

A Collaborative Mining-Based Decision Support Model for Granting Personal Loans in the Banking Sector

Amira M. Idrees, Faculty of Computers and Information, Fayoum University, Egypt*

Ayman E. Khedr, Faculty of Computers and Information Technology, Future University in Egypt, Egypt

ABSTRACT

One main potential objective for financial corporations is to retain long-term customers. Configuring customer knowledge is no doubt mandatory to lower the risk level. Loans and credit card granting are two services that are offered by the banking corporations that can be categorized as high-risk services. Therefore, it is highly recommended for the corporations to have intelligent support for providing an accurate granting decision which naturally leads to minimizing the associated risk. In this research, a decision support model is proposed for loan granting. The proposed model applies a set of data mining techniques in a collaborative environment that aims at applying different techniques with considering their results according to the technique's evaluation weight. The proposed model results present the recommendation for each customer's loan granting a request to be either accepted or rejected. The proposed approach has been applied on a loan granting dataset, and the evaluation results revealed its superiority by 92% success in reaching high accurate decisions.

KEYWORDS

Banking Systems, Classification, Data Mining, Decision Support Systems, Evaluation, Loan Systems, Metrics' Weighting, Recommendation Systems

INTRODUCTION

It is a fact that blocking banks' resources has its major negative impact on the institutions in specific and the economic field in general. For example, it minimizes the bank's capabilities, affects negatively on the bank's productivity, as well as reduces the economic progress (Khedr & Borgman, 2006). Therefore, in addition to the vital objective that ensuring the efficient resources' return for the financial institutions, to avoid any lacking for these resources (Khedr & Kok, 2006), the successful assessment of the organization progress is also strongly recommended (Rehman & Hashim, 2020). Both objectives could be viewed as a collaborative environment. In (Rehman & Hashim, 2020), a relation between assessing the risk level of internal fraudulent with its impact on the organization progress is confirmed which ensures the efficient utilization of the organization resources while the same goal is also confirmed in (Khedr & Kok, 2006) with the perspective of ensuring the efficient utilization of these resources.

On the other hand, data mining techniques have been positively contributed to different fields in general (Idrees, ElSeddawy, & Zeidan, 2019) (Al Mazroi, Khedr, & Idrees, 2021) (Hassan, Dahab,

DOI: 10.4018/IJESMA.296573

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

Bahnasy, Idrees, & Gamal, 2014) and in the financial aspects of the banks in specific which has a positive contribution in both the banks as the main organization as well as the positive economic impact (Khedr & Borgman, 2006). Therefore, the requirement for continuous enhancements in applying these techniques with different perspectives has been highly recommended specially with the success of the collaborative environment approach (Hassan, Dahab, Bahnassy, Idrees, & Gamal, 2015). Although customers' loans are only one of the profitable services in the banking fields that generate an income to the financial institution, however, it is considered one of the highest services' risk. One of these risks is the requirement of correctly identifying the customers' segment for the loans with correctly setting the loan amount. Exposure of these risks leads to lack of control and mismanagement, inability to identify the risk intensity and diversity all along with the ability to reach more dangerous situation such as bankruptcy. Therefore, detecting the targeted customers for loans, identifying the terms, as well as ensuring the payment are three yet interrelated targets for loans granting risk's minimization and consequently, increase the institution profitability while directly leading to economic enhancement. The risks of credits could be modeled in the inability of the customers to pay the loan's installments, therefore, accurate detection to the loans' customers is a critical target to consider. Therefore, validating the suitable customers who demand such a service with this level of high risk could be monitored as one of the highest sector objectives (Khedr, Abdel-Fattah, & Nagm-Aldeen, 2015).

Does adopting intelligent techniques can affect the uncertain environment' business process? What are the most suitable techniques for classifying loans' granting customers? These questions address the research motivation that targets to explore the effective techniques' environment for enhancing the business services' decisions in an uncertain environment such as the banking field. The main success factor of the intelligent techniques offered by different fields such as data mining is the ability to integrate the business information with the proposed intelligent mechanism and method targeting to discover non-intuitive insights hidden information in the institution data (Khedr, Idrees, & El Seddawy, 2016) (El Seddawy, Sultan, & Khedr, 2013) (El Seddawy, Sultan, & Khedr, 2013). This discovered information is considered the spark for refining the corporation process, re-design the corporation policies, and re-organize the corporation relations with its customers as well as the employees (Idrees & Ibrahim, 2018).

In this research, investigating the impact of applying intelligent techniques is performed to explore the most suitable technique for granting loans services. Highlighting the productivity is an effective motivation which increases the support level for the decision maker. Following this motivation, the research aims at successfully determine the suitable customer segment for loans by applying a set of classification techniques in a collaborative environment with considering the contribution of each technique in reaching the most accurate decision. The techniques' recommendations are encountered with a defined weight representing the technique contribution in reaching the final decision, then the proposed model provide the required recommendation for loans' granting according to these contributions.

The remaining of the paper discusses the related work in section 2 with two directions, the business perspective in loans granting services and the technical perspective in demonstrating the success stories in applying the intelligent techniques specially data mining techniques in the financial field. Moreover, section 3 describes the proposed model in detail with an illustration of the detailed stages and data sources while section 4 discusses the experimental case studies with presenting the evaluation results. The research applies the proposed model on a loan granting dataset which is an available source to ensure the proposed model applicability. Finally, the conclusion is presented in section 5 which demonstrates the main keys of the current research and highlights some of the future directions.

LITERATURE REVIEW

The vital impact of the intelligent techniques' support in business fields has been confirmed in different research. For example, the role of developing digital infrastructure to support the government services has been presented in real cases as in (Zoroja, Strugar, & Jaković, 2020). The research presented the positive attitude of a set of participants in Croatia towards using intelligent systems in the government services. The research targeted the participants with an age of range 18 to 34 years who confirmed the significance of the governmental services' electronic availability. Another research was conducted in Egypt (Awad & Dessouki, 2017) which applied an experiment on a set of customers in the Egyptian banks and confirmed their willingness to use mobile banks' services. This conclusion motivates the banking sector for higher mobile banking services' adoption.

A more generic segment has participated in the case study of (Kieźel & Stefańska, 2019) which also highlighted the vital role of technology innovation in enhancing the banking services. The research targeted different age segments and the results showed that although the age of the customers revealed their interest in different services, however, their interest in common characteristics were revealed such as the security, the low services' price, and the comfort in usage. These characteristics are determined to be the main reasons for their satisfaction towards the electronic services. Moreover, the research in (Abdel-Fattah, Khedr, & Nagm Aldeen, 2017) highlighted the perspective that although embedding the technology in business proved to have a positive impact, however, it is a fact that each business process has its own characteristics which requires a business process modeling that meets these characteristics. The research proposed an identifying framework for selecting the suitable business process modeling with its corresponding business process. The previous researches have confirmed the vital role of technology adoption in different fields in general. In specific, as data mining is considered one of the successful fields that provides successful intelligent techniques for different goals in the business organizations, therefore, the following subsections discuss the impact of data mining techniques in different business fields, then a focus on this impact in banking sector is discussed.

DATA MINING

Data mining techniques have been widely applied in different fields targeting different aspects. For example, the research in (Idrees, Ibrahim, & El Seddawy, 2018) proposed a generic approach for exploring trends using different types of data. The research used a dataset for heart diseases mortality which was gathered in Maryland to explore the disease trends and identify the possible infected places. Another research in (Sultan, Khedr, Idrees, & Kholeif, 2017) targeted detecting the key performance indicators using data mining techniques. The research proposed an enhancement in FP-Growth association rules algorithm for successfully detecting key performance indicators with no human intervention. With the increase in the business demands, data has become more vital, therefore, the requirement for data manipulation techniques has been critically considered (Zanin, et al., 2016). In ((Koutanaei, Sajedi, & Khanbabaei, 2015), four data mining algorithms have been applied targeting to select the appropriate algorithm based on a set of parameters. The accuracy results pointed at PCA for the highest accurate results for features' selection, moreover, Neural Network has been nominated for the best classification algorithm. Another research by (Baixauli, Alvarez, & Mónica, 2012) highlighted that each model could be superior to the others with respect to the data source and the problem nature. Data mining also contributed to social networks as different researches targeted manipulating the data in social networks such as in (Khedr, Idrees, & Shabaan, 2020). The research aimed at detecting the fraudulent emails. The research was applied in a real dataset and succeeded to reach

a high classification performance by building an extendable lexicon for spam emails using data mining techniques as well as natural language processing methods. The combination of the users' web data represented in the users' activities over the internet and the users' opinions was explored in

(Helmy, Khedr, Kolief, & Haggag, 2019). The research highlighted the integrated sources targeting the enhancements of business intelligence methods.

