المحور الثالث:

تنقيب البيانات

Data mining

محاضرة المحور الثالث: تنقيب البيانات (الأسبوع 4)

تمهيد:

من خلال هذه المحاضرة سوف يتم التعرف على مفهوم وأهمية تنقيب البيانات، تقنيات تنقيب بيانات، أدوات تنقيب بيانات، تطبيقات تنقيب بيانات، تحديات تنقيب بيانات، وأخيرا أمثلة عملية عن تنقيب البيانات.

أهداف محاضرة المحور الثالث: بعد إكمال المحاضرة على الطالب فهم ما يلي:

- الاطلاع على مفهوم وأهمية تنقيب البيانات.
- التعرف على تقنيات تنقيب بيانات.
- استيعاب أدوات تنقيب بيانات.
- فهم تطبيقات تنقيب بيانات.
- التعرف على تحديات تنقيب بيانات.
- التعرف على أمثلة عملية عن تنقيب البيانات.

أولا: مفهوم وأهمية تنقيب البيانات

1- مفهوم تنقيب البيانات:

توجد العديد من التعريفات التي تحاول حصر مفهومه. وفيما يلي بعض التعريفات:

- "تنقيب البيانات هو عملية اكتشاف المعرفة من خلال تحليل كميات كبيرة من البيانات باستخدام تقنيات إحصائية وخوارزميات تعلم آلي."
- "تنقيب البيانات هو استخراج أنماط مفيدة ومخبأة في مجموعات البيانات الكبيرة باستخدام أدوات تحليلية متقدمة."
- "تنقيب البيانات هو عملية تحويل البيانات الخام إلى معلومات قابلة للاستخدام من خلال اكتشاف الأنماط والاتجاهات المخفية."
- "تنقيب البيانات هو استخدام الخوارزميات الإحصائية والتعلم الآلي لاستخراج المعرفة من البيانات الضخمة والمتنوعة."
- "تنقيب البيانات هو عملية استكشافية تهدف إلى اكتشاف الأنماط غير المعروفة سابقًا في البيانات لتحسين عمليات صنع القرار."
- "تنقيب البيانات هو تحليل البيانات الضخمة باستخدام تقنيات متقدمة لاكتشاف الأنماط والعلاقات التي تساعد في تحسين الأداء التنظيمي."
- "تنقيب البيانات هو عملية تحليلية تهدف إلى استخراج المعلومات المفيدة من البيانات الخام لدعم اتخاذ القرارات الاستراتيجية."
- "تنقيب البيانات هو تطبيق تقنيات الذكاء الاصطناعي والتعلم الآلي لتحليل البيانات واستخراج الأنماط التي تساعد في فهم الظواهر المعقدة."
- "تنقيب البيانات هو عملية تحليل البيانات الضخمة باستخدام أدوات إحصائية وخوارزميات لاكتشاف الأنماط التي يمكن أن تسهم في تحسين العمليات التجارية."

من خلال ما سبق يمكن تقديم التعريف الشامل التالي: **"تنقيب البيانات هو عملية تحليلية تهدف إلى استكشاف وتحليل كميات ضخمة من البيانات بهدف اكتشاف أنماط مخفية، علاقات، واتجاهات ذات قيمة مضافة، يتم باستخدام تقنيات متقدمة من التعلم الآلي، الإحصاء، والذكاء الاصطناعي لفهم الظواهر المعقدة وتحسين اتخاذ القرار."**

2- أهمية تنقيب البيانات:

لتنقيب البيانات أهمية بالغة للمنظمة في إطار ذكاء الأعمال، ويمكن توضيح أهم النقاط فيما يلي:

أ- دعم اتخاذ القرارات الاستراتيجية: يساعد تنقيب البيانات على تحويل كميات هائلة من البيانات إلى رؤى قابلة للتنفيذ، مما يدعم صناع القرار في اتخاذ قرارات استراتيجية مدفوعة بالبيانات بدلاً من الاعتماد على الحدس.

ب- تحسين فهم سلوك العملاء: باستخدام تقنيات تنقيب البيانات، يمكن للمنظمات تحليل الأنماط السلوكية للعملاء لتقديم توصيات شخصية وتحسين استراتيجيات التسويق بناءً على احتياجات العملاء الفعلية.

ج- الكشف عن القيم الشاذة والمخاطر المحتملة: يُستخدم تنقيب البيانات لاكتشاف القيم الشاذة في العمليات أو المعاملات المالية، مما يساعد في تحديد المخاطر مثل الغش أو الاختراقات الأمنية قبل أن تسبب أضرارًا كبيرة.

