مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (39) العدد (2) 2017
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (39) No. (2) 2017

تحسين اكتشاف المعرفة باستخدام تطبيقات التنقيب عن البيانات

د. جبر مخائيل حنا[*1]
فاتن فجر الكردي[**]

(تاريخ الإيداع 3 / 1 / 2017. قُبِل للنشر في 30/ 3 / 2017)

□ ملخّص □

تم من خلال هذه الدراسة استخدام التنقيب عن البيانات Data Mining على موقع ويب مختص بتجارة القطع الالكترونية ومعدات الحواسيب من خلال بعض الخدمات مثل خدمة المنتجات التي تم شراؤها معا Frequently Bought Together تعتمد هذه الخدمة على خوارزمية الترابط Association، وتم استخدام هذه الخدمة على صفحة طلب منتج، أي عندما يقوم زبون بطلب منتج معين تظهر له قائمة بالمنتجات الأكثر مبيعاً مع المنتج المطلوب.
الخدمة الثانية هي خدمة تفضيلات المستخدمين Get Personalized Recommendationsوتعتمد هذه الخدمة على خوارزمية العنقدة Clustering، حيث تظهر مجموعة المنتجات المتوقع أن يطلبها الزبون وذلك حسب بيانات الزبون، تظهر هذه الخدمة على صفحة الزبون. وخدمة عرض تنبؤ مبيعات الموقع تعتمد هذه الخدمة على خوارزمية التسلسل الزمنيTime Series، حيث تُظهر هذه الخدمة لمدير النظام توقع أرباح الموقع لفترة مستقبلية قادمة.

الكلمات المفتاحية: تنقيب عن البيانات، خوارزمية العنقدة، خوارزمية الترابط، التسلسل الزمني.

[*] أستاذ – قسم هندسة الحاسبات والتحكم الآلي –كلية الهندسة الميكانيكية والكهربائية–جامعة تشرين–اللاذقية–سورية
[**]طالب دكتوراه – قسم هندسة الحاسبات والتحكم الآلي –كلية الهندسة الميكانيكية والكهربائية–جامعة تشرين–اللاذقية–سورية

مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (39) العدد (2) 2017
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (39) No. (2) 2017

Improve Knowledge Discovery using Data Mining Applications

Dr. Jabr Hanna*
Faten Alkrdy**

(Received 21 / 6 / 2016. Accepted 21 / 1 / 2017)

□ ABSTRACT □

We Use technologies and methodologies of building the web applications to design and build an E-commerce Website. This website is specialist in trading with electronic parts and computer equipment. Then implement Data Mining techniques on this website by adding the following services:Frequently Bought Together Service: This service depends on Association algorithms, it shows the service on request product page, when a customer requests for a particular product. It shows him the list of best-selling products with the desired product. Next, there is Get Personalized Recommendations Service: This service depends on Clustering algorithms, where the expected range of products shows that the customer requested, according to the specifications, this service appear on the customer page. In addition,Will Be Service: This service relies on Time Series algorithms, where this service showing to the system administrator predicted earnings of website for a future period.

Keywords: Data Mining, Clustering algorithm, Association Rules, Time Series

* **Professor in computer and control engineering in Tishreen University, Lattakia, Syria**
*** **Postgraduate student in computer and control engineering in Tishreen University, Lattakia, Syria**

مقدمة:

أدى الانتشار الواسع للتقنيات وسهولة الحصول على المعلومات إلى تضخم حجم البيانات بصورة كبيرة لم يشهدها التاريخ من قبل، م ما أدى إلى زيادة الحاجة إلى تطوير أدوات من أجل تحليل البيانات واستخراج المعلومات والمعارف منها، فالأساليب التقليدية والإحصائية لا تستطيع أن تتعامل مع هذا الكم اليومي الهائل منها، ولذلك تستخدم أدوات ذكية لمعالجة هذه البيانات بهدف الحصول على توصيات ذكية من النظام، يستفيد منها مستخدمي ومدراء النظام بتوظيفها في عمليات اتخاذ القرار.

تقدم تقنيات التنقيب عن البيانات للشركات والمدراء دعماً في عمليات اتخاذ القرار وتحسين طرق إدارة المنشآت، فيقوم البحث على فكرة تصميم موقع ويب للتجارة الالكترونية لنشر وترويج السلع والمنتجات وبيعها على شبكة الإنترنت من خلال عدة خدمات قائمة على تقنيات التنقيب في قواعد البيانات والتي تهدف لاختزال أو ضغط كمية البيانات الكبيرة التي يحتويها الموقع بما ينعكس إيجابا على عمليات البحث التي ينفذها الزبون.

أهمية البحث وأهدافه:

استخدام تقنيات ومنهجيات التنقيب عن البيانات في موقع تجارة إلكترونية وتوظيفها لمساعدة الزبائن في عمليات الشراء، إضافة إلى خدمات خاصة بالمدير كالتنبؤ بنسبة الأرباح لأحد المنتجات خلال فترة زمنية مستقبلية محددة.، وتوفير بيئة تواصل مرنة ومتاحة بشكل دائم لمستخدمي شبكة الإنترنت من خلال الموقع الالكتروني وإمكانية اختيار المنتجات المطلوبة ببعض الخطوات البسيطة. كما تساعد في بناء قاعدة بيانات مناسبة تحوي كافة المعلومات الضرورية لجميع المنتجات، إضافة إلى إمكانية طلب منتجات معينة للشراء بناء على الكمية المتوفرة منها في المستودع.