In the learning field, data mining has contributed in different perspectives. a research in (Hassouna, Khedr, Idrees, & ElSeddawy, 2020) aimed at applying data mining techniques in detecting students' personalization using the student's behavior and detecting his preferable interaction method with the learning system. Another research in (Mostafa, Helmy, Khedr, & Idrees, 2020) aimed at the same perspective, the research proposed an extendable approach to include the student's social networks behavior. However, the research was only introducing a proposed approach with results that reveal its vital perspective but lacked an experimental study to prove the applicability of the proposed approach. Moreover, the research of (Idrees, El Seddawy, & EL Moaaz, 2019) presented the vital role of data mining techniques in business continuity. The research proposed an approach for business continuity in the educational sector. It aimed at supporting the business continuity management process with raising the stakeholders' awareness. The proposed framework presented the process activities' relations and their impact on the overall process. Moreover, the research investigated the adaptation requirement of NAQAA approach according to the gaps that have been explored using data mining techniques. Another perspective in applying data mining in the educational sector (Helmy, Emam, Khedr, & Bahloul, 2020) which aimed at detecting the key performance indicators for measuring the organization performance. The research presented a survey for the subject on focus which revealed that the topic still requires strong contribution.

Data mining is also proposed as a solution in ensuring food safety in (Abogabal, Ouf, Idrees, & Khedr, 2020). The research proposed a method for detecting the key performance indicators for food safety as well as highlighting its deviation using data mining techniques. The results of the preliminary case study has been evaluated by the experts in the field which confirmed a successful determination of the effective key performance indicators. Moreover, data mining also contributed in the real state field as discussed by the research in (Idrees, ElSeddawy, & Zeidan, 2019). The research proposed an adaptation to Saaty method with the collaboration of data mining techniques targeting to the enhancement of the real state products by considering more concrete and consistent voice of customers. The research confirmed that the results were more accurate than the previous researches by applying the proposed approach on a real life project and the customers' feedback has been measured. More recently, a research in (Bach, Dumičić, Žmuk, Ćurlin, & Zoroja, 2020) presented the success of applying data mining techniques including decision tree and associations to detect the fraud of working-hours claims in the projects based organizations. The research aimed at confirming the effectiveness of applying data mining techniques to control the internal environments' aspects of the organizations. On the other hand, another research in (Bach, Vlahović, & Pivar, 2020) highlighted the same target in different business field. The research proposed a method for developing a profile for fraudulent clients. The method was based on developing self-organizing maps in the leasing business by applying clustering mining techniques. The proposed method of (Bach, Vlahović, & Pivar, 2020) was validated by the experts in the field who confirmed its applicability in positively contributing to the strategic business plans as well as the tactical aspects.

APPLICATIONS OF DATA MINING IN BANKING

Different data mining techniques have been applied in the banking sector to support different business aspects. For example, detecting the banks' failure was one of the main objectives in (Manthoulis, Doumpos, Zopounidis, & Galariotis, 2020). The research proposed a classification model which highlights the bank that belongs to the set of failures to detect the possibility of the bank failure prior to the actual situation. The research was applied on a dataset including the customers' transactions covering a set of US banks. The same objective was highlighted in (Carmona, Climent, & Momparler, 2019) which also applied on US banks dataset but in a different time period with not only detecting the probability of failure but it also succeeded in detecting the indicators for encouraging the failure

situation. Following the same objective with confirming the superiority of applying data mining models, the research in (Jing & Fang, 2018) compared the effectiveness of applying data mining techniques with applying logit models in the bank failure situation detection. The research confirmed that data mining techniques, represented in support vector machine and neural networks has higher prediction results.

The classification data mining techniques not only contributed to predicting the banks' failure, but it also contributed to the customers' classification. An earlier research in (Smeureanu, Ruxanda, & Badea, 2013) aimed at applying Support Vector Machine algorithm and Neural Networks algorithm in customers' classification, the research revealed the higher performance of Neural Networks over Support Vector Machine. Another research in (Moro, Cortez, & Rita, 2014) applied a set of data mining techniques; included Neural Networks, Decision Trees, Support Vector Machine, as well as applying Logistic Regression; over a real dataset of a bank in Portuguese. The aim of the research was to predict the impact of the telemarketing calls in the products' sales. The results of the research also highlighted that Neural Networks reached the highest performance of all the other applied techniques. However, a recent research by (Ilham, Khikmah, Indra, Ulumuddin, & Iswara, 2019) contradicted this conclusion by applying a set of classification techniques over banking customers' data. The set included Neural Networks, K-Nearest Neighbor, Random Forest, and others, then it concluded the higher performance of the Support Vector Machine over all the presented algorithms. Additionally, a research in (Farooqi & Iqbal, 2019) also applied a set of classification techniques targeting to identify the expected deposits as a criteria for telemarketing evaluation. The research approach also included applying a set of classification techniques and revealed the advance of J48 algorithm. This contradiction confirms the current research argument which states that the dataset features can be a vital factor which should be critically considered to identify the suitable technique while the techniques' collaboration could be more effective. The approach of collaboration has been previously examined in (Lahmiri, 2017) but in a narrow perspective. The research aimed at applying a set of Neural Networks for predicting the telemarketing return. The research highlighted that collaboration approach is a promising direction with a requirement for further investigation.

The customer's behavior has been also considered through applying data mining in (Ogwueleka, Misra, Colomo-Palacios, & Fernandez, 2015). The research aimed at monitoring the current bank customers and predicting the new customers' behavior. New customers have been classified by applying mining techniques and the behavior prediction is explored based on their predicted class.

More recently, a research in (Pasha, Meherwar, Abdul Manan, & Furrakh, 2017) have applied a set of classification algorithms over a credit cards dataset and reached maximum of 81% accuracy. The same conclusion was previously proposed by (Hamid & Ahmed, 2016) with a different dataset which applied a set of techniques and proved the superiority of J48 algorithm. Moreover, a research in (Keramati, Ghaneei, & Mirmohammadi, 2016) was also presented. The research discussed the success of data mining in detecting the customers' churn by applying the classification techniques with considering their behavior and detect the customers churn factors. Additionally, a research in (Marinakos & Daskalaki, 2017) highlighted the superiority of data mining techniques to deal with imbalanced data. The research proposed a clustering approach which was applied on imbalanced customers' data for segmentation. The performance of the proposed approach has been measured and according to the research's results, the proposed approach reached a higher performance when compared with statistical approaches. Customers' classification is also targeted in detecting fraudulent customers' accounts such as in (Lv, et al., 2019). The research applied convolution neural network mining technique in detecting the illegal account. The contribution of the research is extended in considering different sources of data with the ability to deal with heterogeneous features, however, the authors did not explicitly define their perspective of the heterogeneous data.

Moreover, a dataset was collected by (Shahari, Zakaria, & Rahman, 2015) within the time period of 2005 to 2012 and included forty banks which were distributed over twelve countries targeting to identify new polices in financing customers using data mining. Another research targeted a hybrid data

mining approach in (Chen, Xiang, Liu, & Wang, 2012), the research approach aimed at determining a score for the customers using two main stages, clustering then classification. The first stage applied the clustering technique with excluding the outliers, then the classification stage applied Support Vector Machine algorithm for labeling new customers. The main contribution in this research that the customers' classes covered different levels. Following this classification, risk management plan was developed according to the customer's class. Other researchers investigated the risk management in banking sector such as (Yurdakul, 2014) which highlight the opposite relation between the risk level and the loan amount. Support Vector Machine algorithm is also applied in another research by (Danenas & Garsva, 2015) who proposed a classification model and compared the results of Support Vector Machine algorithm with other algorithms to prove its superiority. Additionally, a research in (Hsieh & Hung, 2010) focused on the data preprocessing step in which it proposed applying a classification technique to raise the overall classification task accuracy, however, it could be considered as two-steps classification which is an overhead operation. Mining textual data has been also considered in banking sector. For example, the research in (Krishna, et al., 2019) aimed at classifying the customers' complaints by applying text mining techniques as well as natural language processing. The research also introduced the collaboration of two algorithms, Naïve Bays and Random Forest, with text analysis methods for messages' classification to be wither extreme or moderate. More recently, another perspective which highlighted the effectiveness of applying business intelligence tools in the banking sectors is presented in (Bošković & Krstić, 2020). The research presents the contribution of business scorecards tool and data envelopment analysis approach in identifying the business goals and track the organizational business progress. The integration between the business scorecards as a tool and the data envelopment analysis as a reliable analytical approach are proposes as a successful schema for business progress evaluation. Another

