Microfinance Interest Rate Prediction and Automate the Loan Application

I.N. Manawadu[1], G.U. Ganegoda[2], K.M.S. Rathnayake[3], W.A.P. Hansini[4], Hareesha Hilmi[5], W.A. Navodya Pramodi[6]

[1,2,3,4,5,6]*Faculty of Information Technology, University of Moratuwa, Sri Lanka*

[1]imanawadu.uom@gmail.com, [2]upekshag@uom.lk, [3]madhuri.17@itfac.mrt.ac.lk, [4]pavithrah.17@itfac.mrt.ac.lk, [5]hareesha.17@itfac.mrt.ac.lk , [6]pramo.98122@gmail.com

Abstract

In the Sri Lankan context, microfinance is the main banking service provider for the 'unbankable' people who live in rural areas. Due to the informal nature and high-interest rate, microfinance continues to operate in a falling field. In addition, in the appraisal stage, rural borrowers tend to make mistakes while filling out loan applications causing default situations. As a result, microfinance institutions demotivate and people shifting away from traditional forms of borrowing to informal forms is socially problematic. Therefore, in order to address those issues, the researchers collected primary data through ABC bank and analyzed data relating to 10000 random borrowers' records and set of loan forms. The purpose of this research is to automate paper-based loan applications to recognize handwritten characters with the use of information technology through image processing techniques and analyze customers' determinant factors to predict interest rates for microfinance facilities. The finding of our study shows that Regression analysis models can gain the best result in predicting the interest rate. Together Ridge regression analysis and XGBoost regression models gave the most accurate results compared to other models in interest rate prediction. By using this interest rate prediction, microfinance institutions can offer a suitable interest rate that is convenient for the loan borrower. The application automation reduces the paperwork and the manual effort needed to process and increases data accuracy. The developed model can enhance the repayment performance of microfinance firms and prevent defaults by borrowers at the loan appraisal stage.

Keywords: Microfinance, Interest Rate, Handwritten character recognition, Information Technology

I. INTRODUCTION

Microfinance is widely recognized that the exclusion of the poorest lenders, particularly in rural areas from the traditional financial banking system is one of the main obstacles to poverty reduction. Microfinance is also called microcredit. It allows people to take on reasonable small business loans safely and consistently with ethical lending practices. Through the provision of responsive and specific financial services, microfinance institutions (MFIs) allow the financial inclusion of poor entrepreneurs who, for economic reasons, are excluded from the traditional banking system. All around the world, there are several institutions involved in providing microfinance facilities. But in Sri Lanka, this microfinance continues in the descending field because of its risk and the high loan repayment rate. The survival of most MFIs depends entirely on a successful lending program that revolves on funds and loan repayments made to them by the clients. According to that, have to be concerned about the accuracy of the loan application form with accurate data gathering. In this research, we have compared some machine learning models and image processing models to quantify the models' estimation accuracy when measuring individuals' interest rates. Here, the machine learning model type is regression. Moreover, the machine learning target is the interest rate. In the development environment Jupyter Notebook, google colab, and various libraries were used. To find the repayment of a borrower, models such as logistic regression, GradientBoost and CatBoost models were adapted. Furthermore, to predict interest, used the Ridge regression model and Lasso regression model. To identify character recognition, the pre-processing stage consists of applying multiple methods such as grayscale, thresholding, and morphological operations. Consequently, the research study has suggested that to predict the interest rate of the loan per borrower individually. The rest of the paper consists of a literature review, methodology, results, discussion and conclusion.

Faculty of Technology, South Eastern University of Sri Lanka (ISBN 978-624-5736-40-9)

II. LITERATURE REVIEW

Researchers have found various factors, which are affected by MFIs, to be determinants of interest rates. To predict the interest rate, microfinance services, play an important role in identifying low-income groups and providing them with the confidence to start their own businesses while also stabilizing the borrowers' socioeconomic condition.

A. *Interest rate prediction using macroeconomic factors and repayment borrowers*

Microfinance institutions provide lending facilities depending on the borrowers' payback promises. Borrowers have the option of applying for a loan as an individual or as a group (Munene and Guyo, 2013). The repayments of borrowers have a greater impact on the process of granting Microfinance loans. The majority of loan applicants are urban and rural residents who are unable to obtain collateral loans from other banks (Gudde Jote, 2018).