د- تحسين كفاءة العمليات التشغيلية: من خلال تحليل الأنماط في البيانات التشغيلية، يمكن لتنقيب البيانات تعزيز كفاءة العمليات الداخلية مثل إدارة المخزون، التوريد، والإنتاج، مما يؤدي إلى تقليل التكاليف وزيادة الأرباح.

هـ- التنبؤ بالاتجاهات المستقبلية: يتيح تنقيب البيانات إنشاء نماذج تنبؤية تعتمد على البيانات التاريخية لتحديد الاتجاهات المستقبلية مثل الطلب على المنتجات أو حركة السوق، مما يساعد المنظمات على الاستعداد بشكل أفضل.

و- تحسين استراتيجيات التسعير: باستخدام تنقيب البيانات، يمكن للمنظمات تحليل عوامل مثل الطلب، المنافسة، والتفضيلات الخاصة بالعملاء لتحديد أسعار مناسبة تزيد من الإيرادات وتقلل من الخسائر.

ز- تعزيز ولاء العملاء: من خلال تقديم تجارب مخصصة لكل عميل بناءً على سلوكه السابق، يمكن لتنقيب البيانات زيادة رضا العملاء وتعزيز ولائهم للعلامة التجارية.

ح- دعم الابتكار وتطوير المنتجات: يوفر تنقيب البيانات رؤى جديدة حول احتياجات السوق وسلوكيات العملاء، مما يساعد المنظمات على تطوير منتجات وخدمات مبتكرة تلبي هذه الاحتياجات بشكل أفضل.

ط- تحسين إدارة المخاطر: يُستخدم تنقيب البيانات لتقييم المخاطر المتعلقة بالاستثمارات، القروض، أو العمليات المالية الأخرى، مما يساعد المنظمات على اتخاذ قرارات أكثر أمانًا وتحقيق استقرار مالي.

ي- تقليل التكاليف وزيادة الأرباح: من خلال تحسين العمليات، تقليل الهدر، وزيادة الكفاءة التشغيلية، يمكن لتنقيب البيانات المساعدة في خفض التكاليف العامة وزيادة الأرباح على المدى الطويل.

ثانيا: تقنيات تنقيب بيانات

تُعد تقنيات تنقيب البيانات أدوات أساسية لاستخراج الأنماط المخفية والعلاقات من كميات كبيرة من البيانات. هذه التقنيات تستخدم خوارزميات متقدمة مثل التعلم الآلي والإحصاء لتحويل البيانات الخام إلى رؤى قابلة للتنفيذ، مما يساعد المنظمات على اتخاذ قرارات استراتيجية مدفوعة بالبيانات، وفيما يلي أهم التقنيات المستخدمة في تنقيب البيانات:

1- التصنيف:(Classification)

تقسيم البيانات إلى فئات محددة مسبقًا، أمثلة: خوارزمية شجرة القرار، خوارزمية (Support Vector Machine) .

2- التجميع:(Clustering)

تجميع البيانات المتشابهة معًا دون وجود فئات محددة مسبقًا، أمثلة: خوارزمية K-Means ، خوارزمية DBSCAN .

3- الارتباط:(Association Rule Learning)

اكتشاف العلاقات بين المتغيرات في البيانات، أمثلة: خوارزمية Apriori .

4- الانحدار:(Regression)

التنبؤ بقيم متغيرة بناءً على متغيرات أخرى، أمثلة: الانحدار الخطي(Linear Regression) .

5- التنبؤ:(Prediction)

استخدام البيانات التاريخية للتنبؤ بالأحداث المستقبلية، أمثلة: الشبكات العصبية (Neural Networks) .

6- تحليل الانحرافات:(Anomaly Detection)

تحديد البيانات التي تختلف عن النمط العام، أمثلة: خوارزميات اكتشاف القيم الشاذة.

7- تحليل التسلسل:(Sequence Analysis)

تحليل الأنماط الزمنية في البيانات، أمثلة: تحليل سلوك العملاء بمرور الوقت.

8- تحليل النصوص:(Text Mining)

استخراج المعلومات المفيدة من النصوص، أمثلة: تحليل المشاعر(Sentiment Analysis) ، استخراج الكلمات الرئيسية.

9- تحليل الشبكات:(Network Analysis)

دراسة العلاقات بين الكيانات في الشبكات، أمثلة: تحليل شبكات التواصل الاجتماعي.