طرائق البحث ومواده:

بعد أن قامت معظم المؤسسات العالمية الكبرى بأتمتة أعمالها، وبناء نظم معلومات خاصة بها، بدأت بالتفكير جدياً في طرق جديدة للاستفادة من الكم الهائل من المعطيات المتعلقة بها، ومساعدة متخذ القرار على الاستفادة من المعلومات المتوفرة لديه، لأنه بالتأكيد لن يقوم بالاطلاع على التقارير الكثيرة الواردة إليه من نظم المعلومات التي تم بنائها، للوصول إلى نتائج تحسن من أداء الأعمال ضمن مؤسسته، وتمكنه من اتخاذ أنسب القرارات.

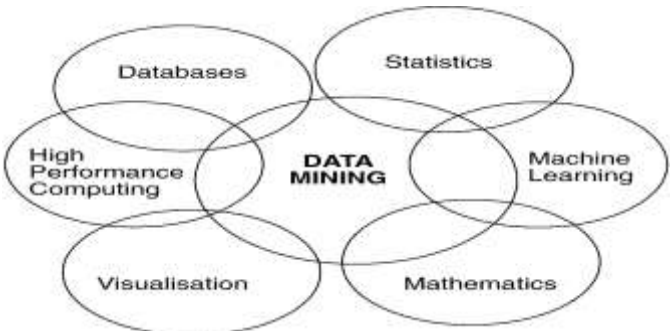
لذلك ظهر مصطلح جديد يعالج المشكلات السابقة هو مصطلح ذكاء الأعمال، طرح مصطلح ذكاء الأعمال أول مرة (غارتنر Gartner) في نهاية العام 1980. يمكن اعتبار ذكاء الأعمال هو مجموعة من الإجراءات الموجهة نحو المستخدم والتي تقوم بالوصول إلى المعلومات واستكشافها، ثم تحليل هذه المعلومات وتطوير طريقة فهمها، مما يؤدي إلى تحسين طريقة اتخاذ القرارات. ويمكن ببساطة تفسير ذكاء الأعمال بأنه الذكاء في فهم أعمالك، فيصبح بإمكانك فهم أعمالك بتحليلك للعمليات التي تقوم بها. إن Microsoft SQL Server Management Studio 2012 هو من البيئات الأساسية التي تستخدم لتطوير حلول الأعمال حيث تتضمن خدمات التحليل [2][1]، ومشاريع خدمات توليد التقارير ومشاريع لخدمات التحليل (Analysis Services Projects)،وهي المستخدمة.

وتبعاً لذلك فإن البيئات الأساسية المستخدمة في هذه الدراسة هي: برنامج Visual Studio 2010، برنامج Microsoft SQL Server Management Studio 2012، خادممعلوماتالإنترنتIIS لغةالاستعلام DMX Queryتم استخدام هذه البيئات لأنها توفر العمل اللازم والضروري لأداء وظائف وطرائق التنقيب عن البيانات في قواعد بيانات مواقع الويب[3] [2] [1]

2. التنقيب عن البيانات:

2.1 تعريف التنقيب عن البيانات

هو عملية بحث عن معرفة من البيانات دون فرضيات مسبقة عما يمكن أن تكون هذه المعرفة [5][4]، كما ويعرف التنقيب البيانات على أنه عملية تحليل كمية بيانات، عادة ما تكون كبيرة، لإيجاد علاقة منطقية تلخص البيانات بطريقة جديدة تكون مفهومة ومفيدة لمالك البيانات. يطلق اسم "نماذج models" على العلاقات المخصصة التي يتم الحصول عليها من التنقيب في البيانات وهو ارتباط عدة علوم منها الرياضيات والإحصاء وقواعد البيانات وغيرها كما يبين الشكل 3.1.



الشكل (2.1) التنقيب عن البيانات

2.1.1عملية اكتشاف المعرفة

اكتشاف المعرفة في قواعد البيانات Knowledge Discovery in Database (KDD) ليس بالعملية السهلة والتي قد يعتقد البعض أنها تتوقف عند تجميع البيانات وإدارتها، بل نراها تمتد إلى التحليل والتوقع والتنبؤ بما سيحدث مستقبلاً.التنقيب في البيانات يشكل جزءا من اكتشاف المعرفة Knowledge Discovery[6][7]، وهذه العملية هي سلسلة متكررة من الخطوات الآتية**[6]** :

1. اكتشاف البيانات Data Discovery: وهي مرحلة جمع البيانات وتشمل كشف وتحديد وتوصيف البيانات المتاحة.
2. تصفية البيانات Data Cleaning: أي حذف البيانات المتضاربة وغير المهمة.
3. تكامل البيانات Data Integration: أي تجميع البيانات التي قد تكون من مصادر متعددة.
4. اختيار البيانات Data Selection: أي استرجاع كل البيانات التي لها علاقة بمهمة التحليل من قاعدة البيانات.
5. تحويل البيانات Data Transformation: أي تحول البيانات إلى نماذج مخصصة للتنقيب بواسطة عمليات التجميع.
6. تنقيب البيانات Data Mining: أي استخدام طرق ذكية تطبق لاستخلاص أنماط البيانات.

7. تقييم النمط Pattern Evaluation: أي لتحديد (تمييز) الأنماط المهمة حيث تمثل قاعدة المعرفة ببعض المقاييس المهمة.

8. عرض (تقديم) المعرفة Knowledge Presentation: أي تقنيات تمثيل المعرفة والرؤية تستخدم لتقديم المعرفة المنقب عنها للمستخدم.