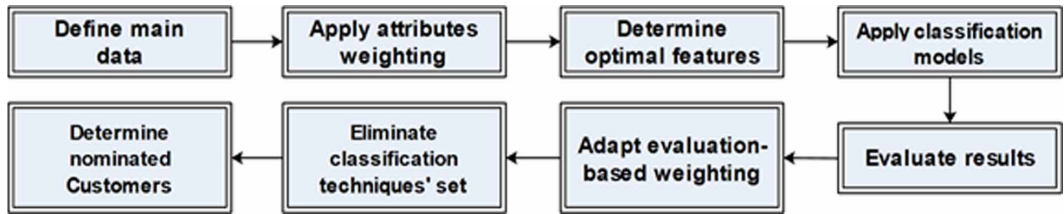
According to the previously discussed research, different approaches introduced the effectiveness in applying data mining techniques in the banking sectors. The collaboration approach has also been introduced and highlighted to be one of the promising directions, however, the perspective of the collaboration was in a narrow perspective and focused only on the technical steps. However, the current research extends the collaboration perspective to also include the algorithm evaluation which highlight the algorithm contribution in reaching accurate decisions. The following subsection demonstrates the proposed model in detail.

THE PROPOSED MODEL FOR CUSTOMERS' PERSONAL LOANS RECOMMENDATION

This research aims at accurately targeting the suitable customers for personal loans granting recommendation. The model assigns the customer to be either eligible for loan granting or not. As highlighted earlier, investigation for the best technique for customers' loans behavior detection is still an open area, therefore, the proposed model provided a different perspective for this investigation by proposing a model which provides the opportunity for different techniques to contribute in the final decision. Each technique will have an associated evaluation weight and accordingly, the final decision is determined using these techniques' weights. The proposed model depends on applying a set of techniques to confirm the model perspective applicability, while it could be further extended for tagging the targeted customers for granting loans recommendations.

The model highlights the strength of providing a collaborative environment among the different techniques for more accurate determination. The proposed model has four main stages, they are: setting the required features, applying the techniques' set, set the evaluation measures, exploring the targeted customers' set with presenting the loan granting status. The following subsections discuss the proposed model stages in detail with illustrating a formal description to the main nodes and steps in these stages. Moreover, while figure 1 illustrates the general steps of the proposed model, the algorithmic steps are demonstrated following the illustrated figure for the detailed model procedure.

Figure 1. A Generic Illustration for the Proposed Model Steps



Stage 1: Explore the Optimal Features

- Define the problem representing the service on focus (P_i)
- Set the customer set members (CUSTOMER)
- Set the service features set members (PROB (P_i))
- Fetch the dataset transactions (DSET (D))
- Set the weighting technique (W_x)
- Determine the weighted features set (PROB'' (P_i))
- Set the weight threshold (TH (F))
- Determine the optimal features' set (OPT_PROB'' (P_i))

Stage 2: Detect the Customers' Preliminary Recommendations

- Set the classification algorithms' set members (CL_Tech)
- Repeat for each classification technique $t_y \hat{=} CL_Tech$
 - Select the classification technique t_y
 - Apply the selected technique (t_y)
- Determine the corresponding recommendations (Pred (t_y))
- End repeat

Stage 3: Detect Final Recommendations in a Collaborative Environment

- Set the evaluation measures' set members (EVAL)
- Fetch the evaluation measures' value set members' weight (P_Eval_Metric)
- Repeat for each classification technique ($t_y \hat{=} CL_Tech$)
- Repeat for each evaluation measure ($eval_{zy} \hat{=} EVAL$)
 - Determine the evaluation measure ($eval_{zy}$)
 - Determine the evaluation measure value (P_Eval_Metric ($eval_{zy}$))
 - Get the evaluation measure value weight (P_z)
 - Set the adapted value measure value ($eval'_{zy}$)
- End repeat
 - Determine the representative classification evaluation value ($f_eval(t_y)$)
- End repeat
- Set the accepted evaluation value threshold ($f_Threshold$)
- Determine the top ranked techniques' set members (Top_Rank_CL_Tech)

// Stage 4: Explore the Targeted Customers Segment

- Repeat for each classification technique ($t_y \hat{=} Top_Rank_CL_Tech$)
 - Set the customers' testing data set (T_CUSTOMER)
 - Apply the classification technique (t_y)
 - Determine the customers' Nominations (Nom_Cust (t_y))
 - Determine c_i nominations status set (Nom_Status (c_i))
 - Determine c_i weighted nominations status set (rank (c_i))
 - Set the Recommendation Threshold (RTH)
 - Determine the final Customers' nominations set (RankedCust)
- End Repeat

In the rest of the chapter, each stage will be described in greater details.

Stage 1: Explore the Optimal Features

The target of this stage is to detect the minimum features' set that is required for exploring the applicable customers for granting loans. The source of this stage is the main customers' dataset. The dataset features could be varied from one organization to another as well as one target to another. Therefore, this stage is a main stage for detecting the highly recommended features for the defined problem. For each defined problem, this stage could be fired once, then the final features' set could be then provided for further investigation to the same problem. However, the stage could be re-fired in case that the organization has a data set with different features. In this research, a recommendation with the most suitable features for the defined target will be provided through the experimental study. The features' set will be determined according to the influence degree of these features which consequently affect the recommendation accuracy. For more clarification, a formal description of the main stage aspects is described as follows:

The set of services of the corporation on focus can be described as follows:

$$\text{PROB} = \{P_1, P_2, \dots, P_i\} \text{ where } i > 0 \quad (1)$$

The set of the corporation customers who are initially issued a loan granting request or are targeted to be offered a loan are described as:

$$\text{CUSTOMER} = \{C_1, C_2, \dots, C_n\} \text{ where } n > 0 \quad (2)$$

The original features' set for each problem that describes the personal and financial profile of the customers; that belong to the segment on focus; can be represented as a set of features as follows:

$$\text{PROB}(P_i) = \{F_1, F_2, \dots, F_j\} \quad (3)$$

Where $j > 0$, F_1 values are unique feature for each customer. That is: $\forall c_q, c_v \in \text{CUSTOMER} \rightarrow \text{Value}(c_q, F_1) \neq \text{value}(c_v, F_1)$

The dataset transactions represent the loans' request / offer. Each transaction corresponds to a single request / offer. The transactions' set can be described as follows:

$$\text{DSET}(D) = \{Tr_1, Tr_2, \dots, Tr_n\} \text{ where } n > 0 \quad (4)$$

Each transaction is described by the values' set of the describing attributes. that is represented as follows:

$$Tr_n = \{<F_1, \text{val}_1>, <F_2, \text{Val}_2>, \dots, <F_j, \text{Val}_j>\} \text{ where } J > 0, |Tr_n| = J \quad (5)$$

Moreover, each customer who is a member in the CUSTOMER set has a corresponding transaction, which is a member in DSET (D) dataset, that is represented as follows:

$$\forall c_q \in \text{CUSTOMER} \rightarrow \exists Tr_q \in \text{DSET}(D) \text{ where } q > 0 \quad (6)$$

Formula 6 illustrates that the customers who belong in the customers' set owns one transaction in the transactions' dataset.

Therefore, the set of all problems with their associated features' set PROB' is described as follows:

$$\text{PROB}^{\setminus} = \{ \langle P_1, \{F_{11}, F_{12}, \dots, F_{1j}\} \rangle, \dots, \langle P_i, \{F_{i1}, F_{i2}, \dots, F_{ij}\} \rangle \} \quad (7)$$

$$\text{PROB}^{\setminus} = \bigcup_{k=1}^i \langle P_k, \{F_{k1}, F_{k2}, \dots, F_{kj}\} \rangle$$

Where $i, j, k > 0$

After applying the weighting measure on the training dataset transactions, then, each feature will have an associated weight which represents its influence degree. So, the members of the features' set for a defined problem with their associated influence measure can be represented formally as follows:

$$\text{PROB}''(P_i) = \{ \langle F_1, W_1 \rangle, \langle F_2, W_2 \rangle, \dots, \langle F_j, W_j \rangle \} \text{ where } i, j > 0 \quad (8)$$

The final features' set of problem P_i which is considered the most representative are the features that have an influence degree above the threshold, it is formally described as follows:

$$\text{OPT_PROB}''(P_i) = \{ \langle F_p, W_p \rangle, \langle F_d, W_d \rangle, \dots, \langle F_m, W_m \rangle \} \quad (9)$$

where $\text{OPT_PROB}''(P_i) \hat{=} \text{PROB}''(P_i)$

$j \geq 1, d, m > 0, W_p, W_d, W_m > \text{Weight_TH}, \text{Weight_TH} = 0.5$

Stage 2: Detect the Customers' Preliminary Recommendations

In this stage, a set of classification techniques are targeted for participating in the collaborative approach. The main idea is to identify the techniques' set members, then apply these techniques on the prepared dataset. This stage has been proposed to participate in identifying the targeted customers for loans granting recommendations. As it is one of the services which can be categorized as high risk, therefore, the research challenge that considering the contribution of more than one algorithm with the respect of each algorithm's associated weight leads to a higher performance in the recommendation process.

