

تنقيب البيانات باستخدام خوارزمية قواعد الارتباط FP-growth لتحسين

ربحية المحلات التجارية

Data mining using the FP-growth association rules algorithm to improve the profitability of shops

مصطفى طويطي

مخبر التطبيقات الكمية والنوعية
للارتقاء الاقتصادي والاجتماعي والبيئي
بالمؤسسات الجزائرية،
جامعة غرداية

touaiti.mustapha@univ-ghardaia.dz

ذهيبة بن عبد الرحمان

مخبر التطبيقات الكمية والنوعية
للارتقاء الاقتصادي والاجتماعي
والبيئي بالمؤسسات الجزائرية،
جامعة غرداية

benabderrahmane.dehiba@univ-ghardaia.dz

كمال موفق

مخبر التطبيقات الكمية والنوعية
للارتقاء الاقتصادي والاجتماعي
والبيئي بالمؤسسات الجزائرية،
جامعة غرداية

mouffok.kamal@univ-ghardaia.dz

المعلومات المقال	الملخص:
تاريخ الارسال: 2024/02/14	تهدف هذه الدراسة إلى إبراز أهم مساهمات تقنيات التنقيب في البيانات بالنسبة للمحلات التجارية، وباستخدام تقنية قواعد الارتباط عن طريق خوارزمية نمو الأنماط FP-growth، خلصنا إلى أن التنقيب في البيانات يساهم في تحسين ربحية المحلات التجارية من خلال استخراج الأنماط المتكررة واكتشاف الترابط ما بين مختلف السلع التي تباع معا، وذلك لتلبية احتياجات الزبائن بالقدر الكافي من السلع والخدمات.
تاريخ القبول: 2024/03/25	
الكلمات المفتاحية: ✓ تنقيب البيانات ✓ قواعد الارتباط ✓ خوارزمية FP-GROWTH	
Article info	Abstract :
Received 14/02/2024	This study aims to highlight the most important contributions of data mining techniques for shops, and by using the technique of association rules through the FP-growth algorithm, we concluded that data mining contributes to improving the profitability of shops by extracting recurring patterns and discovering the correlation between different Goods sold together, in order to adequately meet customer needs for goods and services.
Accepted 25/03/2024	
Keywords: ✓ data mining. ✓ Association Rules. ✓ Algorithm FP-GROWTH.	

. مقدمة:

أدى التحول الرقمي والتكنولوجي الذي شهده العالم في شتى المجالات إلى تراكم قدر هائل من البيانات لا فائدة منها ما لم يتم استغلالها وتحويلها إلى قيمة، من هنا ظهر التنقيب في البيانات لتحويل هذه الأخيرة إلى معلومات مفيدة لاتخاذ قرارات استراتيجية والتي من شأنها أن تقوم بتحسينات أو إضافات للمؤسسات المالكة لهذه البيانات، كما أن عملية التنقيب في البيانات لها آليات وتقنيات تعمل على أساسها وهي تقنيات تنبؤية وأخرى وصفية، إذ تعد تقنية قواعد الارتباط من التقنيات الوصفية التي تقوم على اكتشاف الارتباطات بين القيم لمجموعة الفئات ضمن البيانات الكبيرة، يكون لها عدة استخدامات إلا أن هذه التقنية تستخدم عادة في التسويق.

أدى التحول الرقمي والتكنولوجي الذي شهده العالم في شتى المجالات إلى تراكم قدر هائل من البيانات لا فائدة منها ما لم يتم استغلالها وتحويلها إلى قيمة، من هنا ظهر التنقيب في البيانات لتحويل هذه الأخيرة إلى معلومات مفيدة لاتخاذ قرارات استراتيجية والتي من شأنها أن تقوم بتحسينات أو إضافات للمؤسسات المالكة لهذه البيانات، كما أن عملية التنقيب في البيانات لها آليات وتقنيات تعمل على أساسها وهي تقنيات تنبؤية وأخرى وصفية، إذ تعد تقنية قواعد الارتباط من التقنيات الوصفية التي تقوم على اكتشاف الارتباطات بين القيم لمجموعة الفئات ضمن البيانات الكبيرة، يكون لها عدة استخدامات إلا أن هذه التقنية تستخدم عادة في التسويق.

إشكالية الدراسة:

على ضوء ما تقدم يمكن صياغة إشكالية هذا البحث كالآتي:

كيف تتم الاستفادة المحلات التجارية من تقنيات التنقيب في البيانات لتحقيق رغبات زبائنها بطريقة أفضل ولتحسين ربحيتها؟
الأسئلة الفرعية:

- ما الفائدة من تطبيق خوارزمية FP-growth في المحلات التجارية؟

- وماهي الإجراءات المتخذة لتحسين ربحية المحلات التجارية؟

الفرضيات:

- تطبيق خوارزمية FP-growth مفيد جداً في استخراج الأنماط المتكررة لفهم سلوك المستهلك داخل المحلات التجارية؛

- تتمثل الإجراءات في زيادة توفير السلع المرتبطة معاً، وكذلك في عرض المنتجات.

أهداف الدراسة:

نسعى من خلال هذه الدراسة إلى:

- التعرف على التنقيب في البيانات ومزايا استخدامها؛

- التعرف على خوارزمية FP-growth في استخراجها لقواعد الارتباط؛

- إبراز أهم الدراسات السابقة التي استخدمت خوارزمية FP-growth؛

منهج الدراسة:

اعتمدنا في هذه الدراسة على المنهجين الوصفي والتحليلي.

