دور أدوات التنقيب في استخراج المعارف من البيانات البنكية الضخمة في القطاع المصرفي السعودي

إعداد

هوازن بنت عوض عباس مطير

رسالة مقدمة للاستكمال متطلبات الحصول على درجة الدكتوراه

(قسم علم المعلومات / إدارة المعرفة)

أعضاء لجنة المناقشة

	الاسم	المرتبة العلمية	التخصص	التوقيع
مشرف رئيس	د. إبراهيم محمود العمري	أستاذ	هندسة حاسبات	
مشرف خارجي	د. مشعان العتيبي	أستاذ مشارك	علم المعلومات	
مشرف داخلي	د. خليل عبد الحميد ياغي	أستاذ	نظم معلومات	

المستخلص

تشهد بيئة الأعمال وخصوصاً القطاع المالي تحولات كبيرة وتقلبات اقتصادية وتغير توقعات العملاء والموظفين في ظل تنفيذ رؤية ٢٠٣٠ في المملكة العربية السعودية. كما أصبحت البنوك على دراية بأهمية التكنولوجيا وأعطت أولوية للاستثمارات التقنية لدعم مصالحهم والوصول إلى أهدافهم. وعلى صعيد آخر تعتبر البيانات واحدة من أهم الأصول القيمة لأي مؤسسة خصوصاً إذا استطاعت المؤسسة الكشف عن القيمة المخفية في البيانات الأولية في قواعد بيانات المؤسسة من خلال استخدام أدوات التنقيب في البيانات حيث يسمح التنقيب في البيانات باستخراج المعرفة من البيانات التاريخية والمخبأة على شكل أنماط وارتباطات في قواعد والتنبؤ بنتائج المواقف المستقبلية. كما يساعد التنقيب في البيانات على تحسين قرارات العمل، زيادة قيمة العملاء، التواصل، وتحسين رضا العملاء. تستخدم البنوك الرائدة أدوات التنقيب في البيانات لتجزئة العملاء وربحيتهم، وتسجيل الائتمان والموافقة عليه، والتنبؤ بالتخلف عن السداد، والتسويق، واكتشاف المعاملات الاحتيالية وما إلى ذلك. هذا وتتوفر العديد من أدوات التنقيب في البيانات التي تعمل بكفاءة متساوية ولكن البعض منها يكون ذا قيمة أكبر من غيرها. تعاني البنوك والمؤسسات المالية من العديد من العمليات غير المشروعة والمتعلقة بالنصب والاحتيال والقرصنة وبناء عليه فقد تطلب الأمر إلى إيجاد وسيلة للكشف عن مثل تلك المصاعب بالإضافة إلى معرفة سلوك المقترضين والمتعسرين الأمر الذي يسهل صناعة القرار. تهدف هذه الأطروحة إلى فحص أدوات التنقيب في البيانات والوقوف على دورها في استخراج المعارف بالإضافة إلى ترشيح أداة التنقيب المثلى والأكثر فعالية في استخراج المعارف من البيانات البنكية الضخمة والتي تسهم في كشف مخاطر الائتمان بالبنوك السعودية. ويقدم هذا البحث دراسة تحليلية لثماني أدوات تنقيب في البيانات حيث تشمل الثماني تقنيات والتي تتضمن العنقدة، التلخيص، قواعد الارتباط، اكتشاف التسلسل، تحليل الارتباط، التصنيف، التنبؤ، التراجع، وتحليل السلاسل الزمنية. وقد تم اختيار أداة واحدة من كل نوع للمقارنة والوصول إلى أفضل الأدوات التي تدعم عملية اتخاذ القرارات في بيئة البنوك. وهذه الأدوات هي: Excel، RapidMiner، SPSS، Matlab، Knime، Orange، Weka، Lucidchart. اعتمدت هذه الأطروحة على استخدام ثلاث مناهج بحثية مناسبة لطبيعة الدراسة وهي المنهج الوثائقي النظري، المنهج المقارن والمنهج التجريبي. قامت هذه الدراسة بتحليل ومقارنة هذه الأدوات الثمانية وعرض النتائج وقد