10- التعلم العميق:(Deep Learning)

استخدام الشبكات العصبية العميقة لتحليل البيانات المعقدة، أمثلة: التعرف على الصور، معالجة اللغة الطبيعية.

ثالثا: أدوات تنقيب بيانات

هناك العديد من الأدوات التي تستخدم في تنقيب البيانات، وفيما يلي أهم هذه الأدوات مصنفة حسب دورها:

1- أدوات التحليل الإحصائي والتنبؤي:

- IBM SPSS Modeler: أداة قوية لتحليل البيانات باستخدام خوارزميات إحصائية وتنبؤية، تساعد في بناء نماذج تحليلية معقدة.

- SAS Data Mining: يقدم حلولًا متقدمة للتحليل الإحصائي والاستخباراتي، ويستخدم على نطاق واسع في المؤسسات المالية والصحية.

2- أدوات السحب والإفلات للمستخدمين غير التقنيين:

- RapidMiner: أداة سهلة الاستخدام تعتمد على السحب والإفلات، تتيح بناء نماذج تحليلية بدون الحاجة إلى كتابة التعليمات البرمجية.

- KNIME: أداة مجانية تعتمد على السحب والإفلات، وتتميز بالمرونة العالية وقدرتها على التعامل مع بيانات كبيرة.

3- أدوات التكامل مع قواعد البيانات:

- Microsoft SQL Server Analysis Services (SSAS): جزء من مجموعة SQL Server ، يوفر أدوات لتحليل البيانات وإنشاء نماذج تنبؤية داخل قواعد البيانات.

- Oracle Data Mining: يقدم حلولًا متكاملة لتحليل البيانات داخل قواعد بيانات Oracle، مع دعم لخوارزميات مثل التصنيف والتنبؤ.

4- أدوات التنقيب المفتوحة المصدر للأبحاث الأكاديمية:

- WEKA: أداة مفتوحة المصدر مخصصة للأبحاث الأكاديمية والتعليم، تستخدم مثلا في استخراج المعلومات من المقالات والوثائق، تحتوي على مجموعة كبيرة من الخوارزميات لتحليل البيانات وتنقيبها.

- Apache Mahout: كلمة Mahout هي كلمة هندية معناها سائس أو مدرب الفيلة، لذلك معنى اسم هذه الأداة هو مدرب أنظمة التعلم الآلي لمعالجة البيانات الضخمة، فهي إطار عمل مفتوح المصدر لتطوير خوارزميات التعلم الآلي وتنقيب البيانات، يعتمد على نظام Hadoop لتوفير قدرات معالجة البيانات الضخمة.

5- أدوات إعداد وتنظيف البيانات:

- Tableau Prep: أداة تساعد في إعداد البيانات وتنظيفها قبل استخدامها في عمليات التنقيب، وتتميز بواجهة بصرية سهلة الاستخدام.

6- لغات البرمجة المتخصصة في تنقيب البيانات:

- PythonوR: لغتان برمجيتان شائعتان لتنقيب البيانات، توفران مكتبات متعددة مثل Scikit-learn و TensorFlow لتحليل البيانات وبناء النماذج.

رابعا: تطبيقات تنقيب بيانات

أصبح تنقيب البيانات أداةً حيويةً في مجال ذكاء الأعمال، فمن خلال تحليل الأنماط والاتجاهات، يُمكن لتنقيب البيانات تحسين العمليات، فهم سلوك العملاء، واكتشاف الفرص الجديدة. فيما يلي نستعرض أبرز تطبيقاته:

1- دعم اتخاذ القرارات الاستراتيجية:

توفر تنقيب البيانات رؤى قيمة حول السوق والمنافسة، مما يدعم صناع القرار في وضع خطط استراتيجية طويلة الأمد، ويتم استخدام النتائج لتحسين الأداء العام للمنظمة.

2- تحسين استراتيجيات التسويق:

يستخدم تنقيب البيانات لتحليل سلوك العملاء وتحديد الفئات المستهدفة بناءً على الأنماط الشرائية، ويتم تقديم توصيات شخصية للعملاء لزيادة معدلات التحويل.

3- إدارة المخاطر المالية:

يُستخدم تنقيب البيانات في تحديد القيم الشاذة واكتشاف العمليات غير الطبيعية مثل الغش أو الاختراقات المالية، ويساعد في تقييم المخاطر المرتبطة بالقروض والاستثمارات.