Microfinance institutions become opaque as a result of persistent loan repayment defaults, restricting the availability and accessibility of financial help. Because of that, the government tends to regulate financial acts. The success of a microfinance institution is determined by the efficiency with which loans are collected. Identifying the borrower's background before approving the loan strongly managed the decision-making process to reduce the bank's repayment default cases. Furthermore, research shows that various determinant factors have an impact on loan repayment default. Based on that, the microfinance interest rate is predicted.

In the context of Sri Lanka, there was little Microfinance loan repayment-based research conducted to identify the determinant factors. Current literature reviews have divided factors that affect Microfinance into different categories. (Gudde Jote, 2018) have identified repayment factors as demographic, loan, and institutional categories. (Nanayakkara and Stewart, 2015) classified factors that affect the repayment as lender, borrower, and loan characteristics. Microfinance institutes offer their loans and decide the interest rate in groups and individually. In Sri Lanka, the popular loan type is group lending. Microfinance officers visit village-wise and identify the individuals, group them and grant loans (Nanayakkara and Stewart, 2015). From the side of loan characteristics, the interest rate has been identified as a determinant factor. According to the current situation of microfinance institutes, demand has increased compared to supply. Therefore, the interest charged for the small-size loan has been increased. Indirectly, borrowers' information is asymmetric, and the low repayment frequency has affected the behavior of interest rates (Nanayakkara and Stewart, 2015). (Priyankara and Sumanasiri, 2019) also mentioned as a major factor that affects repayment of the loan. Increased interest rate adversely affects loan repayments. Microfinance Institutes are not the same in interest levels.

In Sri Lanka, the central bank imposes a range where the interest rates can move. No one can violate that range, but microfinance institutes can decide which rate they are going to be followed within that range. Therefore, the interest rate has a significant relationship with loan payments in Sri Lanka. According to the research findings, there are a number of factors that affect loan repayment. Some of the factors mentioned above are some of them.

According to the researchers, it is important to consider the Microfinance Institutions ensuring access to financial services. To generate fair interest rates using technological changes with macroeconomic determinants that are affected positively and negatively. Namely GDP growth rate, unemployment rate, rural population, inflation etc. (Janda and Zetek, 2014). Themistocles Polit off and Dan Ulmer describe how to predict interest rates by comparing the performance of ANN models with that of simple multivariate linear regression (SMLR) models (Priyankara and Sumanasiri, 2019). A variety of factors, both internal and external to MFIs, have been identified as predictors of interest rates charged by MFI researchers. According to Hudon (2007), competition is a key element in influencing interest rates. MFI interest rates are influenced by funding costs, loan size, and efficiency levels, according to Cotler and Almazon (2013), who studied MFIs in a variety of countries. Cost of money, operating costs, provision for bad debts, tax charges, client credit rating, profit, inflation, competition, and customer financial literacy are all factors that influence interest rates, according to Fehmeen (2010). Nirav A. Desai and MBA I.T.Manik Bhatia (Munene and Guyo, 2013) have built a multi-variable regression model to predict the GDP growth rate using key macroeconomic indicators. Inflation, crude oil price, interest rates, services and manufacturing PMI are used as predictors in a multivariable linear regression model to forecast the GDP growth rate (Munene and Guyo, 2013).

Identifying the factors that affect the interest rate of MFIs can be divided into two categories. Such as Microfinance Institutional factors and Macroeconomic factors. Microfinance institutional factors are operating expenses, cost of funds, loan loss expenses, and profit. Operating cost depends on the loan size, age, location and the client's ratings (Janda and Zetek, 2014). The other factor is the cost of funds. When determining a borrower's loan repayment capacity, macroeconomic factors such as pricing levels and economic growth are key considerations (Goncu, 2019). The goal of obtaining a low loan default rate in order to maintain a healthy loan portfolio will lead to MFIs' long-term viability. Interest rates on loans will be unaffected by the size of the microfinance institution and Interest rates on loans will be unaffected by the type of microfinance institution. Another institutional factor is profit. Researchers believe that a higher rate of profit leads to an increase in the interest rate of investors. Loan losses because of borrower default have an effect on MFIs interest rates. Macroeconomic factors which are mainly affected by interest rates are GDP, unemployment rate, inflation rate, agricultural value and rural population (Janda and Zetek, 2014). Basically, MFI interest rates may be caused by increased poverty or an accidentally increase in inflation. In rural areas, the population are highly considerable credit supporters. Therefore, it is also an important factor.

