UNIVERSITY OF MOSUL
COLLEGE OF COMPUTER SCIENCES
AND MATHEMATICS

DESIGN MACHINE LEARNING INTELLIGENT SYSTEM FOR DATA MINING IN BANK FINANCIAL ANALYSIS

Maan Younis Anad Alsaleem

Ph.D.Thesis in
Computer Sciences

Supervised by

Asst. Prof. Dr. Safwan O. Hasoon

2021 A.D. **1443 A.H.**

ABSTRACT

As a result of the development in computer science and machine learning and its entry into many fields, especially the banking and financial field, it has become necessary to build smart banking systems used in decision support, especially bank loans. Bank loans are considered one of the main pillars of financial institutions and the economy in general. These loans are a type of debt offered to individuals and companies, which includes a promise to pay its value during a fixed and agreed period. The decision to grant loans is usually made by committees that study the data of the person requesting the loan. By inserting the computerization in banks, these loans have become registered in the bank databases, so they can be studied and analyzed in addition to making use of statistical and intelligent methods to predict the risks of bank loans by analyzing historical customer data.

In this dissertation, several challenges facing the use of machine learning algorithms to predict the risks of bank loans and machine learning algorithms have been studied. The data sets of bank loans have been studied and compared. The data of the selected group have been analyzed, and the performance of machine learning algorithms has been compared to the methods of compensating that data. Surveying, reviewing, and forecasting the existing business and systems for bank loans. Also testing the performance of a group of supervised classification algorithms (decision trees, logistic regression, support vector machine, K nearest neighbor and Bayes algorithm) on an imbalanced data set for bank loans, where the imbalance is a problem for machine learning algorithms because most models at the training biases towards the majority class in the binary classifications.

It has been found that the use of data balancing methods may be feasible if the model is static and the use of ensemble learning algorithms is a promising solution. Most of the existing systems and studies depend on the use of static models that are not adapted to market requirements and changes in user data. An incremental learning model using Incremental learning of Extreme Gradient Boost (IXGBOOST) algorithm has been proposed to meet these challenges. Using Bayesian optimization to automatically find the optimal parameters of the proposed algorithm, as finding these parameters through the experiment is one of the difficult problems that do not correspond to the proposed incremental model and comparing it with the traditional grid search method. in this study achieved an accuracy of 85 %.

جامعة الموصل
كلية علوم الحاسوب والرياضيات

تصميم نظام تعلم الة ذكائي لتنقيب البيانات في التحليلات المالية المصرفية

أطروحة دكتوراه في
علوم الحاسوب

تقدم بها

معن يونس عناد السليم

بإشراف

أ.م.د. صفوان عمر حسون

1443 هـ 2021 م

الخلاصة

نتيجة للتطور الحاصل في علوم الحاسوب والتعلم الالي ودخوله في الكثير من المجالات وخصوصا المجال المصرفي والمالي اصبح من الضروري بناء انظمة مصرفية ذكية تستخدم في دعم القرار ولاسيما القروض المصرفية التي تعتبر من الاعمدة الاساسية للمؤسسات المالية والاقتصاد بشكل عام. ان تلك القروض هي نوع من انواع الديون التي تقدم الى الاشخاص والشركات والذي يشمل وعدا بسداد قيمته اثناء فترة ثابتة متفق عليها. اذ يتم بالعادة اتخاذ قرار منح القروض عن طريق لجان تقوم بدراسة معطيات الشخص الذي يطلب القرض. وبدخول الحوسبة في المصارف اصبحت تلك القروض مسجلة في قواعد بيانات المصارف والتي يمكن دراستها وتحليلها والاستفادة من الطرق الاحصائية والذكائية للتنبؤ بمخاطر القروض المصرفية عن طريق تحليل بيانات العملاء التاريخية.

تم في هذه الاطروحة دراسة عدد من التحديات التي تواجه استخدام خوارزميات تعلم الالة للتنبؤ بمخاطر القروض المصرفية. اذ تمت دراسة مجموعات بيانات القروض المصرفية ومقارنتها وتحليل بيانات المجموعة المختارة ومقارنة اداء خوارزميات تعلم الالة بالنسبه لطرق تعويض تلك البيانات حيث تم استخدام مجموعتين من بيانات القروض المصرفية باحجام مختلفة وتحديد الطرق الامثل لتعويض البيانات المفقودة حسب حجم مجموعة البيانات. بالاضافة الى مسح واستعراض الاعمال والانظمة القائمة الخاصة بالقروض المصرفية والتنبؤ بها. واختبار اداء مجموعة من خوارزميات التصنيف الخاضعة للاشراف (اشجار القرار, الانحدار اللوجستي, متجه دعم الالة, الجار الاقرب, خوارزمية بايز) على مجموعة بيانات غير متوازنة خاصة بالقروض المصرفية حيث يمثل عدم التوازن مشكلة بالنسبة لخوارزميات تعلم الالة لكون اغلب النماذج عند التدريب تنحاز الى فئة الاغلبية في التصنيفات الثنائية.

توصلت الدراسة الى ان استخدام طرق موازنة البيانات التقليدية قد تكون مجدية في حالة النموذج الثابت (الغير متكيف) وان استخدام الخوارزميات التجميعية يعد حلا واعدا. ان اغلبية الانظمة القائمة حاليا والدراسات تعتمد استخدام نماذج ثابتة غير متكيفة مع متطلبات السوق والتغير في بيانات المستخدم. تم اقتراح نموذج تعلم تراكمي باستخدام خوارزمية (IXGBOOST) لمواجهة تلك التحديات. واستخدام امثلية بايز لايجاد المعلمات المثلى للخوارزمية المقترحة بشكل اوتوماتيكي حيث ان ايجاد تلك المعلمات عن طريق التجربة تعد من المشاكل الصعبة والتي لاتتوافق مع النموذج التراكمي المقترح ومقارنتها مع طريقة البحث الشبكي التقليدية.