4- تحسين عمليات الإنتاج والمخزون:

من خلال تحليل بيانات الإنتاج والمبيعات، يمكن للمنظمات تحسين إدارة المخزون وتقليل التكاليف التشغيلية، ويتم استخدام النماذج التنبؤية لتحديد الطلب المستقبلي على المنتجات.

5- تحليل سلة المشتريات:

يساعد تنقيب البيانات في تحديد المنتجات التي يتم شراؤها معًا، مما يتيح للمنظمات تحسين استراتيجيات العرض والترويج، ويُستخدم في تحسين تخطيط المتجر الإلكتروني أو التقليدي.

6- تحسين تجربة العملاء:

باستخدام تقنيات تنقيب البيانات، يمكن للمنظمات فهم احتياجات العملاء وتوفير تجارب مخصصة تزيد من الرضا والولاء، ويتم استخدام البيانات السلوكية لتطوير خدمات ما بعد البيع.

7- تحسين كفاءة سلاسل التوريد:

يتم استخدام تنقيب البيانات لتحليل عمليات التوريد والإنتاج لتحديد النقاط التي تحتاج إلى تحسين، ويؤدي ذلك إلى تقليل الوقت والتكلفة المرتبطة بسلاسل التوريد.

8- التنبؤ بالاتجاهات المستقبلية:

يعتمد تنقيب البيانات على نماذج تنبؤية لتحديد الاتجاهات المستقبلية في السوق أو الطلب على المنتجات، ويساعد هذا التنبؤ في التحضير للمستقبل وتجنب المخاطر المحتملة.

خامسا: تحديات تنقيب بيانات

رغم أهمية تنقيب البيانات في تحويل كميات كبيرة من البيانات إلى رؤى قابلة للتنفيذ، إلا أنه يواجه العديد من التحديات التي قد تعيق فعاليته. فهم هذه التحديات وتطوير استراتيجيات لمواجهتها يعد ضروريًا لتحقيق أقصى استفادة من تقنيات تنقيب البيانات في مجال ذكاء الأعمال، وفيما يلي أهم التحديات:

1- جودة البيانات:

واحدة من أكبر التحديات هي ضعف جودة البيانات بسبب الأخطاء أو القيم المفقودة، وتتطلب عمليات تنظيف وتصحيح البيانات وقتًا وجهدًا كبيرين قبل الشروع في التنقيب.

2- التعقيد التقني:

استخدام تقنيات تنقيب البيانات يتطلب مهارات متقدمة في البرمجة والإحصاء، مما قد يشكل تحديًا للمنظمات التي تفتقر إلى الكفاءات الفنية.

3- حجم البيانات:

مع ظهور البيانات الكبيرة (Big Data)، أصبحت كميات البيانات الهائلة تحديًا في معالجتها واستخراج الأنماط منها، وتتطلب هذه العملية بنية تحتية قوية وموارد حاسوبية كبيرة.

4- الأمان والخصوصية:

حماية البيانات الحساسة أثناء عملية التنقيب هو تحدي كبير، خاصة مع القوانين التنظيمية مثل: GDPR وCCPA.

يجب على المنظمات اتباع أفضل الممارسات لضمان أمان البيانات وحماية خصوصية العملاء.

5- صعوبة فهم النتائج:

نتائج تنقيب البيانات قد تكون معقدة وغير واضحة للمستخدمين غير الفنيين، ويتطلب الأمر تفسيرًا دقيقًا لهذه النتائج لتكون قابلة للتنفيذ.

6- مقاومة التغيير:

قد تواجه المنظمات مقاومة داخلية من الموظفين تجاه تبني تقنيات جديدة لتنقيب البيانات، ويحتاج الأمر إلى استراتيجيات فعالة لتوعية الموظفين بأهمية هذه التقنيات.

7- ارتفاع التكاليف:

تطبيق أنظمة تنقيب البيانات يتطلب استثمارات كبيرة في البنية التحتية والبرمجيات، وقد تكون هذه التكاليف عائقًا أمام المؤسسات الصغيرة والمتوسطة.

سادسا: أمثلة عملية

1- شركة American Express الكشف عن الغش المالي

طبقت American Express تقنيات تنقيب البيانات لاكتشاف العمليات المالية غير الطبيعية والغش المحتمل. تم استخدام خوارزميات تعلم الآلة لتحليل أنماط المعاملات وتحديد القيم الشاذة.