2.1.2 الاستعلام في قاعدة بيانات عادية وقاعدة بيانات بعد تطبيق التقيب:

على سبيل المثال في قاعدة البيانات العادية (Data Base) يتم الاستعلام عن المعلومات الكاملة للزبائن الذين ينتهي اسمهم الثاني بخالد. أو الاستعلام عن الزبائن الذين اشتروا بمبلغ أكثر من 1000 ل.س في آخر الشهر. أو الاستعلام عن أسماء الزبائن الذين اشتروا الحليب.بعد تطبيق التنقيب عن البيانات(Data Mining) يتم الاستعلام عن كل البطاقات الائتمانية والتي عليها مشاكل، وهذا يسمى التصنيف (Classification). أو الاستعلام عن الزبائن الذين لهم عادات شرائية مماثلة، وهذا يسمى العنقدة أو التجميع (Clustering). أو الاستعلام عن السلع التي يتم شراءها بشكل تزامني مع الحليب، – المقصود في كل عملية شراء يقوم بها الزبائن – وهذه تسمى اكتشاف قواعد وعلاقات الارتباط (Association Rules).

2.2 طرائق التنقيب عن البيانات

يوجد العديد من تقنيات التنقيب عن البيانات سيتم استعراض مايلي:

2.2.1 العنقدةclustering:

وهي تقنية مهمة ومشهورة جداً في التنقيب عن البيانات يتم فيها تجميع مجموعات ضخمة من البيانات في عناقيد تحوي مجموعات أصغر ببيانات متشابهة بحيث يكون العنقود مفيد وذو معنى وذلك باستخدام تقنيات مختلفة. العنقود هو مجموعة أغراض البيانات متشابهة في نفس العنقود (قد يكون التشابه كبير أو صغير) ومختلفة عن الأغراض الموجودة في العناقيد الأخرى حالة التشابه تقاس بالمسافة بين الأغراض بحيث تكون المسافة أصغرية بين أغراض العنقود الواحد (الأغراض المتشابهة) والمسافة كبيرة نسبياً بين الأغراض من عناقيد مختلفة (الأغراض غير المتشابهة)[9] [8] .

2.2.2 قواعد الارتباط Association Rules

هي إحدى الاتجاهات الواعدة من Data Mining كأداة من أدوات اكتشاف المعرفة KDD ولديها القدرة على تصفح كميات هائلة من البيانات [10]، وهي تسمح بالتقاط كل القوانين الممكنة التي تشرح بعض الصفات الموجودة اعتماداً على وجود الصفات الأخرى وبمعنى آخر مبدأ عملها هو إيجاد روابط بين البيانات وجعلها ذات ارتباطات تقضي إلى تحقيق استنتاجات ودلائل من بين مجموعات من العناصر الموجودة في قواعد بيانات المعاملات أو مستودعات البيانات الأخرى، وتستخدم في التسويق الموجه TargetedMarketing [11]و[12] .

القواعد التي تعرفها قواعد الارتباط ممكن أن تستخدم للتنبؤ عن مشتريات الزبون المستقبلية المحتملة، مستندة على المواد الموجودة في –عربة التسوق– الخاصة بكل زبون. فعلى سبيل المثال 98% من الناس الذين يشترون إطارات وملحقات السيارة يرغبون بالحصول على خدمات التركيب من قبل عامل الصيانة الموجود في المخزن ذاته.

2.2.3 تحليل السلاسل الزمنية Time Sequence Analysis

السلسلة الزمنية هي عبارة عن قيم ظاهرة من الظواهر في سلسلة تواريخ متلاحقة، أياما أو أشهرا أو سنوات والهدف من وراء تحليل سلسلة زمنية هو دراسة التغيرات التي قد طرأت على الظاهرة التي تمثلها خلال فترة زمنية وتحليل أسبابها ونتائجها أو التنبؤ اعتمادا على فكرة من حوادث الماضي للمستقبل[12].

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 لهذه السلسلة وتحديد القيمة التي يمكن أن يصل إليها هذا الاتجاه في المستقبل.وتحليل السلاسل الزمنية هو نموذج رياضي يمكنه أن يحاكي تقريبا التدرج التاريخي لتلك الظاهرة بحيث يمكنه أن يقدر بدقة قيم السلسلة الزمنية ويمكن استخدامه بالتنبؤ بقيم مستقبلية لهذه الظاهرة.