While in this research, the classification techniques' set members are determined during the experimental study, however, the techniques' set members could be further extended to include more available techniques.

The set of contributed classification techniques are described as follows:

$$\text{CL_Tech} = \{t_1, t_2, \dots, t_y\} \text{ where } y > 0 \quad (10)$$

For each classification technique, the recommendations that are resulted by applying the techniques' set members on the testing dataset transactions are documented. Each record representing a loan granting request from a defined customer earns a recommendation to be either accepted or rejected. Therefore, each technique with its associated recommendation for the each of the dataset transactions is theoretically represented as follows:

$$\text{Pred}(t_y) = \{ \langle \text{Tr}_1, \text{pred}_1 \rangle, \langle \text{Tr}_2, \text{pred}_2 \rangle, \dots, \langle \text{Tr}_c, \text{pred}_c \rangle \} \quad (11)$$

The set of all techniques associated with their labeled dataset is theoretically represented as follows:

$$\text{PRED} = \bigcup_{k=1}^y \{ \langle t_y, \text{Pred}(t_y) \rangle \} \text{ where } k, y > 0 \quad (12)$$

Formula 12 illustrates that the PRED set is the mother of all predictions' set. It includes a set of two-members' vectors. Each vector represent the applied technique with its associated prediction results.

Stage 3: Detect Final Recommendations in a collaborative environment

In this stage, the evaluation of each classification algorithm is performed according to the algorithms' recommendation results. The evaluation criteria set members are determined and then the classification algorithms have been evaluated against these measures. Each algorithm will have a value for each measure. Additionally, it is also applicable that the evaluation criteria set could be further extended targeting for more accurate comparison.

Formally describing the evaluation criteria set could be described as follows:

$$\text{EVAL} = \{ \text{eval}_1, \text{eval}_2, \dots, \text{eval}_z \} \text{ where } z > 0 \quad (13)$$

The set of evaluation criteria values for each of the classification techniques is illustrated as follows:

$$\text{Eval_CL_Tech}(t_y) = \{ \langle \text{eval}_{1y}, \text{val}_{1y} \rangle, \langle \text{eval}_{2y}, \text{val}_{2y} \rangle, \dots, \langle \text{eval}_{zy}, \text{val}_{zy} \rangle \} \quad (14)$$

where $z, y > 0, t_y \in \text{CL_Tech}, |\text{Eval_CL_Tech}(t_y)| = |\text{EVAL}|, 0 > = \text{val}_{zy} > = 100$ (as val_{zy} represents the evaluation measure percentage).

The parent set that includes members representing all the evaluated classification algorithms associated with their evaluation criteria corresponding values is described as follows:

$$\text{Eval_CL_Tech} = \sum_{k=1}^y \text{Eval_CL_Tech}(t_y) \text{ where } k, y > 0 \quad (15)$$

The next step is to determine the contributed classification algorithm's performance. This step is performed by considering the values of the algorithms' granted evaluation measures. Then, the research follows the proposed model in (Idrees & Alsherif, 2020) (Idrees & Alsherif, 2020) for considering these values as weights to the contributed algorithms. In more details, although the proposed method in the research of Idrees and Alsherif (Idrees & Alsherif, 2020) (Idrees & Alsherif, 2020) focused on the evaluation metrics ranking, however, in this research, the scope of applying the proposed method in (Idrees & Alsherif, 2020) (Idrees & Alsherif, 2020) is enlarged targeting to explore the most suitable classification algorithm results. Each evaluation measure value is multiplied with its weight to reveal the updated measure value for the algorithm. Applying this step ensures that each measure value reflects the right contribution in evaluating the classification algorithm. The metrics ranking has been determined according to the variance level of the metric. The research of presented that the metric with less non-variance is more accurate in the classification tasks. This conclusion is presented according to different research. The metrics' variance is measured according to the ability of the item to move from true/false negative to true/false positive and vice versa. Five invariance measures presents these relations and monitor these item's transition status between them.

Following the previous description, the rule of ranking the classification algorithm which is proposed in (Idrees & Alsherif, 2020) can be re-formulated as follows:

$$\forall t_y \in CL_Tech, \exists Eval_CL_Tech(t_y)$$

$$\forall eval_z \in P_Eval_Metric(eval_z), eval_z = \langle eval_z, p_z \rangle \rightarrow eval'_z = eval_z \times p_z$$

$$f_eval(t_y) = \left(\sum_{s=1}^j eval'_z \right) / j \text{ Where } y, j, s, z \in N, y, j, s, z > 0 \quad (15)$$

Formula 15 describes how each algorithm evaluation result is performed. For each of the contributing algorithms, each evaluation measure's result is weighted by its weighting value, then the overall algorithm evaluation is determined by the average of all criteria values.

Additionally, the set of final ranking for all the classification algorithms can be described as follows:

$$F_Eval(CL_Tech) = \bigcup_{i=1}^y \{ \langle t_y, f_{eval(t_y)} \rangle \} \text{ where } y \in N, y > 0 \quad (16)$$

Then the set that includes the of final ranking for all the classification algorithms that earned a rank above the threshold is determined, it can be described as follows:

$$ACC_F_Eval(CL_Tech) = \{ \langle t_r, f_{eval(t_r)} \rangle, \langle t_b, f_{eval(t_b)} \rangle, \dots, \langle t_d, f_{eval(t_d)} \rangle \} \quad (17)$$

Where $r, b, d \in N, y > r, b, d > 0, \langle t_y, f_{eval(t_y)} \rangle \in F_Eval(CL_Tech)$,

$$f_{eval(t_y)} > f_Threshold, ACC_F_Eval(CL_Tech) \subseteq F_Eval(CL_Tech)$$

Finally, according to the members of the $ACC_F_Eval(CL_Tech)$ set, then the selected classification algorithms which will be considered for customers' ranking are determined. Therefore, the set of final algorithms that are targeted for the final recommendation decision contribution is described as follows:

$$Top_Rank_CL_Tech = \{ t_a, t_r, \dots, t_f \} \quad (18)$$

Where $a, r, f \in \{1, \dots, y\}$ and $Top_Rank_CL_Tech \subseteq CL_Tech$

$$\exists \langle t_x, f_{eval(t_x)} \rangle \in ACC_F_Eval(CL_Tech), x = a \text{ r l f}$$

Stage 4: Set the Targeted Customers' Set

The final stage targets to explore the set of customers who are suitable to be granted the requested loans. In other words, in this step, the loan granting recommendation is provided to the decision maker to be either accepted or rejected. The loans' requests are gathered representing the testing dataset. Then the set of algorithms are determined according to the problem type and the defined features' set. The selected algorithms are the algorithms which have a rank that is above the determined threshold for the problem with the same description.

Following the dataset description in equation 1, it is hypothesized that each transaction describes the personal and financial profile for one of the customers. It is also hypothesized that each customer can only issue one loan request which means that each customer has only one corresponding record describing his situation. Therefore, the output of this step is the set of customers as loan nominators which is a subset of the original dataset and is recommended to be granted a loan. It is worth highlighting that the output of this stage is only to accept or reject the loan with no determination of the risk percentage, however, this objective could be further investigated.