2- ماهية التنقيب في البيانات :

في عصر المعلومات ومع توفر قدر هائل من البيانات في مختلف الصناعات والمنظمات والتي لا تكون ذات فائدة ما لم يتم تحويلها إلى معلومة قيّمة، ظهرت الحاجة إلى تطوير أدوات تمتاز بالقوة لتحليل البيانات واستخراج المعلومات والمعارف منها.

1.2. تعريف التنقيب في البيانات:

يُعرّف التنقيب عن البيانات على أنه عملية اكتشاف أنماط ومعرفة مثيرة للاهتمام من كميات كبيرة من البيانات (Jiawei Han, 2012, p. 8). ويعرّفه ديفيد هاند وآخرون DAVID Hand et al على أنه "عبارة عن عملية استكشاف مؤتمتة لبيانات مفيدة وتحليلها للعثور على علاقات وتلخيص البيانات بطرائق جديدة تكون مفهومة ومفيدة، بحيث يشار إلى العلاقات والملخصات المستخلصة من عملية التنقيب عن البيانات بالنماذج Models أو الأنماط. أما وفقا لمجموعة غارتنر Gartner Group التنقيب عن البيانات هي عملية اكتشاف روابط جديدة ذات مغزى وأنماط واتجاهات من خلال غرلة كميات كبيرة من البيانات المخزنة في مخازن البيانات باستخدام تقنيات التعرف على الأنماط فضلا عن الاحصاءات والتقنيات الرياضية. أما بالنسبة لايفانجيلوس سيموديس وآخرون Evangelos Simoudis et al التنقيب عن البيانات هو حقل متعدد التخصصات جمع تقنيات من التعلم الآلي والتعرف على الأنماط والإحصاء وقواعد البيانات والإظهار المرئي Visualization لمعالجة مسألة استخراج المعلومات من قواعد البيانات الكبيرة (رفاعي، 2017، صفحة 43 بتصرف).

من خلال التعريفات السابقة نستخلص أن التنقيب في البيانات هو عبارة عن مجموعة من التقنيات لاكتشاف آلي وفعال لما لم يكن معروفاً من قبل، وصالح لأنماط جديدة ومفيدة ومفهومة في قواعد البيانات الكبيرة، بحيث يجب أن تكون الأنماط قابلة للتنفيذ ليتم استخدامها في صنع القرار بالمؤسسة.

2.2. مزايا استخدام التنقيب في البيانات:

توفر عملية اكتشاف المعرفة أو ما يعرف بعملية التنقيب في البيانات عدة مزايا نستعرض أهمها كالاتي (رايح، 2017، صفحة 3 بتصرف):

- تساعد متخذ القرار على تفعيل الترابط بين الأقسام والأعمال المختلفة، لأن ذلك ينعكس على مصلحة العمل في المؤسسة بشكل إيجابي؛
- تسهل التعامل مع تقنيات المعلومات المتطورة، وتساعد على قياس فاعلية نظم المعلومات الفرعية المختلفة وإنتاجيتها، من خلال توفير المعلومات الدقيقة والصحيحة؛
- تساعد على الاستخدام الفعال لمصادر البيانات والموارد المتاحة، وفي التخطيط لتحسين نظم المعلومات المحاسبية والمصرفية المستخدمة وتطويرها؛
- تساعد على الكشف عن قدرة المؤسسة على النمو ومواكبة التطور، وحاجتها إلى تطوير المصادر الفنية والبشرية لنظم المعلومات المستخدمة، ومن ثم فإن المعلومات تساعد على زيادة المعرفة، والحد من البدائل، وهذا الأمر يؤدي إلى التخلص من حالة عدم التأكد التي تتمثل في هذه البدائل؛
- تمكن متخذ القرار من التخطيط لاستمرارية إجراء الفحص وتحديد المشكلة وعناصرها وإتمام الرقابة على ملفات المعلومات والبيانات اللازمة لقراراته.

3. تقنيات التنقيب في البيانات:

تستخدم عملية تنقيب البيانات عدة تقنيات لاستنتاج المعرفة المخبأة داخل العديد من البيانات والمعلومات (رفاعي، 2017، صفحة 44 بتصرف) واختيار التقنية المناسبة يعتمد على طبيعة هذه البيانات وحجمها وهناك نموذجان لتنقيب البيانات (المبارك، 2017، صفحة 37 بتصرف) نستعرضهما كالآتي:

1.3 التنقيب التنبؤي:

وهو النموذج الذي يستخدم النتائج المعروفة المستنبطة من البيانات المختلفة لأجل التنبؤ بقيم لاحقة للبيانات أي يحاول إيجاد أفضل التنبؤات اعتمادا على المعطيات، ويعتمد هذا التنقيب على استخدام المعلومات القديمة لتوقع ما سيحدث في المستقبل وتكون لدى مثل هذه البيانات هدف، ويتضمن هذا النموذج تقنيات وأدوات من أهمها ما يلي (المبارك، 2017، الصفحات 38 - 39 بتصرف):

1.1.3 التصنيف Classification :

يستخدم التصنيف بشكل واسع في حل الكثير من المشكلات خاصة تلك التي تتعلق بالأعمال من خلال تحليل مجموعة من البيانات ووضعها في شكل أصناف أو أقسام يمكن استخدامها فيما بعد لتصنيف البيانات مستقبلا، أي أن التصنيف يقصد به تقسيم البيانات إلى مجاميع يتم تحديدها مسبقا، وللإشارة هناك عدد من الطرق التي يمكن استخدامها في تصنيف البيانات باستخدام خوارزميات مختلفة مثل الخوارزميات الإحصائية والشبكات العصبية والخوارزميات الجينية وطريقة الجار الأقرب وطريقة شجرة القرار، وهي هيكل شجري يقدم مجموعة من القرارات التي تولد قواعد لمجموعة البيانات المصنفة تشتت هذه التقنية وجود حقل قرار يتم تصنيف البيانات بناء عليه.