B. Handwritten character recognition of the loan application

In the character recognition process, it detects and recognizes the character from the input image. Then it converts into a machine-editable form like ASCII (Kai Ding *et al.*, 2007). The Handwriting Recognition System is a computerized system that can recognize the characters and the symbols written by hand in natural handwriting. There are two types of handwritten recognition known as Offline Handwriting Recognition which scans the image and is recognized by the computer and Online Handwriting Recognition which recognizes the characters while writing on the touchpad (Ayush and Singh Chauhan, 2016).

1) Hidden Markov Models (HMM):

To recognize cursive handwriting and online handwriting, Hidden Markov Model is used and can identify isolated characters. This model processes two main sections: preliminary classification and recognition using HMMs (Hewavitharana, Fernando and Kodikara, 2002).

2) Kohonen Artificial Neural Network (KANN):

The address identification process of Sri Lankan posts continues manually and most of the time, the address is written in natural handwriting. The aim of automating this separate letter into divisions is to identify the address in the envelope, Kohonen Artificial Neural Network is used. Using 32x32 input neurons and one output neuron were used to build this network. After pre-processing, the data characters are recognized by using the character group. These groups can be customized based on user requirements. Then the system creates a pattern for the received characters. After that, it recognizes the received characters such as the city name by retrieving the city name from the database (Ifhaam and Jayalal, 2019).

3) Thinning Algorithm:

The combination of Thinning Algorithm and character recognition process can identify the Sinhala handwritten characters. This research uses the curvature histogram-based method to identify the characters. In the pre-processing stage, the Sinhala character set is scanned using a Red Green Blue (RGB) color space scanner and converted to Grayscale. Here the characters to identify are in vertical or horizontal projection. After that, resize and find the effective area of the character and convert it to a color image in Grayscale and apply binarization. After that, apply the thinning algorithm to remove the pixels which are not contributing to the character horizontally and vertically. After that, the proposed method identifies the curvature by considering the pixel contribution by the predefined curvature patterns. The pre-processing stage in this method enhances the accuracy of the character recognition (Madushanka, Bandara and Ranathunga, 2017).

Many researchers around the world stated that recognizing a handwritten character achieved a higher level of accuracy as a result of deep learning techniques rather than other character recognition algorithms (Chaudhuri Arindam et al., 2016). Among the classification methods, CNN has the ability to extract the most suitable features from the image and recognize the object among the other patterns (Saqib et al., 2022).

The Work (Ebrahimzadeh and Jampour, no date) proposed system has been generated in three steps. These are, 1) Pre-processing, 2) HOG features extraction and 3) Support vector machine classification. Pre-processing includes some basic image processing to separate numbers from real samples or preparing data from a dataset (reshaped data from images to vectors) and then extracting HOG features, which is a very distinguishable descriptor for digit recognition that divides an input image into 99 cells and computes the histogram of gradient orientations, which represents each digit with a vector of 81 features. Finally, a linear multiclass support vector machine was used to categorize digits in the third stage. The use of HOG features with SVM is the model's key contribution. HOG is a descriptor that can execute distinguishing characteristics quickly and accurately.

In work (Jiang and Zhang, 2020) proposed a system including two novel deep learning models, named EdgeSiamNet and Edge-TripleNet, for handwritten digit recognition. First, use the canny edge extraction method to extract the edge photos according to their model. They propose employing a Siamese/Triple network topology to extract features from both the original input image and the edge image rather than using edge features as residual connections. The strategy increases the network's performance in recognizing distinct numbers without introducing more parameters than EdgeNet. To extract edge pictures, they employ the Canny function provided by OpenCV.A technique for off-line recognition of handwritten Devanagari numbers is proposed in this study (Lakshmi, Jain and Patvardhan, 2007). This research offers a new approach for recognizing handwritten Devnagari numerals utilizing edge histogram features on images that have been pre-processed with splines to improve their quality. Splines have proven effective in a variety of signal-processing applications. PCA is then used to improve the results. For their study, they used 9800 samples of handwritten data. They present a novel strategy for solving the problem of images with thin edges based on detecting edges in continuous domain images via cubic spline interpolation in this study.