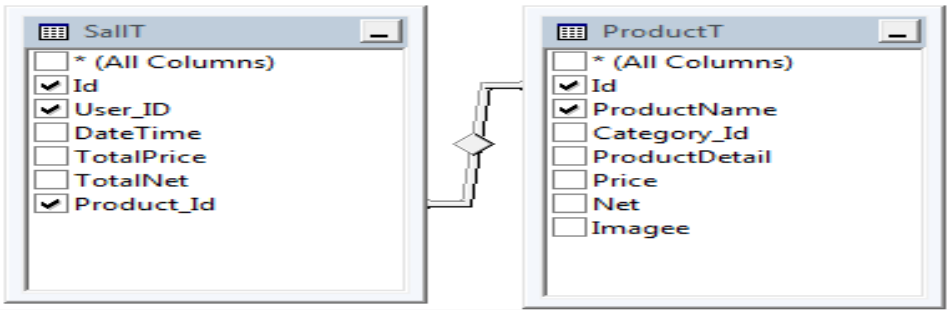
2.2.3.1 تغيرات السلسلة الزمنية

تتعرض أي سلسلة زمنية لنوعين من التغيرات وهذه التغيرات يطلق عليها عناصر السلسة وهي نوعين التغيرات المنتظمةهي التغيرات التي يتكرر ظهورها في السلسلة في مواضع ذات صفات محددة وتشمل الاتجاه العام والتغيرات الموسمية والتغيرات الدورية. تمثل الاتجاه العام (SecularTrend): وهو العنصر الذي يقصد به الحركة المنتظمة للسلسلة عبر فترة زمنية طويلة نسبيا. ويقال إن الاتجاه العام للسلسلة موجب إذا كان الاتجاه نحو التزايد بمرور الزمن ويقال إن الاتجاه العام سالب إذا اتجهت نحو التناقص بمرور الزمن. أما التغيرات الموسمية (SeasonalVariations): هي التي تمثل التغيرات المنتظمة القصيرة الأجل والتي تحدث خلال الفترة الزمنية الواحدة التي لا يزيد طولها عن السنة، فقد تكون أسبوعية أو شهرية أو فصلية.التغيرات الدورية (Cyclical Variations) وهي التي تمثل التغيرات التي تطرأ على قيم السلسلة الزمنية بصورة منتظمة ويزيد أمدها عن السنة، وتتكون من دوال تشبه دوال الجيب وجيب التمام ولكن بأطوال وسعات مختلفة. النوع الثاني هو التغيرات العشوائية أو العرضية (Irregular Variations): تشمل التغيرات العرضية أو الفجائية التي تحدث فجائية لايمكن التنبؤ بها، ومن أمثلتها ما يحدث للنشاط الاقتصادي في بلد ما بسبب الزلازل أو الحروب غير المتوقعة.

2.3 تطبيق خوارزميات التنقيب عن البيانات:

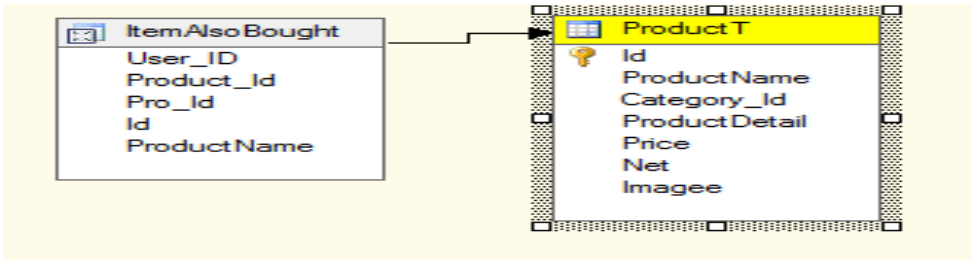
2.3.1 خوارزميات الترابطAssociation

تفيد هذه الخدمة في إظهار المنتجات التي تم شراؤها معا Frequently Bought Together وتظهر هذه الخدمة على صفحة طلب منتج، أي عندما يقوم زبون بطلب منتج معين تظهر له قائمة بالمنتجات الأكثر مبيعا مع المنتج المطلوب. في صفحة استعراض معلومات منتج ما تم إضافة حقل خاص بعرض المنتجات الأكثر طلبا مع المنتج الذي يتم استعراض معلوماته. لتحقيق هذه الخدمة عبر تقنيات التنقيب عن البيانات تم أولا إنشاء View جديد يجمع بين عدد من جداول قاعدة البيانات، بهدف الحصول على مجموعة البيانات المناسبة لخوارزمية التنقيب عن البيانات التي سيتم استخدامها. يتم ربط جدول المنتجات وجدول المبيعات بالشكل لبناء view المطلوب كما هو في الشكل (2.2)



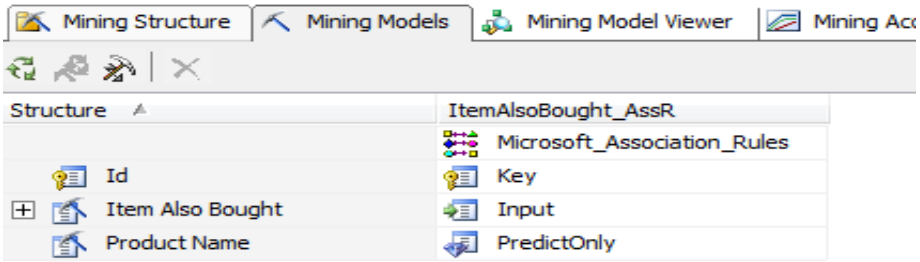
الشكل (2.2) View خدمة Frequently Bought Together

يبين الشكل التالي (2.3) ربط هذا الـ View مع جدول المنتجات:



الشكل (2.3) ربط Viewمع جدول المنتجات

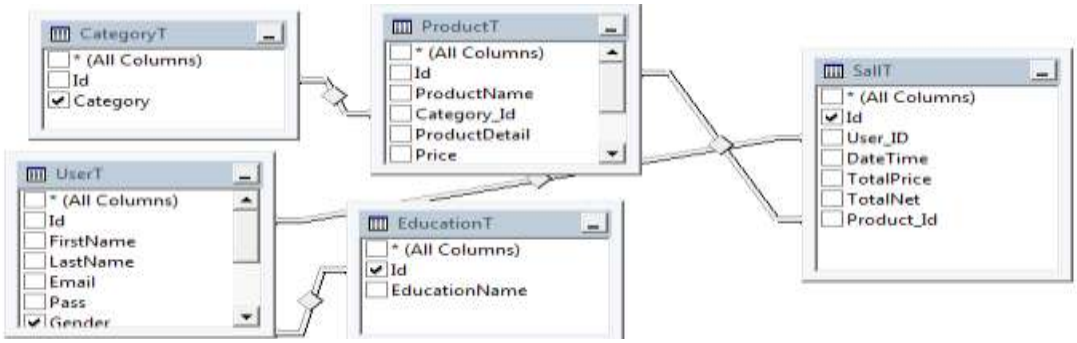
تم استخدام خوارزمية Microsoft Association Rules المتاحة ضمن sql server لإظهار المنتجات التي تم بيعها مع المنتج المطلوب كما في الشكل (2.4) وفيه المعلومات التي ستظهر من تطبيق الخوارزمية



