Abstract

Banks deal with huge amounts of customer's data and thus needs tremendous efforts to improve the understanding of the accumulated data in order to detect customer's behavior and accordingly enable the executive mangers to make the right decision and avoid any possible losses, wasting time and effort .

The main aim of this thesis is to distinguish between borrowers who pay back loan from those who don't . therefore the executive mangers can easily reduce the costs of non-payment borrowers and decrease the high number of bad loans in order to serve the bank and its customers by using data mining techniques.

The dataset of this research was obtained from the UCI machine learning repository website. In order to improve the accuracy of our classification and gain useful results some preprocessing techniques were applied such as : removed any irrelevant and correlated data , implemented data discretization ,data cleaning ,and target class balancing as well to achieve a suitable dataset for our Algorithms. Then five data mining classification techniques were conducted which are: Naive bayes , J48, IBK, Multilayer Perceptron (MLP) and Sequential minimal optimization (SMO).The Weka software from Waikato university with (10-cross validation) was used to model and validate the proposed models.

Experiments in this research were conducted in two stages. Firstly, J48 classifier was applied on full dataset, the results carried out in this stage show that: applying of the preprocessing techniques on the data set improved the performance of the classifier. Secondly, five classification techniques were applied to the preprocessed datasets. The results carried out in this stage showed that the performance of the five classification algorithms are nearly same . Out of these five classification algorithms, J48classifier had the highest accuracy (84.35%) .

مستخلص البحث

تتعامل البنوك مع كميات هائلة من بيانات العملاء وهذا يحتاج إلى جهود هائلة لتحسين فهم البيانات المتراكمة ، لاكتشاف سلوك العملاء وبالتالي تمكن المدراء من اتخاذ القرار الصحيح ، وتجنب أي خسائر محتملة ،من إضاعة الوقت والجهد.

الهدف الرئيسي من هذه الرسالة هو التمييز بين أصحاب القروض المستردة وأصحاب القروض الهالكة مما يمكن المدراء من تقليل تكاليف أصحاب القروض غير المستردة وبالتالي العدد الكبير من القروض المعدومة من أجل خدمة البنك وعملائه باستخدام تقنيات تنقيب البيانات.

تم الحصول على مجموعة البيانات لهذا البحث من موقع (UCI). من أجل تحسين دقة تصنيفنا والحصول على نتائج مفيدة ، تم تطبيق بعض تقنيات المعالجة المسبقة مثل: إزالة أي بيانات غير ذات صلة ومترابطة ، وتقدير البيانات المنفصلة ، وتنظيف البيانات ، وموازنة الطبقات المستهدفة أيضًا لتحقيق مجموعة بيانات مناسبة لخوارزمياتنا. ثم تم إجراء خمس تقنيات لتصنيف بيانات التعدين وهي Naive Bayes و J48 و IBK و Multilayer Perceptron (MLP)و Sequential minimal optimization (SMO)تم استخدام برنامج Weka من جامعة Waikato مع استخدام الدالة (10-valid validation) لعمل النماذج المقترحة والتحقق من صحتها.

أجريت التجارب في هذا البحث على مرحلتين. في المرحلة الأولى ، تم تطبيق التصنيف J48 على مجموعة البيانات الكاملة من البيانات ، وأظهرت النتائج التي أجريت في هذه المرحلة ما يلي: أن تطبيق تقنيات المعالجة المسبقة على مجموعة البيانات قد حسن أداء المصنف. في المرحلة الثانية ، تم تطبيق خمسة أساليب تصنيف على مجموعات البيانات التي تم تجهيزها مسبقًا. أظهرت النتائج التي أجريت في هذه المرحلة أن أداء خوارزميات التصنيف الخمسة مماثل تقريبا. من خوارزميات التصنيف الخمسة هذه ، كان التصنيف J48 يتمتع بأعلى دقة (84.35٪) .

Dedications

This dissertation is lovingly dedicated to :

My mother

A strong and gentle soul who taught me to trust in ALLLAH , believe in hard work and that so much could be done with little

My father

for earning an honest living for us and for being my first teacher who support and encourage me to believe in myself

Acknowledgement

Thanks first and foremost to God Almighty who honored me in accomplishing this work, and I would like to express my sincere gratitude to Sudan University of Science and Technology Collage of Graduate Studies for letting me fulfill my dream of being a student , my supervisor, Dr. Shaza Mirghani , I am extremely grateful for her assistances and suggestions throughout this research .I would also like to thank my parents and friends who helped and supported me a lot in finalizing this research.

Table of Contents

List of Tables

List of Figures