات تطوير المجموعة .
15. تحسين البحث في موارد المكتبة - مشروع يستخدم تقنيات علوم البيانات لتحسين خوارزمية البحث لموارد المكتبة لتحسين تجربة البحث لدى المستخدم .
16. تقييم مجموعات المكتبة - مشروع يستخدم تقنيات تحليل البيانات لتقييم مجموعات المكتبة وتحديد الموارد غير المستغلة أو التي لا يتم استخدامها على الإطلاق.
17. تحليل بيانات خدمة الإعارة بين المكتبات - مشروع يستخدم تقنيات تحليل البيانات لفهم أنماط استخدام خدمة الإعارة بين المكتبات ولإبلاغ عملية صنع القرار بشأن تقاسم الموارد واتفاقيات الشراكة.

## المراجع

1. "The Hundred-Page Machine Learning Book" by Andriy Burkov (2019)
2. "Python Machine Learning" by Sebastian Raschka and Vahid Mirjalili (2015)
3. "Deep Learning" by Yoshua Bengio, Ian Goodfellow and Aaron Courville (2016)
4. "Data Science from Scratch" by O'Reilly Media Inc (2015)
5. "Data-Driven: Creating a Data Culture" by Hilary Mason and DJ Patil (2016)
6. "Big Data: A Revolution That Will Transform How We Live, Work, and Think" by Viktor Mayer-Schönberger (2013)
7. "Data Science for Business" by Foster Provost and Tom Fawcett (2013)
8. "Applied Predictive Modeling" by Max Kuhn and Kjell Johnson (2013)
9. "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling" by Ralph Kimball and Margy Ross (2002)
10. "Data Mining: Concepts and Techniques" by Jiawei Han, Micheline Kamber, and Jian Pei (2011)

11. "Machine Learning: A Probabilistic Perspective" by Kevin P. Murphy (2012)
12. "Introduction to Machine Learning" by Alpaydin E. (2010)
13. "Data Science Handbook" by Carl Shan, William Chen, and Henry Wang (2018)
14. "Data Science for Dummies" by Lillian Pierson (2015)
15. "Data Science for Business" by Foster Provost and Tom Fawcett (2013)
16. "Data Wrangling with Python" by Jacqueline Kazil and Katharine Jarmul (2018)
17. "Data Science with R" by Hadley Wickham and Garrett Grolemund (2017)
18. "Data-Driven: Creating a Data Culture" by Hilary Mason and DJ Patil (2016)
19. "Data Science from Scratch" by O'Reilly Media Inc (2015)
20. "Python for Data Science Handbook" by Jake VanderPlas (2016)
21. "Data Science for Business" by Foster Provost and Tom Fawcett (2013)
22. "Data Science for Dummies" by Lillian Pierson (2015)
23. "Data Science from Scratch" by O'Reilly Media Inc (2015)