2.1.3 الانحدار Regression :

هو تقنية تسمح بتحليل البيانات لوصف العلاقة بين متغيرين أو أكثر، ويستخدم في تقدير قيمة أحد المتغيرين إذا عرف المتغير الآخر. ويفترض أن توضع البيانات بنوع معروف من الدوال ومن ثم يتم تحديد أفضل دالة للبيانات المعطاة.

3.1.3 تحليل السلاسل الزمنية Time Series Analysis :

السلاسل الزمنية هي عبارة عن قيم ظاهرة من الظواهر في سلسلة تواريخ متلاحقة، أياما أو أشهر أو سنوات أو حتى فصول والهدف من وراء تحليل سلسلة زمنية وتحليلها وهو التعرف على التغيرات التي طرأت على الظاهرة التي تمثلها في مدة من الزمن، ثم تحليل أسبابها ونتائجها وتحديد اتجاهها حتى يمكن استخدامها للتقدير والتنبؤ بالمستقبل، وللتنبؤ بسلوك مسار الاتجاه العام للظاهرة في المستقبل يجب استخدام أحد منحنيات النمو المعبر عنها بعلاقة رياضية أو بنموذج رياضي وباستخدام هذا النموذج الرياضي يمكن تحديد معدل نمو السلسلة الزمنية موضوع الدراسة وتحديد الاتجاه العام للسلسلة وتحديد القيمة التي يمكن أن يصل إليها هذا الاتجاه في المستقبل، وتحليل السلاسل الزمنية هي تقدير نموذج رياضي يمكنه أن يحاكي تقريبا التدرج التاريخي لتلك الظاهرة بحيث يمكنه أن يقدر بدقة قيم السلسلة الزمنية ويمكن استخدامه للتنبؤ بقيم مستقبلية لهذه الظاهرة. (فتوح و الشفيق جعفر محمود، 2014، الصفحات 6-7)

4.1.3 التنبؤ Prediction :

بعد التنبؤ من الأدوات التي تجذب الانتباه لأنها تتمكن من إعطاء مغزى التوقع الناجح في سياق العمل لذا فإنه يمكن النظر إلى العديد من تطبيقات تنقيب بيانات العالم الحقيقي كأنها تنبؤ بحالة بيانات مستقبلية معتمدة على بيانات سابقة وحالية (المبارك، 2017، صفحة 39).

2.3 التنقيب الوصفي:

يقوم على إيجاد أنماط قابلة للتفسير بواسطة الإنسان والتي يمكن أن تستخدم لوصف البيانات أو ظاهرة معينة، أي تعتمد على إعادة تنظيم البيانات والتنقيب في أعماقها لاستخراج المؤشرات الموجودة فيها (رابح، 2017، صفحة 4 بتصرف) ومن بين الأدوات والتقنيات ما يلي:

1.2.3 العنقدة Clustering :

يتمثل الهدف منها في تحديد الاتجاهات داخل البيانات، ويحاول هذا الأسلوب العثور على مجموعات من العناصر التي توجد عادة معا، وهي عملية تقسيم البيانات إلى مجموعة من الأصناف اعتماداً على تشاركها في الخواص المتشابهة، بحيث يكون التقسيم غير موجه للبيانات، وبالتالي العنقدة هي عكس التصنيف، كما أنها تساعد المستفيد على فهم التركيب الطبيعي للمجموعات من البيانات (سليمان، 2017، صفحة 10 بتصرف).

2.2.3 قواعد الارتباط Association Rules

الهدف من قواعد الارتباط هو اكتشاف العلاقات أو الارتباطات بين القيم المحددة للمتغيرات الفئوية في مجموعات البيانات الكبيرة، تسمح هذه التقنية للمحللين والباحثين للكشف عن الأنماط المخفية في مجموعات البيانات الكبيرة (Gary، Nisbet، و Ken ، 2018، صفحة 124).

3.2.3 اكتشاف التسلسل Sequence Discovery

يستخدم تحليل أو اكتشاف التسلسل لتحديد أنماط متسلسلة في البيانات وهذه الأنماط معتمدة على تسلسل زمن التأثيرات، ويتم في هذه الطريقة البحث لاكتشاف نماذج تحدث بالتسلسل، إذ تكون المدخلات عبارة عن بيانات تشكل مجموعة متسلسلة وكل سلسلة من البيانات هي قائمة منظمة من العمليات أو المصطلحات وعندما تكون العملية عبارة عن مجموعات من المصطلحات لا بد أن يحتسب معها الوقت المصاحب لكل عملية، ولكن مشكلة هذا النموذج تكمن في إيجاد كل النماذج المتسلسلة مع أقل دعم يخصصه المستخدم عندما يكون الدعم لهذا النموذج هو نسبة تسلسل البيانات التي يتضمنها النموذج (المبارك، 2017، صفحة 40 بتصرف).

4.2.3 التلخيص Summarization:

يشير التلخيص إلى أساليب تفتت كتل البيانات الكبيرة إلى مقاييس موجزة، توفر وصفا عاما للمتغيرات وعلاقتها، ومن الأمثلة على أساليب التلخيص نذكر: المتوسطات، والمجاميع، والاحصائيات الوصفية التي تتضمن مقاييس النزعة المركزية مثل المتوسط الحسابي والوسيط والمنوال، ومقاييس التشتت مثل الانحراف المعياري (فتوح و الشفيق جعفر محمود، 2014، صفحة 7).