In general, the nominated customers' set for loan granting can be represented as follows:

$$\text{Nom_Cust} = \{C_f, C_r, \dots, C_z\} \text{ where } f, r, z \in \{1, \dots, n\}, \text{Nom_Cust} \subseteq \text{CUSTOMER} \quad (19)$$

Therefore, each classification algorithm in the top ranked list Top_Rank_CL_Tech has then an associated loan nominated customers. The set of algorithms with their associated nominated customers can be represented as follows:

$$\text{Class_Cust} = \{ \langle t_y, \text{Nom_Cust}(t_y) \rangle \text{ where } t_y \in \text{Top_Rank_CL_Tech} \quad (20)$$

After applying the top ranked classification techniques and determining if the customer is nominated or not, then each customer will have a set of attributes which is equal to the top ranked classification algorithms. Each of these attributes represent the classification algorithm and its associated rank. Then the result of this record for each algorithm is recorded as an attribute with respect to the algorithm rank. This step can be described as follows:

$$\forall c_i \in \text{CUSTOMER}, \text{Nom_Status}(c_i) = \{\text{Nom}_{a_i}, \text{Nom}_{r_i}, \text{Nom}_{f_i}\} \quad (21)$$

Where: Nom_{a_i} is the nomination status for c_i using algorithm t_a ,

$$t_a \in \text{Top_Rank_CL_Tech}, a, r, f, i > 0$$

$\text{Nom}_a = \{-1, 1\}$ where -1 indicates that the customer is not nominated and 1 indicates that the customer is nominated $\forall c_i \in \text{CUSTOMER}, \forall t_y \in \text{Top_Rank_CL_Tech} \rightarrow \text{rank}(c_i, t_y) = \text{Nom}_{y_i} * r_y$

Then, calculating final ranking for each customer to be either negative or positive follows the following rule:

$$\forall c_i \in \text{CUSTOMER}, \forall t_y \in \text{Top_Rank_CL_Tech}$$

$$\text{Ranked_Status}(c_i) = \frac{\sum_{d=1}^y \text{rank}(c_i, t_d)}{| \text{Top_Rank_CL_Tech} |} \quad (22)$$

The ranking of each customer is determined as the average of the ranking that the customer earned from all the algorithms.

The final step is detecting if the final recommended decision for the customer c_i is either a nomination for loan granting if the customer earned a positive value above the threshold, otherwise, the customer is not recommended for loan granting. This step process follows the following rule:

- if $\text{Ranked_Status}(c_i) < 0$ then c_i is not recommended (illustrated as $\text{Final_Rank}(c_i) = -1$)
- else If $\text{Ranked_Status}(c_i) < 0.5$ then c_i is not recommended (illustrated as $\text{Final_Rank}(c_i) = -1$)

else c_i is recommended (illustrated as $\text{Final_Rank}(c_i) = +1$) (23)

EXPERIMENTAL CASE STUDY

This section discusses the applied experiments which confirm the applicability of the proposed model.

Dataset

The experimental case study is applied on a dataset from the “Dream Housing Finance company” which specialty in house loans. The dataset is uploaded on the GitHub website as a benchmark and is available to use (Loan Prediction Dataset, 2018) for public use. The dataset transactions describe the customer’s data with an associated decision if his loan request is accepted or rejected. The original dataset included thirteen attributes, however, following the proposed steps in the first stage, only twelve attributes are included describing the customer’s personal and financial profile. The personal profile included the customer’s marital status, if there are any dependents as children or parents, the education level and the customer’s gender. The financial profile included the customer’s average income, if he owns his workplace or is an employee, the total required loan amount and the installment per month according to this amount and the interest percentage, a history of his credit score indicating his payment regularity, the area code that the customer’s house exists in. only one attribute is eliminated, it is the income of the guaranteed person which highlighted to have a weigh below the threshold. Finally, each transaction includes an attribute (Loan status) indicating if the loan has been granted or not. This property is used as a label for the training data set and as a validation label in the testing data set. Table 1 illustrates a brief description of the contributing attributes.

Table1. Brief Description of the Contributed Attributes

	Feature Name	Description	Type
1	ID	Describe a unique ID for each loan transaction	Continuous
2	Customer Gender	Male/Female	Discrete
3	Customer Martial Status	Married/ Not Married	Discrete
4	Has Dependents	Count of the dependents (0, 1, 2, more than 2)	Discrete
5	Education Level	Graduate / Not Graduate	Discrete
6	Own Workplace	Yes / No	Discrete
7	Average Income	Amount of income	Continuous
8	The Current Loan Amount	Total amount of the loan	Continuous
9	The Loan Duration	duration of the granted loan in months	Continuous
10	Credit	The history of the customer in his bank credits	Discrete
11	Property Area	The district type of the customer’s home	Discrete
12	Loan Status	State if the loan is approved or not	Discrete

The dataset included 614 records for customers’ loan requests. Table 2 illustrates a sample of the dataset. Statistical measures are also presented in table 3 and table 4. Table 3 illustrates the count of discrete attributes while table 4 illustrates the maximum, minimum, mean, number of records above mean, number of records below mean, and the standard deviation for the continuous valued attributes.

Table 2. Sample of Dataset

ID	Customer Gender	Customer Martial Status	Has Dependents	Education Level	Own Workplace	Average Income	The Current Loan Amount	The Loan Duration	Credit	Property Area	Loan Status
LP001002	Male	No	0	Graduate	No	5849	100	360	1	Urban	Y
LP001003	Male	Yes	1	Graduate	No	4583	128	360	1	Rural	N
LP001005	Male	Yes	0	Graduate	Yes	3000	66	360	1	Urban	Y
LP001006	Male	Yes	0	Not Graduate	No	2583	120	360	1	Urban	Y
LP001008	Male	No	0	Graduate	No	6000	141	360	1	Urban	Y
LP001011	Male	Yes	2	Graduate	Yes	5417	267	360	1	Urban	Y

Table 3. Statistical Illustration for the Profile's Discrete Attributes

Feature Name	Value	Count
Customer Gender	Male	489
	Female	125
Customer Martial Status	Married	398
	Not Married	216
Has dependents	0	345
	1	102
	2	102
	Above 2	65
Education Level	Graduate	480
	Not Graduate	134
Own Workplace	Yes	81
	No	533
Property Area	Urban	202
	Rural	233
	Semiurban	179
Credit	Yes	475
	No	139
Granted the Loan	Yes	422
	No	192

Table 4. Statistical Illustration for the Profile's Continuous Attributes

Fzeature Name	Maximum value	Minimum Value	Median	Mean	No of Customers below Mean	No of Customers above Mean	Standard Deviation
Average Income	81000	150	3812.5	5403.5	438	176	6109.5
The current Loan Amount (in Thousand)	700	9	128	150.4	392	222	86.6
Loan Duration	480	12	360	340	87	527	65.7

Methods

In this research, seven algorithms have been applied in this experiment, they are Decision Tree, Random Forest, Naïve Bayes, Logistic Model Tree (LMT), K Nearest Neighbor, Support Vector Machine, and Iterative Dichotomiser 3 (ID3) algorithms. The contributing algorithms were determined to be the most widely used with a variety in the accuracy. Different researches have highlighted the accuracy variation such as in (Bach, Šarlija, Zoroja, Jaković, & Ćosić, 2019) (Sen, Hajra, & Ghosh, 2020) which presented a literature for classification algorithms and highlighted that contributed algorithms.

On the other hand, The contribution of the evaluation measures is considered one of the main pillars in the proposed model, therefore, the set of evaluation measures that have been proposed in this research followed the research in (Idrees & Alsherif, 2020) which proposed a method to determine a weight each evaluation measure that reflects its accurate evaluation level of the classification algorithms. The set of evaluation measures are accuracy, precision, recall, f-measure, specificity, YOUDEN, and error rate. Each of these measures highlight one of the evaluation perspectives as illustrated in table 5, therefore, considering the results of each measure would provide higher reliable recommendations. The following section discusses the applied classification algorithms' evaluation based on the weighted evaluation metrics.