الشكل (2.4) خوارزمية Microsoft Association Rules

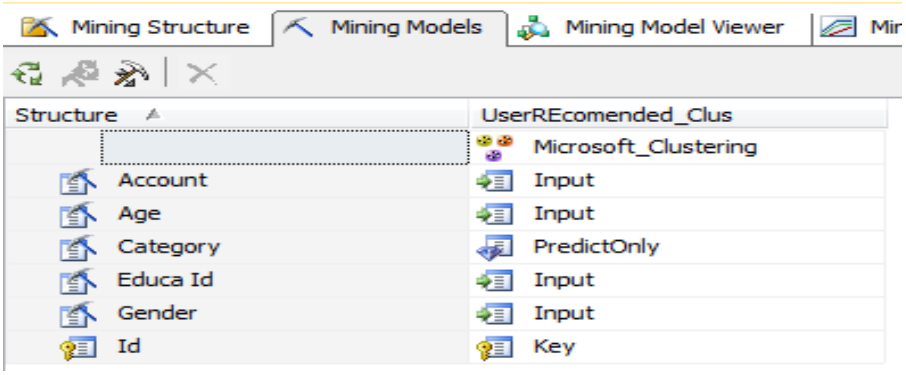
2.3.2خوارزميات العنقدة Clustering

تفيد خوارزميات العنقدة هنا بعرض خدمة تفضيلات المستخدمين Get Personalized Recommendations تعتمد هذه الخدمة على أن تظهر مجموعة المنتجات المتوقع أن يطلبها الزبون وذلك حسب بيانات الزبون، تظهر هذه الخدمة على صفحة الزبون. بعد قيام المستخدم بتسجيل دخوله الى الموقع يقوم الموقع بعرض ملاحظة للمستخدم تشير إلى صنف المنتجات الذي قد يناسب توجهاته في المثال التالي المنتجات المقترحة له هي قائمة من الكمبيوترات المحمولة وذلك اعتمادا على عمره ومستواه التعليمي. لتحقيق هذه الخدمة عبر تقنيات التنقيب عن البيانات تم إنشاء View جديد في الشكل (2.5) يجمع بين عدد من جداول قاعدة البيانات بهدف الحصول على مجموعة البيانات المناسبة لخوارزمية التنقيب عن البيانات التي سيتم استخدامها.



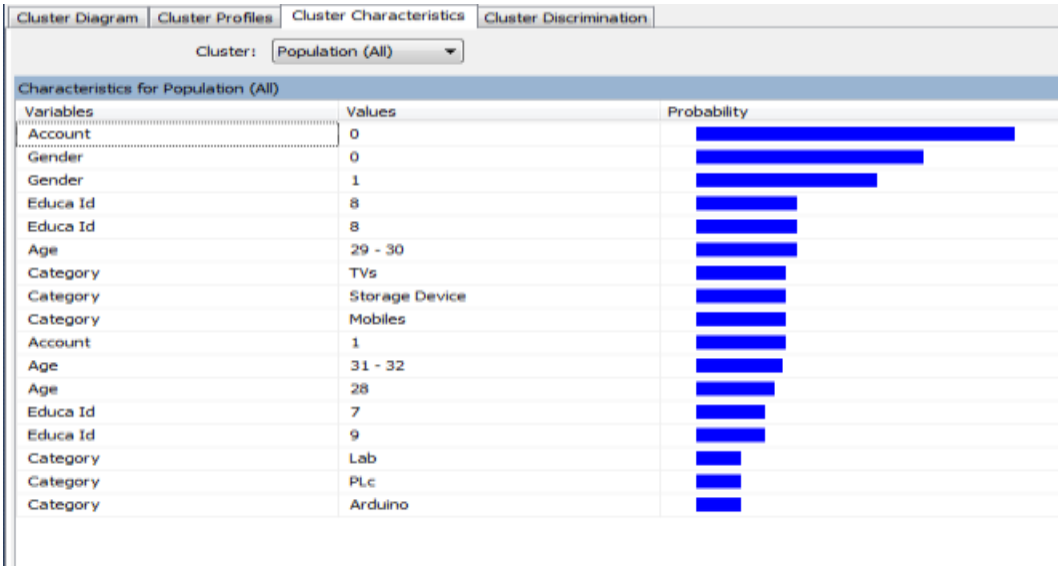
الشكل (2.5) View خدمة تفضيلات المستخدمين

تم استخدام خوارزمية Microsoft Clustering كما في الشكل (2.6) بإدخال الصفات التالية العمر، نوع الشهادة والجنس، ونوع المنتج جميعها كدخل input فكان الخرج عناقيد بحيث تم تجميع الزبائن حسب صفات مشتركة فيما بينهم.