4- تطبيق خوارزمية قواعد الارتباط FP-Growth في الدراسات الميدانية:

سنستعرض فيما يلي بعض الدراسات الميدانية التي استخدمت خوارزمية قواعد الارتباط FP-Growth.

1.4 دراسة زكريا مهروسة وديما مفتي الشوافعة وحسن فزاز، سنة 2021 بعنوان:

التنقيب عن العناصر المتكررة بالاعتماد على خوارزمية FP-growth واستخدام تقنية MapReduce

هدفت إلى معالجة التحديات المتعلقة بتنقيب مجموعات العناصر المتكررة من البيانات الضخمة، وركزت على دراسة خوارزميات التنقيب عن العناصر المتكررة خاصة التطويرات الموزعة لخوارزمية FP-growth، حيث تم استعراض إيجابيات وسلبيات كل منها واقترح خوارزمية للتنقيب عن العناصر المتكررة بالاعتماد على المخطط الموجه، تم تطبيق الخوارزمية المحسنة من خلال تقنية

MapReduce على Hadoop الموزعة، وذلك بهدف تمكين الخوارزمية المقترحة من التعامل مع البيانات الضخمة بتخفيض درجة تعقيد التنقيب عن العناصر المتكررة وتقليل وقت تنفيذها، حيث تم اختبار هذه الخوارزمية على قواعد بيانات قياسية. أظهرت النتائج أن الخوارزمية المقترحة يمكنها التوسع بشكل جيد وفعال في معالجة مجموعات البيانات الضخمة والقيام بعملية التنقيب فيها مع تحقيق تحسين كبير في معدل استهلاك الذاكرة من حيث تخزين الأنماط المتكررة وتعقيد الزمن.

2.4. دراسة Ali Ikhwan & al، سنة 2018 بعنوان:

A NOVELTY OF DATA MINING FOR PROMOTING EDUCATION BASED ON FP-GROWTH ALGORITHM

هدفت الدراسة إلى إنشاء استراتيجية لدعم تعزيز التعليم بالجامعة من خلال تحديد استراتيجية ترويجية لاستقطاب الطلاب خصوصا في ظل اشتداد المنافسة بين الجامعات، حيث اعتمدت على إنشاء نظام برمجي يمكن من التنفيذ وبشكل فعال لخوارزمية FP-Growth التي تعد إحدى تقنيات التنقيب عن البيانات، وركزت الدراسة على مفهوم تطوير شجرة FP-Tree للحصول على مجموعة العناصر المتكررة، واستخدمت المتغيرات التالية: التعليم الأخير وعنوان المنزل والقسم واختيار Prodi. كشفت نتائج الدراسة أن الخوارزمية تعمل بشكل جيد في البحث عن العناصر المتكررة لشجرة القرار من خلال انشاء قاعدة من بيانات نموذجية للطلاب الجدد، كما أن تحديد متغيرات البيانات بدقة يساهم في صنع الخوارزمية بشكل جيد.

3.4. دراسة Jian Pei و Jiawei Han سنة 2000 بعنوان:

Mining Frequent Patterns by Pattern-Growth: Methodology and Implications

هدفت إلى تقديم منهجية جديدة تسمى "النمو المتكرر للأنماط" للتنقيب بالأنماط المتكررة بدون توليد مرشح، والتي يمكن أن تقلل بشكل كبير من عدد المجموعات المرشحة التي سيتم إنشاؤها وكذلك تقليل حجم قاعدة البيانات لتكون بشكل متكرر لئتم معالجتها ولتؤدي إلى أداء عالٍ. حيث قدمت هذه الدراسة نظرة عامة على هذا النهج وفحصت منهجيته وآثاره على التنقيب في عدة أنواع من الأنماط المتكررة، بما في ذلك الارتباط ومجموعات العناصر المغلقة المتكررة، والأنماط القصوى، والأنماط المتسلسلة، والتنقيب القائم على القيود للأنماط المتكررة، أظهرت الدراسة أن نمو الأنماط المتكرر فعال في التنقيب ضمن قواعد البيانات الكبيرة وقد يؤدي تطويره الإضافي إلى تنقيب قابل للتطوير للعديد من أنواع الأنماط الأخرى أيضاً، وقد أظهرت النتائج أن الخوارزميات المشتقة من منهجية نمط النمو تُعد أكثر كفاءة وقابلية للتطوير من أساليب التنقيب العديدة ذات الأنماط المتكررة الأخرى، كما أنها تقوم على استراتيجية التجزئة من خلال تقسيم قاعدة بيانات كبيرة إلى مجموعة قواعد بيانات أصغر بشكل متكرر وتدرجي، حيث تنخفض الأنماط محل البحث في كل قاعدة بيانات مقابلة متوقعة وذلك بدرجة كبيرة، كما خلصت الدراسة أيضاً إلى أن أحد الاتجاهات يتمثل في تطوير تقنيات تهدف لزيادة تحسين اثبات كفاءة التنقيب، مثل التجسيد والحساب المتزايد لأشجار FP أو قواعد البيانات المتوقعة، التنقيب المتوازي والموزع لنمو الأنماط، وتخفيض تكلفة الاسقاط وغيرها، وخلصت كذلك إلى وجود اتجاه آخر يتمثل في توسيع نطاق تنقيب نمو الأنماط نحو التنقيب في أنماط أكثر تعقيداً مثل التنقيب عن الارتباطات متعددة الأبعاد و/أو متعددة المستويات أو الأنماط المتسلسلة، وارتباطات التنقيب والهياكل السببية، وأنماط المطابقة الجزئية للتنقيب، دورية التنقيب الجزئية، والتصنيف الترابطي، والتنقيب القائم على القيود للأنماط التسلسلية وغيرها من المهام المتعددة الأخرى. وللإشارة فقد تمت دراسة العديد من مهام التنقيب هذه اعتماداً على Apriori. وأوصت الدراسة بضرورة إعادة فحص مهام التنقيب هذه في إطار نمو الأنماط المتكرر لكون ذلك قد يؤدي إلى تطوير خوارزميات تنقيب أكثر كفاءة بالإضافة إلى طرق جديدة محتملة بسبب تقسيم قاعدة البيانات والطبيعة الهيكلية لنهج نمو الأنماط (Han & Jian , 2000, pp. 30-36).