"An Improved Algorithm for Segmenting and Recognizing Connected Handwritten Characters" (Zhao, Chi and Feng, 2010) research paper proposed a solution for the segmentation and recognition of handwritten character strings. The proposed method includes a gradient descent mechanism employed in the approach to weight the distance measure while applying KNN for segmenting/recognizing connected characters (numerals and Chinese characters) in the left-to-right scanning direction. A high-quality segmentation technique is required for recognizing related characters. In many circumstances, traditional approaches attempt to segment the string into individual characters without recognition and then apply a recognition algorithm to each isolated character, resulting in incorrect segmentation and poor recognition results. This suggested method mimics the human process of identifying related character strings, in which segmentation and recognition are combined. The technique is resilient and efficient, as evidenced by experimental results on 1959-character strings from the USPS database of postal envelopes.

III. METHODOLOGY

The methodologies used in this work contain two parts. Those are as follows, 1). Loan interest rate prediction 2). Handwritten character recognition of the loan application. The solution is used to predict the interest rate using an approved loan request and look to correctly extract Ref.No, NIC and name of the application and display it and write it into the CSV file. Figure 1 shows the flow of interest rate prediction and loan application recognition in a process diagram.

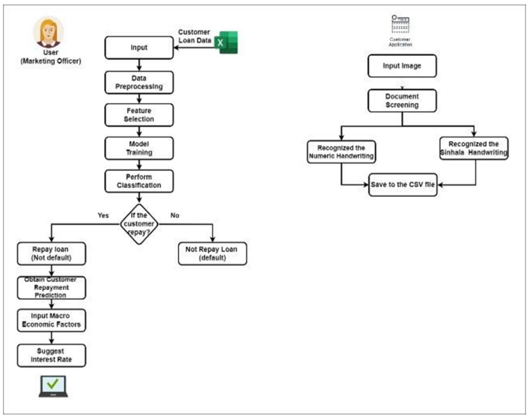

Figure 01: Process Diagram.

A. *Predict the interest rate using macroeconomic factors and repayment borrowers.*

Every year hundreds of loans are granted to different borrowers. Some loans will be repaid and settled correctly; some loans are not. Many factors affect deciding loan repayment or default. The dataset contained 10000 applicants and 10000 applicant-relevant records between the years 2014 to 2020 from island-wide branches in the ABC bank.

The key novelty in this work is to take into account several factors such as loan amount, repayment mode, population density (to identify potential market), income, and expenses that were overlooked in previous studies. In that case, the factors that have a significant impact on loan repayment were identified. Those variables were only used in model development.

1) Data Extraction

The initial dataset used was collected from 10000 loan datasets from ABC bank. Among the dataset, we selected only Microfinance loans. This facility is given to low-income generating people who have minimum access to banking services for borrowing for consumption purposes and also who work as entrepreneurs. Extract only microfinance-relevant loans as Microfinance from 10000 datasets for research purposes. After extraction, the dataset appears to contain 5933 records.

2) Data pre-processing

In the bank, these data were entered by microfinance marketing officers who are working in the field. According to the discussion we had with top management, "they told us these staff members are not well educated and while they enter data, they make a lot of mistakes". In that case, in the pre-processing stage, recognize that there are many custom details that do not have all data attributes available. Such as GENDER, MARITAL STATUS and EFFECTIVE_RATE. These missing data are replaced using missing value imputation with the "mode" of the relevant column. Because removing rows of data-frame could affect the data set. Also, perform basic data cleaning dealing with categorical variables using the dummies approach. Because dummy encoding allows to encode as many category columns as there are in the data frame and choosing how to label the columns, Dataset mainly consists of more than four categorical columns such as DUE_FREQUENCY (Payment mode), MARITAL_STATUS, and GENDER. Also, perform basic data cleaning dealing with data types and converting categorical to numeric using a dummy. Dataset mainly consists of more than five categorical columns such as due frequency, gender, marital status and loan cycle.