Table 5. Evaluation Metrics Description (Idrees & Alsherif, 2020)

Evaluation Metric	Definition
Accuracy	The percentage indicating the level of successful classification
Precision	The ratio between the number of customers that are classified to the correct class and the number of all customers that determined to belong to that class.
recall	The ratio between the number of customers that are classified to the correct class and the number of all customers that determined to belong to their correct class
F measure	A relation of precision and recall ($2 * \text{Precision} * \text{recall} / (\text{precision} + \text{recall})$)
AUC	The probability of a correctly classified positive customer and a correctly classified negative customer
Youden	sensitivity + specificity - 1
Sensitivity	The percentage of setting the accepted customer classification correctly
Specificity	The percentage of setting the rejected customer classification correctly
Error Rate	The percentage of occurring errors

Table 6. Evaluation Results for the Classification Algorithms in percentage

	Decision Tree	Random Forest	Naïve Bayes	LMT	KNN	SVM	ID3
accuracy	80.94	69.06	98.37	80.94	98.75	95.28	97.88
precision	92.13	75	95.05	92.13	100	100	100
recall	42.71	1.56	100	42.71	93.20	84.9	93.23
F_measure	58.36	3.06	97.46	58.36	96.5	91.83	96.5
specificity	98.34	99.76	97.63	98.34	100	100	100
YOUDEN	0.41	0.013	0.976	0.41	0.97	0.849	0.932
Error Rate	19.06	30.94	1.63	19.06	1.25	4.72	2.12

Table 7. Evaluation Metrics Weight (Idrees & Alsherif, 2020)

	Weight	%
Accuracy	4	0.8
Precision	3	0.6
recall	3	0.6
F measure	4	0.8
AUC	5	1
Youden	5	1
Sensitivity	3	0.6
Specificity	4	0.8
Error Rate	5	1

Table 8. Final Evaluation Results for the Classification Algorithms with the Evaluation Measures' Weight Consideration

Measure (Weight)	Decision Tree	Random Forest	Naïve Bayes	LMT	KNN	SVM	ID3
Accuracy (0.8)	64.752	55.248	78.696	64.752	79	76.224	78.304
Precision (0.6)	55.278	45	57.03	55.278	60	60	60
Recall (0.6)	25.626	0.936	60	25.626	55.92	50.94	55.938
F_measure (0.8)	46.688	2.448	77.968	46.688	77.2	73.464	77.2
Specificity (0.8)	78.672	79.808	78.104	78.672	80	80	80
YOUDEN (1)	41	1.3	97.6	41	97	84.9	93.2
Error Rate (1)	19.06	30.94	1.63	19.06	1.25	4.72	2.12
Average	47.29657	30.81143	64.43257	47.29657	64.3857	61.464	63.82314

RESULTS

The classification algorithms are applied on the house loans dataset and the evaluation for each algorithm has been performed individually according to the contributing evaluation metrics, these results are illustrated in table 6. As previously mentioned, each evaluation measure owns a contributing weight in measuring the algorithm performance as proposed in (Idrees & Alsherif, 2020), these weights are illustrated in table 7. Consequently, the evaluation results of the classification algorithms are then updated according to the evaluation measure’s weight. This adaptation in the classification algorithms’ evaluation have been illustrated in table 8 while the average evaluation percentages are illustrated in figure 2 for more clarification.

Figure 2. Average Evaluation Results for the Classification Algorithms

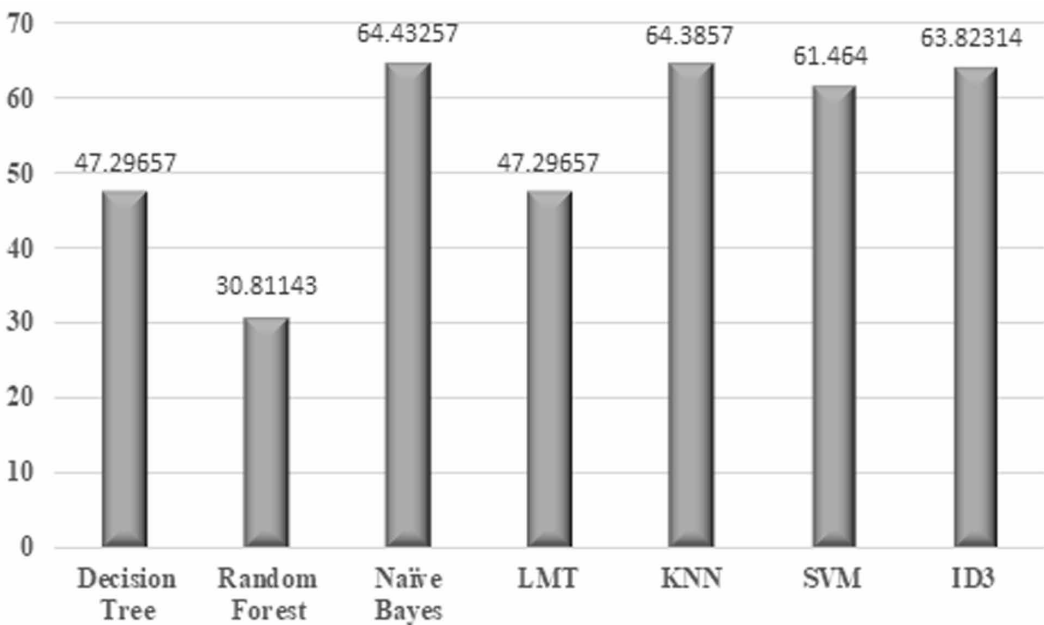


Table 8 highlights the final evaluation for each classification algorithm. The evaluation metrics have been considered and the measure’s result with respect to each algorithm has been adapted according to the measure’s weight. For example, the accuracy of the Decision Tree algorithm has been adapted from 80.94 as mentioned in table 5 to be $(80.94 * 0.8 = 64.752)$ as mentioned in table 8. Following equation 15, the average evaluation for each classification algorithm has been calculated and the final evaluation for each of the classification algorithms have been determined as illustrated in table 7. It is worth mentioning that the final evaluation percentage follows a defined threshold, so, the results of the algorithms that earned an evaluation below the threshold are excluded and their classification was not considered in the final recommendation step.

According to the illustrated results, the classification algorithms which will be considered are those which average validation is above 50%, they are Naïve Bayes, KNN, SVM, and ID3. The next step is generating a dataset including the customer ID and the prediction of each algorithm with respect to its validation weight, respectively. Finally, the rank for each customer is calculated and the decision recommendation is provided according to the final customer nomination. Each customer earns four individual recommendations, one from each algorithm. The algorithm recommendation

has two aspects, the classification algorithm evaluation percentage with an associated sign to be either positive or negative. The positive sign indicates that the loan recommendation is accepted while the negative sign indicates that it is not accepted. Finally, the final decision is recommended according to the average value for all algorithms. In case that the average is negative, then the decision is rejecting the loan request, in case that the average is positive, then the loan is only accepted in case that the value is above 50%.

Model Utilization

Table 9 presents a sample of the calculated ranking for each customer, then a final decision is reached according to the algorithms' contribution in the ranking step as illustrated in table 8. The proposed model results have been monitored and linearly measured against the results' accuracy which reached 92% when compared with the original decisions that are provided by the experts. These original decisions are illustrated in the dataset source. The illustrated results reveal that the proposed model could be utilized as a decision support model to minimize the risk level of granting loans in banking sectors. The authors argue that the proposed model could provide the decision maker with a wider perspective with the customers' segment that could be accepted for loans' granting. Moreover, speeding up the selection process of the accepted customers for loan granting could be highlighted, as embedding the proposed model in the customers evaluation process could provide the decision maker with a higher perspective for the customer on examination in a less time compared with the normal evaluation process. Although the proposed model is applied on a loan granting dataset, however, the authors argue that the model could also successfully contribute to other tasks such as credit cards granting and financial investment requests.

Table 9. A Sample of the Customers' Requests Decision

Loan_ID	ID3	SVM	NB	KNN	Final Nomination	Decision
LP001002	-63.8231	-61.464	-64.4326	65.71429	-31.0014	No
LP001003	-63.8231	-61.464	-64.4326	-65.7143	-63.8585	No
LP001005	63.82314	61.464	64.43257	65.71429	63.8585	Yes
LP001006	63.82314	61.464	64.43257	65.71429	63.8585	Yes
LP001008	63.82314	61.464	64.43257	65.71429	63.8585	Yes
LP001011	63.82314	61.464	64.43257	65.71429	63.8585	Yes

CONCLUSION

Supporting financial institutions with intelligent techniques could provide a positive impact on the participated stakeholders despite the difference in their requirements. Retaining the institutions' resources not only it preserves its economic situation, but it also provides the service to the most suitable customer. Loans granting is classified as a one of the highest risk services, therefore, proposing a successful contribution for intelligent techniques could become a supportive pillar to the financial institutions. The main aim is a successful classification of the customers who applied for loan granting. A direct classification model could positively contribute in the required task, however, this research opened the argument of the possibility of providing a contribution of a set of models in order to raise the accuracy in the classification task which in turn minimize the risk level.

In this research, a data mining-based model is proposed which aims at recommending the customers' segment for loans granting according to a set of criteria which describe the customer's situation. The research introduces a set of classification mining techniques as well as a set of evaluation criteria as two main sets in the model base. Then applying the classification mining techniques set members is performed. The contributing algorithms are evaluated by the evaluation criteria set members to determine the algorithms' performance level. The evaluation criteria are determined with considering the weight of each of these criteria as the criteria weight reflects the contribution of this criteria in the evaluation process. Finally, the evaluation results are analyzed in a collaborative perspective to identify the situation of the customers' requests.