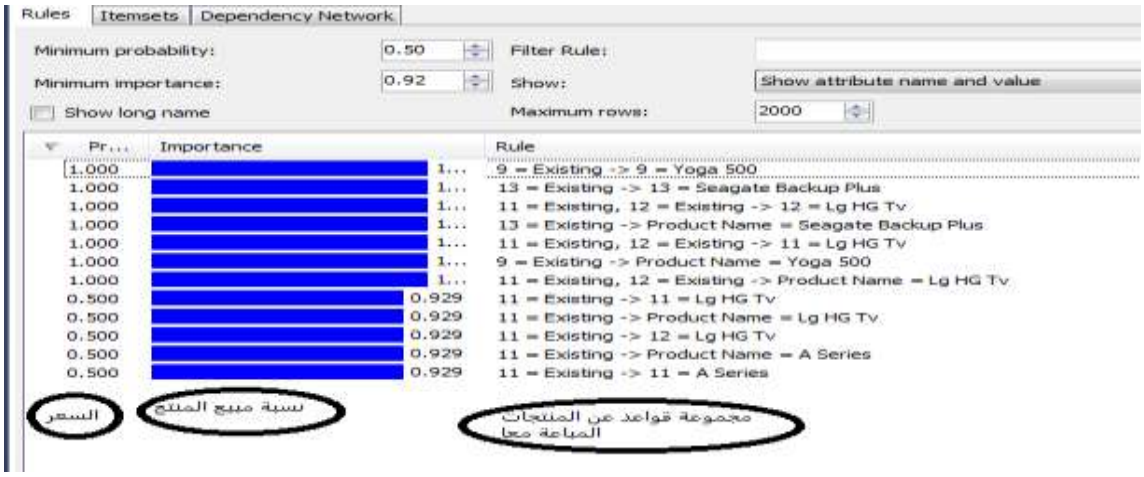
الشكل (2.6) خوارزمية Microsoft Clustering

نتيجة تطبيق الخوارزمية كانت على شكل مجموعة عناقيد تظهر في الشكل (2.7) تحتوي كل منها مجموعة المستخدمين وفقا لما يتشاركون به من توجهات شرائية.



الشكل (2.7) ناتج خوارزمية Microsoft Clustering ويمثل أهم مكونات العناقيد ونسبة تواجدها في العناقيد

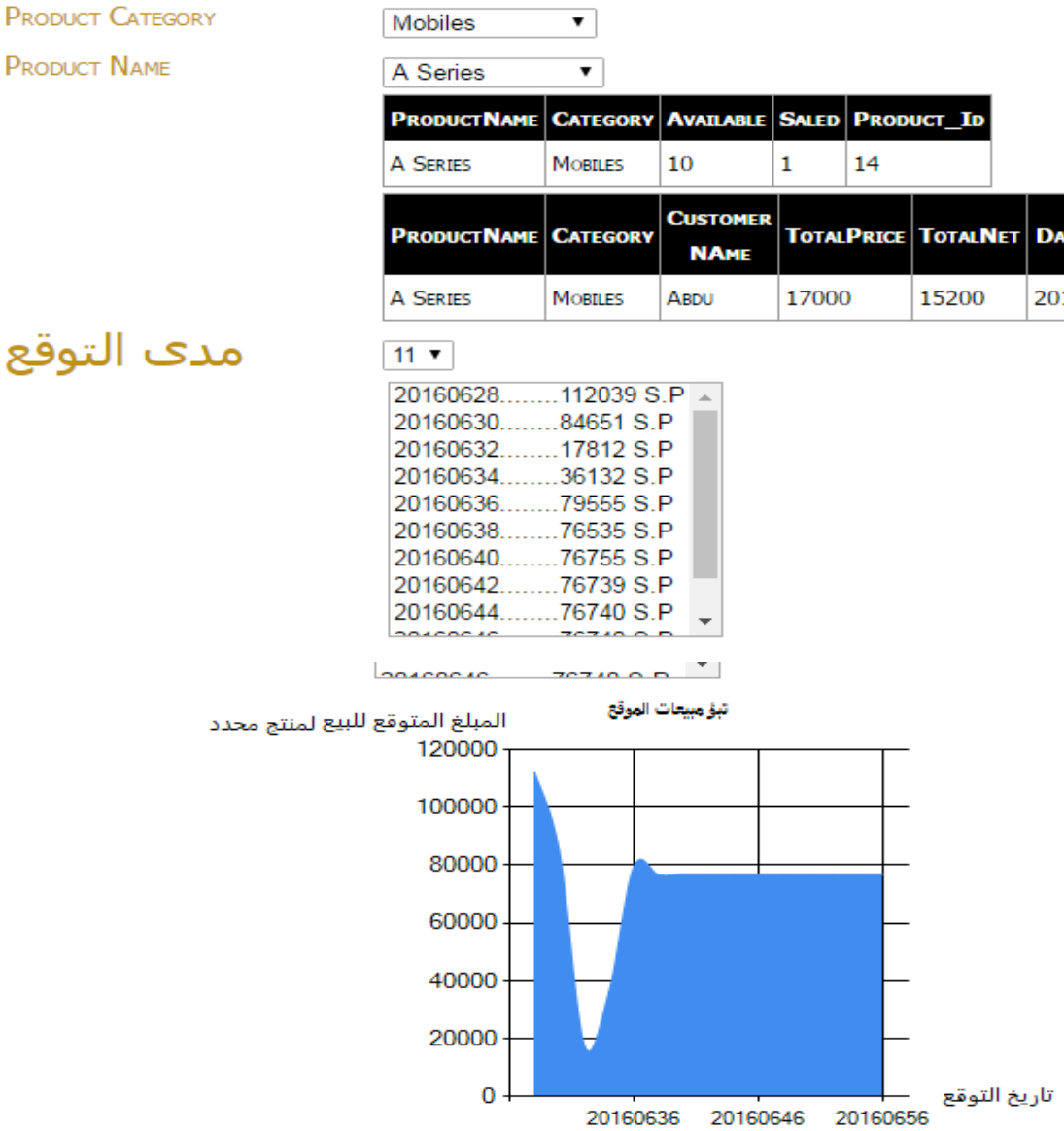
يمثل الشكل (2.8) نتيجة تطبيق هذه الخوارزمية هو عرض مخطط ترابط المنتجات فيما بينها وفقا للمنتجات التي تم طلبها معا.مثلا عندما تم طلب المنتج المتوفر 11 فإن المنتجات التي تطلب معه هي 12 وهو عبارة عن LgHGTv وهو المنتج الأكثر مبيعا مع المنتج المطلوب وعند طلب هذا المنتج سوف يتم عرض بيانات المنتج المرافق.