5- خوارزمية نمو الأنماط fp-growth – Growth Approach :A Frequent pattern

1.5 مفهومها:

تعتبر خوارزمية FP- Growth إحدى خوارزميات تقنية قواعد الارتباط وجاءت لتحسين أو بحل بعض المشاكل المتواجدة في خوارزمية Apriori التي تعتمد كلياً على استخراج جميع عناصر الشراء واحداً واحداً ثم معرفة إذا كان كل واحد على حدة متكرراً أم لا وعناصر الشراء المتكررة تحصل في الأجيال القادمة (وهي الأجيال التي تكون 2-2، 3-3، 4-4.... وهكذا)، أي أن الخوارزمية تعتمد اعتماداً كلياً على التكرار من البداية. (Ikhwan, et al., 2018, p. 1661)

ومن بين مشاكل Apriori أنها تعتمد في بحثها على التحليل في المستوى أولاً Breadth-first عملية البحث تكون عن طريق المستوى وهذا يتطلب وقت كبير للبحث وكذلك توليد الأجيال Candidate generation and test حيث عملت خوارزمية FP- Growth بحلها عن طريق البحث في:

- العمق أولاً Depth-first search أي أن عملية البحث تكون في العمق مما يمكن أن نجد النمط الذي يكون له الصدارة في عملية التكرار وهو أطول شيء يبحث حتى النهاية، أي أنه ممكن أن يغطي أكبر عدد من العناصر، أي أن البحث في العمق من خلاله نجد الحلقة الطويلة المتكررة ومعناها أن كل ما تحتها متكرر.

- تجنب خلق الأجيال

الفكرة الأساسية لخوارزمية FP- Growth : البحث عن النمط الطويل أولاً ومن ثم فحص الأنماط التي تكون أصغر منها

2.5 خطوات خوارزمية نمو الأنماط FP- growth :

تعمل الخوارزمية على مرحلتين وهما (مهروسة، ديما مفتي الشوافعة ، و حسن قزاز، 2021، الصفحات 85-86 بتصرف):

▪ مرحلة بناء شجرة النموذج المتكرر FP- Tree:

في هذه المرحلة يتم تتبع الخطوات التالية:

- يتم المسح الأول لقاعدة البيانات لإيجاد مجموعة تتضمن كل عنصر بشكل منفرد أي استخراج العناصر فرداً فرداً مع تكرارها؛
- إقصاء العناصر المتكررة التي يكون تكرارها أقل من العتبة المحددة من طرف المستخدم min_sup وترتيب العناصر المتكررة الباقية حسب قيمة تكرارها ترتيباً تنازلياً؛

- إعادة صياغة الجدول من جديد لعناصر كل مداولة والمصنفة حسب عدد التكرار؛

- المسح الثاني لقاعدة البيانات لبناء شجرة النموذج المتكرر FP- Tree ويتم بناء الشجرة كما يلي:

• إنشاء جذر الشجرة وتسميته ب Null؛

• لكل سجل من قاعدة البيانات يتم ترتيب العناصر المتكررة فيه؛

• إنشاء الفروع لكل سجل مع مراعاة إذا كانت العقدة لها نفس الاسم للعنصر الحالي نزيد الأعداد الموجودة فيها بمقدار واحد وهكذا؛

• ربط جدول الدعم مع الشجرة.

▪ مرحلة استخراج الأنماط المتكررة من الشجرة FP- Tree:

في هذه المرحلة يتم بناء النموذج الشرطي وقاعدة بيانات جزئية تشكل مسارات مرتبطة بشجرة FP- Tree الشرطية وذلك لإيجاد العناصر المتكررة والحصول على النموذج الابتدائي والأنماط المتكررة المولدة من شجرة FP-Tree الشرطية.

نقوم بالمشح أو المرور على شجرة FP- Tree واستخراج المسارات لمعالجتها وذلك حسب جدول الدعم ونبدأ بالمنتج الذي له أقل تكرار ثم الذي يليه إلى غاية إتمام جميع العناصر.