The proposed approach has been applied on the "Dream Housing Finance company" which specialty is in house loans. The dataset is uploaded on the GitHub website as a benchmark and is available to use. The dataset included 614 loans granting requests, while it was characterized by twelve attributes. these attributes describe the customer personal profile in six attributes such as the gender, workplace ownership, the number of dependents, and others. Moreover, the remaining of the attributes describe the financial situation of the customer including the loan amount, the loan duration, and others. It is noticed that more than 79% of the customers were Males. The same percentage represents the education level of the customers while more than 64% are married. Moreover, 50% of the customers had no dependents. The highest percentage was that more than 86% of the customers do not own their workplace while the customers have an acceptable distribution with respect to their property area. The dataset included 422 successful requests in granting the requested loans while 192 requests are refused.

In this research, seven of the most widely classification mining algorithms have been applied in the experiment, they are Decision Tree (DT), Random Forest (RF), Naïve Bayes (NB), Logistic Model Tree (LMT), K Nearest Neighbor (KNN), Support Vector Machine (SVM), and Iterative Dichotomiser 3 (ID3) algorithms. On the other hand, the set of evaluation measures that have been considered were accuracy, precision, recall, f-measure, specificity, YOUDEN, and error rate. The selection of the contributing algorithms as well as the evaluation criteria follows recent researches. The proposed model reached a final accuracy percentage above 92% in its main target with successful recommendation of the loan granting customers' requests. Consequently, the proposed model is considered a decision support model which could successfully contribute to the financial institutions for having accurate loan granting recommendations with a limited time frame. The authors argue that the proposed model does not only contribute to accurate recommendations, but the time frame could also be a contribution compared with the traditional customer's evaluation method.

While different researches have discussed the loan granting subject, however, these researches focused on applying a determined algorithm. In this research, the collaboration perspective between a set of classification algorithms is proposed as well as considering the weighting evaluation criteria. The proposed model idea has emerged from the fact that each algorithm has its strength points and weaknesses, therefore, the proposed collaboration perspective could hinder the weaknesses of these contributed algorithms. In addition, as evaluation metrics traditional perspective was the evaluation of the applied algorithm with no contribution to the main goal, however, in this research, considering the evaluation metrics weight is argued to reach more accurate recommendations.

Although the accuracy percentage is considered acceptable, however, investigating in considering more criteria would further contribute to raise the accuracy level. Another direction is the possible contribution of more algorithms which could enrich the process. Moreover, social data is a rich source that could also provide a positive impact in the same direction. Another mining perspective could be further provided in applying other techniques' set such as associations rules mining which could further highlight the most contributing descriptive attributes for the recommendation task. Finally, and more widely, an enrichment approach could be proposed to continuously feed the model base with successful members in each contributing set.

REFERENCES

- Abdel-Fattah, M. A., Khedr, A. E., & Nagm Aldeen, Y. (2017). An Evaluation Framework for Business Process Modeling Techniques. *International Journal of Computer Science and Information Security*, 15(5), 382–392.
- Abogabal, F., Ouf, S. M., Idrees, A. M., & Khedr, A. E. (2020). An Architectural Framework For Generating Food Safety Key Performance Indicators. *Journal of Southwest Jiaotong University*, 55(5), 5. doi:10.35741/issn.0258-2724.55.5.5
- Al Mazroi, A., Khedr, A. E., & Idrees, A. M. (2021). A Proposed Customer Relationship Framework based on Information Retrieval. *Expert Systems with Applications*, 176.
- Awad, T. A., & Dessouki, Y. S. (2017). Mobile Banking Roll-out in Egypt: Antecedences of Mobile Banking Adoption. *International Journal of E-Services and Mobile Applications*, 9(2), 1–22. doi:10.4018/IJESMA.2017040101
- Bach, M. P., Vlahović, N., & Pivar, J. (2020). Fraud Prevention in the Leasing Industry Using the Kohonen Self-Organising Maps. *Organizacija*, 53.
- Bach, M. P., Šarlija, N., Zoroja, J., Jaković, B., & Ćosić, D. (2019). Credit Risk Scoring in Entrepreneurship: Feature Selection. *Managing Global Transitions*, 17(4), 265–287.
- Bach, P. M., Dumičić, K., Žmuk, B., Ćurlin, T., & Zoroja, J. (2020). Data mining approach to internal fraud in a project-based organization. *International Journal of Information Systems and Project Management*, 8(2).
- Baixauli, J., Alvarez, S., & Módica, A. (2012). Combining structural models and accounting based models for measuring credit risk in real estate companies. *International Journal of Managerial Finance*, 8(1), 73–95. doi:10.1108/17439131211201040
- Bayomy, N. A., Abd-Elmegid, L. A., Khedr, A. E., & Idrees, A. M. (2021). A Literature Review for Contributing Mining Approaches for Business Process Reengineering. *Future Computing and Informatics Journal*, 5(2), 60–78. doi:10.54623/fue.fcij.5.2.1
- Bošković, A., & Krstić, A. (2020). The Combined Use of Balanced Scorecard and Data Envelopment Analysis in the Banking Industry. *Business Systems Research*, 11(1), 4–15. doi:10.2478/bsrj-2020-0001
- Carmona, P., Climent, F., & Momparler, A. (2019). Predicting failure in the U.S. banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*, 61, 304–323. doi:10.1016/j.iref.2018.03.008
- Chen, W., Xiang, G., Liu, Y., & Wang, K. (2012). Credit risk evaluation by hybrid data mining technique. *Systems Engineering Procedia*, 3, 194–200. doi:10.1016/j.sepro.2011.10.029
- Danenas, P., & Garsva, G. (2015). Selection of Support Vector Machines based classifiers for credit risk domain. *Expert Systems with Applications*, 42(6), 3194–3204. doi:10.1016/j.eswa.2014.12.001
- El Seddawy, A. B., Sultan, T., & Khedr, A. E. (2013). A Proposed Data Mining Technique to Improve Decision Support System in an Uncertain Situation. *International Journal of Engineering Research and Development*, 8(7), 56–61.
- El Seddawy, A. B., Sultan, T., & Khedr, A. E. (2013). Enhanced K-mean Algorithm to Improve Decision Support System under Uncertain Situations. *International Journal of Computer Science and Network Security*, 13(7).
- Farooqi, R., & Iqbal, N. (2019). Performance evaluation for competency of bank telemarketing prediction using data mining techniques. *International Journal of Recent Technology and Engineering*, 8(2), 5666–5674.
- Hamid, A. J., & Ahmed, T. M. (2016). Developing Prediction Model Of Loan Risk In Banks Using Data Mining. *Machine Learning and Applications: An International Journal*, 3(1), 1–9. doi:10.5121/mlaij.2016.3101
- Hassan, H. A., Dahab, M. Y., Bahnassy, K., Idrees, A. M., & Gamal, F. (2015). Arabic Documents Classification Method a Step towards Efficient Documents Summarization. *International Journal on Recent and Innovation Trends in Computing and Communication*, 3(1), 351–359. doi:10.17762/ijritcc2321-8169.150171
- Hassan, H. A., Dahab, M. Y., Bahnassy, K., Idrees, A. M., & Gamal, F. (2014). Query answering approach based on document summarization. *International Open Access Journal of Modern Engineering Research*, 4(12).