النتائج:

- تقليل حالات الغش بنسبة 40% باستخدام النماذج التحليلية.
- توفير مليارات الدولارات سنويًا نتيجة للكشف المبكر عن العمليات المشبوهة.
- بحلول عام 2023، بلغ عدد العملاء النشطين لدى American Express أكثر من 120مليون عميل.

2- شركة LinkedIn تحسين توصيات الوظائف

استخدمت LinkedIn تقنيات تنقيب البيانات لتحليل سلوك المستخدمين وتقديم توصيات شخصية للوظائف بناءً على المهارات والخبرات السابقة. كما تم تحليل بيانات الشركات لتحديد الفرص الوظيفية المناسبة.

النتائج:

- زيادة نسبة التفاعل مع التوصيات الوظيفية بنسبة 30%.
- تحسين معدلات التوظيف عبر المنصة بنسبة 25%.
- بحلول عام 2023، بلغ عدد المستخدمين النشطين على LinkedIn أكثر من 900 مليون مستخدم.

3- شركة McDonald's تحسين تجربة العملاء في المطاعم

استثمرت McDonald's في تقنيات تنقيب البيانات لتحليل بيانات العملاء مثل الطلبات السابقة والموقع الجغرافي. تم استخدام هذه البيانات لتخصيص العروض وإدارة المخزون بشكل ديناميكي.

النتائج:

- زيادة المبيعات بنسبة 10% بسبب تقديم عروض مخصصة لكل عميل.
- تحسين وقت الخدمة في المطاعم بنسبة 20% باستخدام التنقيب عن الأنماط الزمنية.
- بحلول عام 2023، بلغت الإيرادات السنوية لمكدونالدز 24 مليار دولار.

4- شركة Alibaba تحسين التجارة الإلكترونية

استخدمت Alibaba تقنيات تنقيب البيانات لتحليل سلوك البائعين والمشترين على منصة التجارة الإلكترونية الخاصة بها. تم تطوير نماذج تنبؤية لتحديد المنتجات التي قد يهتم بها العملاء وتحسين تجربة الشراء.

النتائج:

- زيادة مبيعات المنصة بنسبة 30% بسبب التوصيات الشخصية.
- تحسين كفاءة البحث عن المنتجات بنسبة 25% باستخدام الخوارزميات الذكية.
- بحلول عام 2023، بلغت الإيرادات السنوية لعلي بابا 148 مليار دولار.

5- شركة Adidas تخصيص المنتجات للعملاء

استثمرت Adidas في تقنيات تنقيب البيانات لتحليل سلوك العملاء وتفضيلاتهم في شراء المنتجات الرياضية. تم استخدام هذه البيانات لتخصيص التصاميم وتحسين استراتيجيات التسويق.

النتائج:

- زيادة المبيعات بنسبة 18% بسبب تقديم منتجات مخصصة لكل عميل.
- تحسين معدلات الاحتفاظ بالعملاء بنسبة 25% باستخدام التحليلات السلوكية.
- بحلول عام 2023، بلغت الإيرادات السنوية لأديداس 23 مليار يورو.

سابعا: أسئلة عن المحاضرة: للتأكد من الفهم الجيد للمحاضرة حاول الإجابة عن الأسئلة التالية:

- ما المقصود بتنقيب البيانات في مجال ذكاء الأعمال؟
- تكلم عن أهمية تنقيب البيانات بالنسبة للمنظمات؟
- أذكر التقنيات المختلفة في مجال تنقيب البيانات؟
- أذكر أدوات تنقيب البيانات؟
- هناك العديد من التطبيقات التي يمكن أن تستفاد منها المنظمة من تنقيب البيانات، أذكرها؟
- يواجه تنقيب البيانات العديد من الصعوبات، أذكرها؟

ثامنا: أسئلة للبحث: لمزيد من التعمق في موضوع المحاضرة، ابحث عن أجوبة للأسئلة التالية:

- **السؤال:** ما الفرق بين تنقيب البيانات وتحليل البيانات التقليدي؟
- **السؤال:** ما دور الذكاء الاصطناعي والتعلم الآلي في تنقيب البيانات؟
- **السؤال:** كيف يساهم تنقيب البيانات في تحسين تجربة العملاء؟
- **السؤال:** كيف يمكن لتنقيب البيانات تحسين إدارة سلسلة التوريد؟
- **السؤال:** كيف يمكن استخدام تنقيب البيانات في اكتشاف الاحتيال؟