تم التوصل إلى عدة نتائج أهمها أن تقنية شجرة القرار هي أفضل تقنية تستخدم للوصول إلى نتائج دقيقة في تطبيقات التنقيب في البيانات البنكية. كذلك أن الأداة RapidMiner تحتل المركز الأول بين هذه الأدوات في تقييم معايير الأداء والوظيفة، سهولة الاستخدام والدعم الفني لمستخدميها. وباستخدام الأدوات المذكورة وخاصة RapidMiner فإن نسبة المخاطرة قد تنخفض بدرجات عالية حيث تقوم RapidMiner بتقليص الوقت المستغرق في الكشف عن أنماط المخاطر في طلبات الائتمان وبالتالي ترفع نسبة الحذر لدى البنوك وتحسين الأداء المالي والعلاقة مع العملاء مما يعكس إجابياً على المجتمع ككل بسهولة تقديم القروض وتمكين البنوك في مساعدة الفرد والمجتمع.

The Role of Data Mining Tools in Extracting Knowledge from Big Data in Saudi Banking Sector

By Hawazen Awad Abbas Metair

A thesis submitted for the requirements of the degree of PhD

Information Science / Knowledge Science

Supervised By

Prof. Dr. Ibrahim AlOmary

FACULTY OF ARTS AND HUMANITIES

KING ABDULAZIZ UNIVERSITY

SAUDI ARABIA

1444 / 2023

دور أدوات التنقيب في استخراج المعارف من البيانات البنكية الضخمة في القطاع المصرفي السعودي

إعداد

هوازن بنت عوض عباس مطير

رسالة مقدمة للاستكمال متطلبات الحصول على درجة الدكتوراه

(قسم علم المعلومات / إدارة المعرفة)

أعضاء لجنة المناقشة

	الاسم	المرتبة العلمية	التخصص	التوقيع
مشرف رئيس	د. إبراهيم محمود العمري	أستاذ	هندسة حاسبات	
مشرف خارجي	د. مشعان العتيبي	أستاذ مشارك	علم المعلومات	
مشرف داخلي	د. خليل عبد الحميد ياغي	أستاذ	نظم معلومات	

Abstract

In light of the implementation of Vision 2030 in the Kingdom of Saudi Arabia, business environment, especially the financial sector, is witnessing major transformations changing customer and employee expectations. Banks became aware of the importance of technology and gave priority to technical investments to support their interests and reach their goals. On the other hand, data is considered one of the most valuable assets for any organization, especially if the organization is able to reveal the hidden value in the primary data in the organization's databases through the use of data mining tools, as data mining allows extracting knowledge from historical and hidden data in the form of patterns and correlations in rules and predict the outcome of future situations. Data mining also helps improve business decisions, increase customer value, communicate, and improve customer satisfaction. Leading banks use data mining tools for customer segmentation and profitability, credit scoring and approval, default prediction, marketing, fraudulent transaction detection and so on. There are many data mining tools out there that work equally well but some are more valuable than others. Banks and financial institutions suffer from many illegal operations related to fraud and piracy. Accordingly, it was necessary to find a way to detect such difficulties, in addition to knowing the behavior of borrowers and defaulters, which facilitates decision-making. This thesis aims to examine data mining tools and stand on their role in extracting knowledge, in addition to nominating the optimal and most effective data mining tool for extracting knowledge from big data, which contributes to detecting credit risks in Saudi banks. This research presents an analytical study of eight data mining tools, which include eight techniques such as clustering, summarization, correlation rules/ analysis, sequence discovery, classification, prediction, regression, and time series analysis. This thesis used the following research approaches: documentary-theoretical approach, comparative approach, and experimental approach. Several results were reached, the most important of which is that the decision tree technique is the best technique used to reach accurate results in banking data mining applications. RapidMiner also ranks first among these tools in evaluating performance and functionality standards, ease of use and

technical support for its users. And by using the mentioned tools, especially RapidMiner, the risk ratio may decrease by high degrees, as RapidMiner reduces the time it takes to detect risk patterns in credit applications. This raises the percentage of caution among banks and improves financial performance and the relationship with customers, which benefit the society as a whole with the ease of providing loans and enabling banks to helping individuals and communities.

Key Words

Knowledge acquisition - Knowledge extraction - Data mining – Data Mining Tools - Big Data.