الشكل (2.8) القواعد الناتجة لخوارزمية Microsoft Association Rules

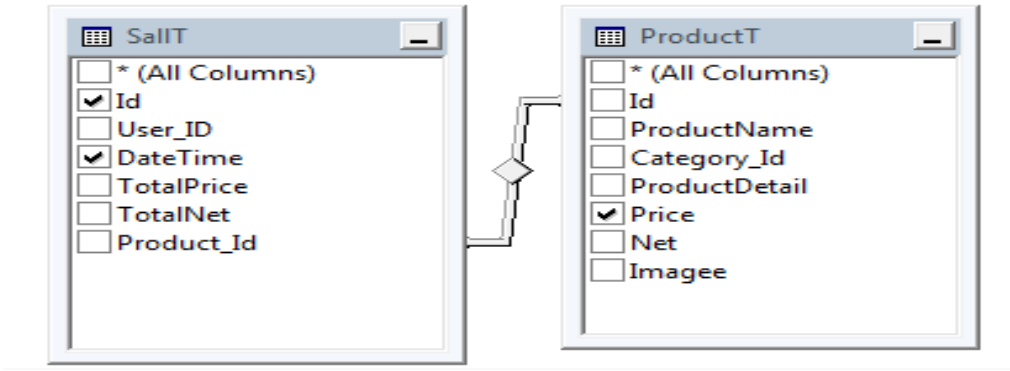
2.3.3 خوارزميات التسلسل الزمني Time Series:

تفيد هذه الخوارزميات في عرض تنبؤ مبيعات الموقع، حيث تُظهر هذه الخدمة لمدير النظام توقع أرباح الموقع لفترة مستقبلية قادمة. في صفحة لوحة التحكم الخاصة بالمدير، عند النقر على تتبع مبيعات المنتجات يمكن للمستخدم تحديد الفترة الزمنية المراد التنبؤ ضمنها كما هو واضح في الشكل (2.9) والذي يظهر توقعات مبيع منتج معين في تواريخ معينة وذلك بأكثر من طريقة ضمن الموقع.



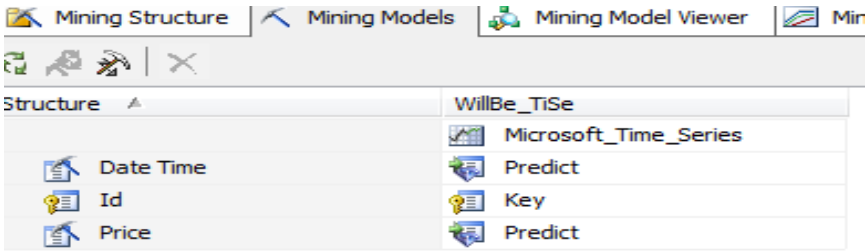
الشكل (2.9) تنبؤ مبيعات منتج ضمن الموقع

لتحقيق هذه الخدمة عبر تقنيات التنقيب عن البيانات قمنا أولا بإنشاء View جديد كما في الشكل (2.10) يجمع بين عدد من جداول قاعدة البيانات بهدف الحصول على مجموعة البيانات المناسبة لخوارزمية التنقيب عن البيانات التي سيتم استخدامها.



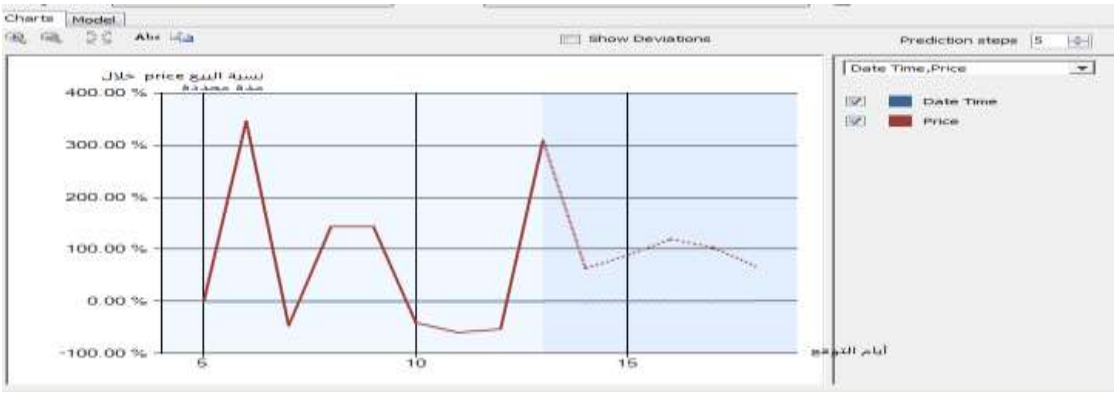
الشكل (2.10) Viewمن جدولي المنتجات والمبيعات