وفي المثال التالي نعمل على توضيح خطوات الخوارزمية مع افتراض عتبة التكرار هي $\min \sup=3$ لنفترض عناصر الشراء لمجموعة من الزبائن لأحد المحلات التجارية كما يلي:

الجدول رقم 01: مداولات قاعدة البيانات

رقم الوصل	عناصر الشراء
T1	حليب، بيض، كسكس، بسكويت، زيت، مشروبات غازية، سكر، جبن
T2	حليب، بيض، كسكس، سكر، سمن، عصير، طماطم مصبرة
T3	حليب، سمن، طماطم مصبرة، ملح، فاصولياء، أرز
T4	كسكس، جبن، سمن، عدس، زبدة
T5	حليب، بيض، كسكس، سكر، جبن، عصير، حمص مصر، خل

المصدر: من اعداد الباحثين

نقوم بترميز عناصر الشراء لتسهيل وتبسيط الدراسة

الجدول رقم 02: ترميز عناصر الشراء

رمز المنتج	اسم المنتج
A	حليب
B	بيض
C	كسكس
D	بسكويت
E	زيت
F	مشروبات غازية
G	سكر
H	جبن
I	سمن
J	عصير
K	طماطم مصبرة
L	ملح
M	فاصولياء
N	أرز
O	عدس
P	زبدة
Q	حمص مصر
R	خل

المصدر: من اعداد الباحثين

الجدول رقم 03: عناصر الشراء بعد الترميز

رقم الوصل	عناصر الشراء
T1	A, B, C, D, E, F, G, H
T2	A, B, C, G, I, J, K
T3	A, I, K, L, M, N
T4	C, H, I, O, P
T5	A, B, C, G, H, J, Q, R

المصدر: من اعداد الباحثين

استخراج العناصر فرداً فرداً مع تكرارها

الجدول رقم 04: عناصر الشراء فرداً فرداً

المنتج	التكرار
A	4
B	3
C	4
D	1
E	1
F	1
G	3
H	3
I	3
J	2
K	2
L	1
M	1
N	1
O	1
P	1
Q	1
R	1

المصدر: من اعداد الباحثين

نقوم بإنشاء جدول العناصر المتكررة فرداً فرداً بعد حذف العناصر التي لها أقل تكرار من العتبة المحددة 3=Min support