3) Data analyzing

This data set has different attributes. This data analysis process tries to identify the correlation between those features. Specially Correlation with the LOAN _STATUS column. Because of having contained more than 50 columns after doing a dummy. During the data analysis process, multicollinearity among independent factors in the data frame was discovered. IBM statistical tools were also used to determine the significance of the factors influencing loan repayment. Also, the multicollinearity among independent variables was identified using a correlation map.

4) Perform Classification

This data set contains so many factors which affect the repayment of loans which are identified based on significance and previous research. But every value is not important to predict the output. The classification process tries to identify feature significance. To predict microfinance repayment, a few machine learning models have been implemented because various studies suggest the best models as different models based on context and input variables. The Boosting algorithms show significant improvement. So, in the beginning, to find the most accurate model for the prediction of repayment, Logistic regression, Gradient boosting, and CatBoosting algorithms are used with true data. The research about loan default prediction conducted in 2019 shows that the XGBoost algorithm has the best performance compared to other algorithms like AdaBoosting, ANN and Support vector. In that case, as a based model, choose XGB and evaluate with other models After comparing the accuracy of models and finding the most suitable model. Then the output of this model sends for use of the interest prediction module. The accuracy of the model was measured using the following equations and the summary of the models shown in Table 01. Among the selected models, the best accuracy scores were recorded in the XGB model. In order to achieve a swap between losing money on default customers, this study examined how various methods affect sensitivity, F1 and precision.

- Accuracy =(TP+TN)/(TP+TN+FP+FN)
- F1-score =2TP/(2TP+FP+FN)

- Precision =TP/(TP+FP)

Table 01: Models accuracy, f1 -score and precision

Model	Accuracy	F1-score	Precision
Logistic Regression	0.70	0.70	0.70
CatBoost	0.88	0.88	0.89
XGBoost	0.91	0.91	0.91
GradientBoost	0.90	0.92	0.92

In this model, the F1 and Precision values appear to be the same. The accuracy column was used to evaluate the model in that case. In comparison to the other three models, the XGBoost model accurately predicts loan repayment with 91 % accuracy.

By identifying the loan approval data and macroeconomic data, the model is developed. To predict the interest rate of the borrowers, various regression models can be used. Among them, for evaluation purposes, the lasso regression model and ridge regression model have been used. Ridge regression is used as the extension of linear regression. To determine the average of the absolute difference between the actual and predicted values in the dataset Mean Absolute Error (MAE) was suitable for that. It measures the average of the residuals of the dataset. When the models are predicting the loan approval data and macroeconomic data, build and evaluate the model by applying Classification and Regression Machine Learning Algorithms. If the loan is approved, Predict the interest rate according to the macroeconomic factors of the country. For the dataset classification, the following steps are being used.

Minimizing the prediction error of all data points is important. The Mean Absolute Error rate is performed. In the Lasso Regression model, the Mean Absolute Error rate is -1.091. In the Ridge Regression model, the Mean Absolute Error is -1.615, and the summary is shown in Table 02. To evaluate the module, one way is to identify the error of all data points, and the Mean Squared Error can be used to compare regression models.

Table 02: Summary of Lasso Regression and Ridge Regression

Methods	Lasso Regression	Ridge Regression
MAE	-1.091	-1.615
MSE	19.79	15.50

When considering the above error data points, a low error rate exists during Ridge Regression.

B. Handwritten character recognition of the loan application

The loan borrower's data is gathered from the loan application form. This form contains both text and numeric data related to the loan borrower. Also, the loan borrower must write in the given format. The form format is in a traditional way that it does not allocate the data-filled boxes to add the text data and only allows for numbers. So before proceeding, need to identify the particular areas separately. Before going to that required to train the model. For the model, use VGG19 Convolutional Neural Network. This recognition process will continue in two phases. The first phase is to train the model and the second one is the recognition part. In the model, training generates the dataset and the labels and train.