- Hassan, H. A., & Idrees, A. M. (2010). Sampling Technique Selection Framework for Knowledge Discovery. *7th International Conference on Informatics and Systems (INFOS)*, 1 - 8.
- Hassouna, D. H., Khedr, A. E., Idrees, A. M., & ElSeddawy, A. I. (2020). Intelligent Personalized System for Enhancing the Quality of Learning. *Journal of Theoretical and Applied Information Technology*, 98(13), 2199–2213.
- Helmy, Y., Emam, O., Khedr, A., & Bahloul, M. (2020). A Survey on Effect of KPIs in Higher Education based on Text Mining Techniques. *International Journal of Scientific and Engineering Research*, 11(3).
- Helmy, Y., Khedr, A. E., Kolief, S., & Haggag, E. (2019). An Enhanced Business Intelligence Approach for Increasing Customer Satisfaction Using Mining Techniques. *International Journal of Computer Science and Information Security*, 17(4).
- Hsieh, N., & Hung, L. (2010). data driven ensemble classifier for credit scoring analysis. *Expert Systems with Applications*, 37(1), 534–545. doi:10.1016/j.eswa.2009.05.059
- Idrees, A. M., & Alsherif, F. K. (2020). A Collaborative Evaluation Metrics Approach for Classification Algorithms. *Journal of Southwest Jiaotong University*, 55(1), 1–14. doi:10.35741/issn.0258-2724.55.1.1
- Idrees, A. M., El Seddawy, A. I., & Moaaz, E. L. (2019). A Proposed Mining Based Business Continuity Information System for Educational Institutes. *Journal of Computational Science*, 15(8), 1133–1149. doi:10.3844/jcssp.2019.1133.1149
- Idrees, A. M., ElSeddawy, A. I., & Zeidan, M. O. (2019). Knowledge Discovery based Framework for Enhancing the House of Quality. *International Journal of Advanced Computer Science and Applications*, 10(7), 324–331. doi:10.14569/IJACSA.2019.0100745
- Idrees, A. M., & Ibrahim, M. H. (2018). A Proposed Framework Targeting the Enhancement of Students' Performance in Fayoum University. *International Journal of Scientific and Engineering Research*, 9(11).
- Idrees, A. M., Ibrahim, M. H., & El Seddawy, A. I. (2018). Applying spatial intelligence for decision support systems. *Future Computing and Informatics Journal*, 3, 384-390.
- Ilham, A., Khikmah, L., Indra, A., Ulumuddin, A., & Iswara, I. (2019). Long-term deposits prediction: A comparative framework of classification model for predict the success of bank telemarketing. *Journal of Physics: Conference Series*, 1175(1), 1–6. doi:10.1088/1742-6596/1175/1/012035
- Jing, Z., & Fang, Y. (2018). Predicting US bank failures: A comparison of logit and data mining models. *Journal of Forecasting*, 37(2), 235–256. doi:10.1002/for.2487
- Keramati, A., Ghaneei, H., & Mirmohammadi, S. (2016). Developing a prediction model for customer churn from electronic banking services using data mining. *Financial Innovation*, 2(1), 1–13. doi:10.1186/s40854-016-0029-6
- Khedr, A. E., Abdel-Fattah, M. A., & Nagm-Aldeen, Y. (2015). A Literature Review of Business Process Modeling Techniques. *International Journal of Advanced Research in Computer Science and Software Engineering*, 5(3), 43–47.
- Khedr, A. E., & Borgman, H. (2006). Adoption of New Technologies in a Highly Uncertain Environment: The Case of Egyptian Public Banks. *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, 223-227.
- Khedr, A. E., & Borgman, H. (2006). Processes and Obstacles in Adoption of New Technologies in Uncertain Environment: The Case of Egyptian Public Banks. In *International Arab Conference on Information Technology (ACIT)*. Amman, Jordan: Al Yarmouk University.
- Khedr, A. E., Idrees, A. M., & El Seddawy, A. I. (2016). Enhancing Iterative Dichotomiser 3 algorithm for classification decision tree. *WIREs Data Mining Knowledge Discovery*, 6(2), 70–79. doi:10.1002/widm.1177
- Khedr, A. E., Idrees, A. M., & Shabaan, E. (2020). Automated Ham-Spam Lexicon Generation Based on Semantic Relations Extraction. *International Journal of e-Collaboration*, 16(2), 45–64. doi:10.4018/IJeC.2020040104
- Khedr, A. E., & Kok, J. (2006). Adopting Knowledge Discovery in Databases for Customer Relationship Management in Egyptian Public Banks. *IFIP World Computer Congress, TC 12*, 201-208. doi:10.1007/978-0-387-34749-3_21

- Kiezel, M., & Stefańska, M. (2019). Generations on New Technologies in Retail Banking in the Context of Knowledge Economy. *International Journal of E-Services and Mobile Applications*, 11(3), 19–33. doi:10.4018/IJESMA.2019070102
- Koutanaei, F., Sajedi, H., & Khanbabaei, M. (2015). A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. *Journal of Retailing and Consumer Services*, 27, 11–23. doi:10.1016/j.jretconser.2015.07.003
- Krishna, G., Ravi, V., Reddy, B., Zaheeruddin, M., Jaiswal, H., & Sai Ravi Teja, P. (2019). Sentiment classification of Indian Banks' Customer Complaints. In *Proceedings of IEEE Region 10 Annual International Conference* (pp. 429-434). IEEE. doi:10.1109/TENCON.2019.8929703
- Lahmiri, S. (2017). A two-step system for direct bank telemarketing outcome classification. *Intelligent Systems in Accounting, Finance & Management*, 24(1), 49–55. doi:10.1002/isaf.1403
- Loan Prediction Dataset. (2018). Retrieved from GitHub: <https://github.com/shrikant-temburwar/Loan-Prediction-Dataset/>
- Lv, F., Huang, J., Wang, W., Wei, Y., Sun, Y., & Wang, B. (2019). A two-route CNN model for bank account classification with heterogeneous data. *PLoS One*, 14(8), 1–22. doi:10.1371/journal.pone.0220631 PMID:31425545
- Manthoulis, G., Doumpos, M., Zopounidis, C., & Galarotis, E. (2020). An ordinal classification framework for bank failure prediction: Methodology and empirical evidence for US banks. *European Journal of Operational Research*, 282(2), 786–801. doi:10.1016/j.ejor.2019.09.040
- Marinakos, G., & Daskalaki, S. (2017). Imbalanced customer classification for bank direct marketing. *Journal of Marketing Analytics*, 5(1), 14–30. doi:10.1057/s41270-017-0013-7
- Mohsen, A. M., Hassan, H. A., & Idrees, A. M. (2016). Documents Emotions Classification Model Based on TF IDF Weighting. *International Journal of Computer Electrical Automation Control and Information Engineering*, 10(1).
- Moro, S., Cortez, P., & Rita, P. (2014). A datadriven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62, 22–31. doi:10.1016/j.dss.2014.03.001
- Mostafa, A. M., Helmy, Y. M., Khedr, A. E., & Idrees, A. M. (2020). A Proposed Architectural Framework For Generating Personalized Users' Query Response. *Journal of Southwest Jiaotong University*, 55(5), 3. doi:10.35741/issn.0258-2724.55.5.3
- Ogwueleka, F., Misra, S., Colomo-Palacios, R., & Fernandez, L. (2015). Neural network and classification approach in identifying customer behavior in the banking sector: A case study of an international bank. *Human Factors and Ergonomics in Manufacturing*, 25(1), 28–42.
- Pasha, M., Meherwar, F., Abdul Manan, D., & Furrakh, S. (2017). Performance Comparison of Data Mining Algorithms for the Predictive Accuracy of Credit Card Defaulters. *IJCSNS International Journal of Computer Science and Network Security*, 17(3), 178–183.
- Rehman, A., & Hashim, F. (2020). Impact of Fraud Risk Assessment on Good Corporate Governance: Case of Public Listed Companies in Oman. *Business Systems Research*, 11(1), 16–30. doi:10.2478/bsrj-2020-0002
- Sen, P. C., Hajra, M., & Ghosh, M. (2020). Supervised Classification Algorithms in Machine Learning: A Survey and Review. *Advances in Intelligent Systems and Computing*, 937, 99-111.
- Shahari, F., Zakaria, R., & Rahman, S. (2015). Investigation of the expected loss of sharia credit instruments. *International Journal of Managerial Finance*, 11(4), 503–512. doi:10.1108/IJMF-12-2014-0196
- Smeureanu, I., Ruxanda, G., & Badea, L. (2013). Customer segmentation in private banking sector using machine learning techniques. *Journal of Business Economics and Management*, 14(5), 923–939. doi:10.3846/16111699.2012.749807
- Sultan, N., Khedr, A. E., Idrees, A. M., & Kholeif, S. (2017). Data Mining Approach for Detecting Key Performance Indicators. *Journal of Artificial Intelligence*, 10(2), 59–65. doi:10.3923/jai.2017.59.65

Yurdakul, F. (2014). Macroeconomic Modelling of Credit Risk for Banks. *Procedia: Social and Behavioral Sciences*, 109, 784–793. doi:10.1016/j.sbspro.2013.12.544

Zanin, M., Papo, D., Sousa, P., Menasalvas, E., Nicchi, A., Kubik, E., & Boccaletti, S. (2016). Combining complex networks and data mining: Why and how. *Physics Reports*, 635, 1–44. doi:10.1016/j.physrep.2016.04.005

Zoroja, J., Strugar, I., & Jaković, B. (2020). E-Government in Croatia: Percieved Quality and Obstacles Among Younger Users. *International Journal of E-Services and Mobile Applications*, 12(3), 85–96. doi:10.4018/IJESMA.2020070105