الجدول رقم 05: جدول الدعم

المنتج	التكرار
A	4
B	3
C	4
G	3
H	3
I	3

المصدر: من اعداد الباحثين

الجدول رقم 06: ترتيب جدول الدعم ترتيبا تنازليا

المنتج	التكرار
A	4
C	4
B	3
G	3
H	3
I	3

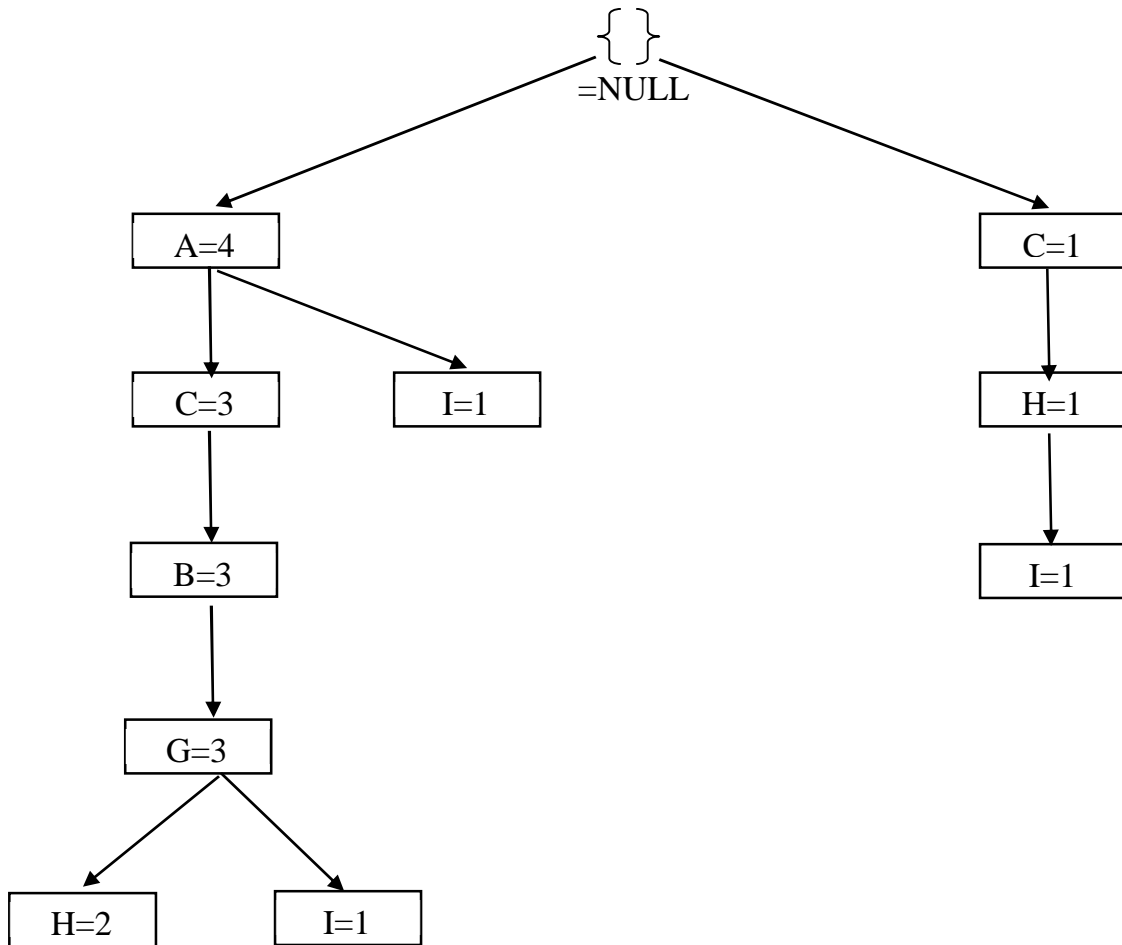
المصدر: من اعداد الباحثين

الجدول رقم 06: العناصر لكل مداولة المصنفة حسب عدد التكرار

رقم الوصل	عناصر الشراء
T1	A,C,B,G,H
T2	A, C,B, G,I
T3	A,I
T4	C,H,I
T5	A,C,B,G,H

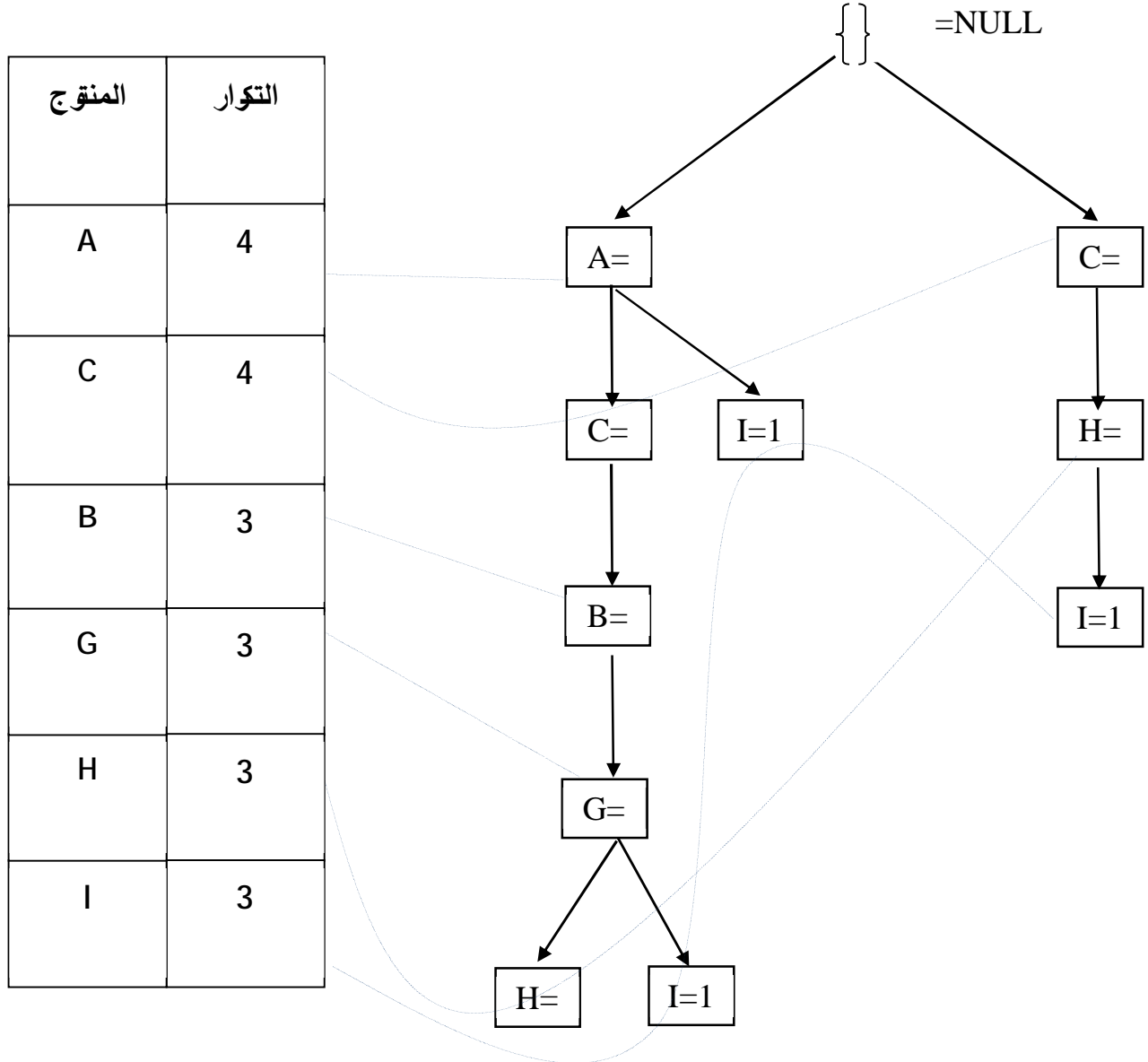
المصدر: من اعداد الباحثين

الشكل رقم 01: بناء الشجرة FP-tree



المصدر: من اعداد الباحثين

الشكل رقم 02: ربط جدول الدعم مع الشجرة FP-tree



المصدر: من اعداد الباحثين

الجدول رقم 07: استخراج الأنماط المتكررة

عناصر الشراء	قاعدة النمط الشرطي	شرطية الشجرة fp-tree	النمط المتكرر
I	{C,H:1}{A:1}{A,C,B,G:1}	A :2, C :2, } {B :1 ,G :1}	العناصر أقل من تكرار العتبة لذا لا تدخل
H	{A:1}{A,C,B,G:2}	{A:3}	{A,H:3}
G	{A,C,B:3}	{A:3,C:3,B:3}	{A,C,B,G:3}{A,C,G:3}{A,B,G:3}{C,B,G:3}{A,G:3}{C,G:3}{B,G:3}
B	{A,C:3}	{A:3,C:3}	{A,C,B:3}
C	{A:3}	{A:3}	{A,C:3}

المصدر: من اعداد الباحثين

ومنه نستخرج مجموعة الأنماط المتكررة كما يلي:

A,C,B,G,H,I,BG,CG,AG,AC,CBG,ABG,ACG,ACB,ACBG

نقوم برفع الترميز على هذه الأنماط وفق الجدول التالي:

رمز المنتج	عناصر الشراء
A	حليب
C	كسكس
B	بيض
G	سكر
H	جن
I	سمن
BG	بيض، سكر
CG	كسكس، سكر
AG	حليب، سكر
AC	حليب، كسكس
CBG	كسكس، بيض، سكر
ABG	حليب، بيض، سكر
ACG	حليب، كسكس، سكر
ACB	حليب، كسكس، بيض
ACBG	حليب، كسكس، بيض، سكر