تم استخدام خوارزمية Microsoft Time Series باستخدام البارامترات السابقة (تاريخ الفاتورة ومعرف الفاتورة والسعر):



الشكل (2.11) بارامترات خوارزمية Microsoft Time Series

يبين الشكل (2.12) نتيجة تطبيق هذه الخوارزمية هو عرض توقع قيمة الحقل Price حتى فترة زمنية محددة مقدرة ب 15 يوم:



الشكل (2.12) توقع قيمة الحقل Price حتى فترة زمنية محدد عدة أيام

النتائج والمناقشة:

تم استخدام تقنيات ومنهجيات بناء تطبيقات الويب في هذا العمل من أجل بناء وتصميم موقع ويب مختص بتجارة القطع الالكترونية ومعدات الحواسيب، وتمت الاستفادة من تقنيات التنقيب عن البيانات Data Mining على هذا الموقع من خلال استخدام الخوارزميات الآتية: خوارزميات الترابط Association ومن خلالها وجدنا المنتجات التي تم شراؤها معا Frequently Bought Together، وتظهر هذه الخدمة على صفحة طلب منتج، أي عندما يقوم

زبون بطلب منتج معين تظهر له قائمة بالمنتجات الأكثر مبيعاً مع المنتج المطلوب، مما أدى لزيادة نسبة المبيعات مقارنة مع المواقع الالكترونية الأخرى.

من خلال خوارزميات العنقدة Clustering تم إظهار تفضيلات المستخدمين Get Personalized Recommendations، حيث تظهر مجموعة المنتجات المتوقع أن يطلبها الزبون وذلك حسب مواصفاته، وتظهر هذه الخدمة على صفحة الزبون وهذا ما يؤدي إلى توجيه الزبون لاقتناء بعضها لأنها توافق اهتماماته.

من خلال خوارزميات التسلسل الزمني Time Series تم عرض تنبؤ مبيعات الموقع، حيث تُظهر هذه الخدمة لمدير النظام توقع أرباح الموقع لفترة مستقبلية قادمة وحركة المنتجات ضمن السوق وبالتالي تحديد نسبة نجاح أو فشل منتج معين.

الاستنتاجات والتوصيات

يمكن أن يتم تطوير هذه الدراسة من أجل خدمات أخرى مثل: إضافة خدمة لعرض المنتجات التي يتم شراؤها من قبل الزبائن معا، حيث تعرض هذه المنتجات في صفحة واحدة عند اختيار أحدها فيقلل الوقت اللازم لإيجادها.أوإضافة خدمات الدفع الإلكتروني مثل Paypal، Skrill، Payoneerو Stripe لتسمح للمستخدمين بإجراء المعاملات النقدية بعيدا عن التحويلات المصرفية الروتينية. كما يمكن توظيف البحث وخدمات التنقيب عن البيانات في الشركات والمؤسسات الإنتاجية الحكومية لعرض منتجاتها للمواطنين والتنبؤ بنجاح خطتها الإنتاجية أم فشلها خلال فترة زمنية مستقبلية بما يؤمن عدم كساد المنتجات في المستودعات.

المراجع

1. ROSS MISTRY, STACIA MISNER, "*Introducing Microsoft Sql Server* 2012", Microsoft Corporation, Pp.267.
2. LEBLANCE, PATRICK, "Microsoft Sql Server 2012 Step By Step Online Book And Practice Files", Microsoft Corporation, 2012, Pp.364.
3. Sql Server 2012 Tutorials: *Writing Transact-Sql Statements, Quick Step-By-Step*, Microsoft Corporation, Pp.21, 2102.
4. MOH AMMED J ZAKI, WAGNER MEIRA JR, "*Dataminingand Analysis*", Cambridge University Press, 2014, Pp.606.
5. IAN H.WITTTEN, EIBE FRANK, MARK A HALL, "*Data Mining Practical Machine Learning Tools And Techniques*", Third Edition, Elsevier, 2011, Pp.665.
6. USAMA FAYYAD, GREGORY PIATETSKY, PADHRAIC SMYTH, "*From Data Mining To Knowledge Discovery In Database*", American Association For Artificial Intelligence, Pp.18, 1996.
7. CIOS K.J, PEDRYCZ W, SWINIARSKI R.W, KURGAN L, "*The Knowledge Discovery Process*", Springer, 2007, Pp.17.
8. JERZY STEFANOWSKI, "*data mining clustering*", Poznan University of technology, 2009, pp.113.
9. JURE LESKOVEC, ANAND RAJARAMAN, "*clustering algorithms*", Stanford University, 2016, pp.46.
10. GORDON S. LINOFF AND MICHAEL J. BERRY, "*Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*", third Edition 1 April 2011.

11. TROY RAEDER, NITESH V. CHAWLA, "*Market basket analysis with networks"*, Springer, 2011, pp.17.

12. SAM M WALTON, "*data mining with sql server data tools*", university of ARKANASAS, college of business, 2013, pp.44.

13. HOLGER KANTZ, THOMAS SCHREIBER, "*Time Series Analysis*", Cambridge University Press, 2014.