المصدر: من اعداد الباحثين

من خلال الجدول السابق نلاحظ أن مادة الحليب والكسكس والبيض والسكر لها ارتباط قوي مع بعضها البعض وأن الزبون إذا اشترى الحليب والكسكس اشترى البيض معهما، وإذا اشترى الحليب والكسكس والبيض وجب عليه أن يشتري كذلك السكر وهكذا، وإن أفضل النتائج هي تلك التي تتكون من سلعتين إلى أربع سلع مختلفة التي تباع معا ومن هنا وجب اتخاذ القرار المناسب لتطوير وتحسين المبيعات لكل صنف وذلك بزيادة توفير وبصفة دائمة بعض الأصناف التي تباع معا، وكذلك هندسة عرض المنتجات بما يتوافق مع سلوك المستهلك من خلال هاته النماذج.

في هذا المثال اقتصرنا على 5 زبائن فقط، أما في حالة زبائن أكثر من ذلك فتتراكم البيانات وتصبح هناك بيانات ضخمة ويتطلب في ذلك استعمال برمجيات في تحليل ذلك ومن بين هذه البرمجيات برنامج weka وبرمجة R وبرمجة python.... وغيرها من البرمجيات.

خاتمة

حاولنا في دراستنا إبراز أهم مساهمات تقنيات التنقيب في البيانات بالنسبة للمحلات التجارية، وباستخدام تقنية قواعد

الارتباط عن طريق خوارزمية نمو الأنماط FP-growth، خلصنا إلى النتائج التالية:

- يعتبر تطبيق تقنية قواعد الارتباط أنسب في مجال التسويق، حيث تقوم هذه التقنية بالكشف عن الأنماط المخفية في مجموعات البيانات الكبيرة، من خلال عدة خوارزميات من أهمها خوارزمية نمو الأنماط FP-Growth التي تعمل على استخراج الأنماط المتكررة؛
 - مادة الحليب والكسكس والبيض والسكر لها ارتباط قوي مع بعضها البعض
 - ان الزبون إذا اشترى الحليب والكسكس اشترى البيض معهما
 - اذا اشترى الحليب والكسكس والبيض وجب عليه أن يشتري كذلك السكر
- ومن بين أهم التوصيات:
- الزيادة في توفير بعض الأصناف من السلع التي تباع معا بصفة دائمة؛
 - ضرورة عرض المنتجات داخل المحلات التجارية وفق الأنماط المستخرجة التي تتوافق وسلوك المستهلك.

قائمة المراجع

- البدوي سعد البدوي المبارك، 2017، استخدام تقنيات تنقيب البيانات لإستكشاف أنماط مؤثرات التحصيل الأكاديمي لطلاب المرحلة الثانوية، كلية علوم الحاسوب وتقانة المعلومات، جامعة النيلين، السودان.
- الجيلي كباشي ابراهيم سليمان، 2017، تنقيب بيانات المحلات التجارية بواسطة قاعدة الارتباط (خوارزمية البدهة) دراسة حالة: محلات تجارية بمدينة ام درمان، كلية الدراسات العليا، جامعة السودان للعلوم والتكنولوجيا، السودان.
- رهف حريري رفاعي، 2017، استخدام تقنيات التنقيب عن البيانات في دعم قرارات شركات التأمين، قسم الإحصاء، كلية الاقتصاد، جامعة حلب، سوريا.
- سيف الدين عثمان فتوح، و الشفيق جعفر محمود، 2014، التنقيب في البيانات واتخاذ القرارات (نموذج تطبيقي لخزان خشم القرية)، مجلة النيل الأبيض للدراسات والبحوث، العدد الثالث مارس 2014، 1-17.
- زكرياء مهروسة، ديمما مفتي الشوافعة وحسن قزاز، 2021، التنقيب عن العناصر المتكررة بالاعتماد على تطوير خوارزمية FP-Growth واستخدام تقنية MapReduce، مجلة اتحاد الجامعات العربية للدراسات والبحوث الهندسية، 98-83.
- طلال زغبة، محمد صلاح ، رابح بلعباس، يومي 12 و 13 نوفمبر 2017، التنقيب في البيانات واكتشاف المعرفة كآلية لتحسين جودة القرار في المؤسسة الاقتصادية، الملتقى الدولي الثاني حول التحول الرقمي للمؤسسات و النماذج التنبؤية على المعطيات الكبيرة، جامعة محمد بوضياف المسيلة، الجزائر.
- Han, J., & Jian , P. (2000). Mining Frequent Patterns by Pattern-Growth: Methodology and Implications. ACM SIGKDD, volume 2, Issue 2, pp 30-36.
- Ikhwan, A., Yetri, M., Syahra, Y., halim, J., Siahaan, A. P., Aryza, S., & Yacob, Y. M. (2018). A NOVELTY OF DATA MINING FOR PROMOTING EDUCATION BASED ON FP-GROWTH ALGORITHM. International Journal of Civil Engineering and Technology (IJCIET), volume 9, Issue 7 , pp1660-1669.
- Jiawei Han, Micheline Kamber, Jian Pei, (2012) Data Mining Concepts and Techniques, Third Edition, Morgan Kaufmann, Elsevier Inc, USA.
- Nisbet, R., Gary , M., & Ken , Y. (2018). HANDBOOK OF STATISTICAL ANALYSIS AND DATA MINING APPLICATIONS. London: Academic Press is an imprint of Elsevier.