VGG19 is combined with 19 layers and contains 16 convolution layers and 3 fully connected layers. The 16 convolutional layers work for feature extraction and 3 fully connected layers 37 specialize in image classification. Those can classify 1000 object categories. When considering the Sinhala language, a total of 736 characters (Chamikara *et al.*, 2014) are available in the language. So that to separately identify their categories can use this. Another thing is that can add a million images to the model. Also, in the Sinhala language because of the writing style, the shape of the character varies. for that also can be addressed in this model (Bansal *et al.*, 2021).

In the pre-processing stage, the scanned image is made in a computer to a readable format. This directly affects the performance of the system.

A convolutional matrix is used here because the shape, thickness and writing style of the Sinhala language are similar to the letters and their technical methods of edge recognition and sharpening help to recognize correctly. Then the image is converted to grayscale because the RGB type of image contains a lot of data and when processing it takes more time, and most of the time, it loses the information. Then do the Thresholding to identify the shape of the image.

In the character recognition phase, to segment, the characters take the sliding window, and it has a fixed value to segment the image. This value is decided by considering the space available for writing. Here, the segmented characters match the classes and their predicted values, which satisfy the conditions assigned to the class value. Finally, after finishing this recognition process, save the recognized field to the CSV file.

Through this part, identify numbers of data from the loan application and increase the accuracy of those data. The application has NIC, DOB, Telephone Number, number of the address, registration number of the business place (if available), duration with year and month etc., with different formats. This module aims to recognize handwritten numbers and increase the accuracy of those data. The after-effects of probably the most broadly utilized tesseract OCR technology. Sensitive information about related banking customers was difficult to get during data collection because the bank's policies only permitted loan applications with no associated data. In order to reduce accuracy difficulties in the data set, the research was carried out using little data that was developed by us and with the help of a bank marketing officer. In that case, use only Name, NIC number, Date of Birth, and application reference number. This module has mainly four stages: pre-processing, segmentation, feature extraction, classification, and recognition (Rosebrock, 2018). The pre-processing step's job is to execute a variety of activities on the supplied image (Ebrahimzadeh and Jampour, no date). It essentially improves the image by making it segmentation friendly. Pre-processing is primarily motivated by the desire to extract a fascinating example from the backdrop. This stage mostly consists of noise filtering, smoothing, and standardization. After the input photos have been pre-processed, the sequence of images is divided into sub-images of individual digits. Individual digits are assigned images and are divided into sub-image of individual digits. Each digit gets resized into pixels on its own. The dataset photos are segmented using an edge detection algorithm in this step. Then, the pre-processing and segmentation stages are completed, and the pre-processed images are represented in the form of a matrix that contains pixels from very large images. In this approach, representing the digits in the photos that carry the necessary information would be beneficial. Feature extraction is the term for this activity. Redundancy in the data is removed during the feature extraction stage.

The flow of both Sinhala and numeric character recognition models is shown in Figure 02.

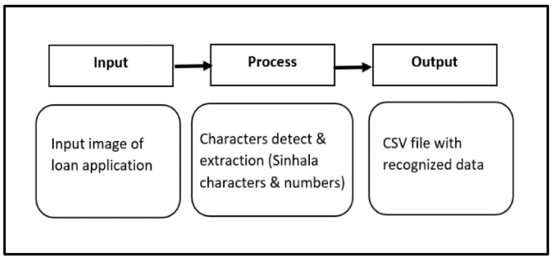

Figure 02: Flow of the character recognition model

IV. RESULTS AND DISCUSSION

This study aims to predict the interest rate according to repayment determinant factors and macroeconomic factors for each customer. Through this work, we were able to achieve the core of the aim. Firstly, the collected dataset from the ABC bank was used for data extraction. Missing data is replaced by using missing value imputation. During data analyzation multi correlation was identified to obtain the best relationship among the variables. And evaluated the regression models using various models to identify the better way to repay and the interest rate. For the loan approval, an accuracy column was used to evaluate the model. In comparison to the other three models, the XGBoost model accurately predicts loan repayment with 91 % accuracy. By analyzing the error rate of the Ridge Regression Lasso Regression models, the least error was identified in the Ridge Regression model.

According to the findings, we are predicting the interest rate for each customer considering several factors that affect the interest rate and ability to loan repayment. The regression models actually suggest a better interest rate in prediction with the help of the Boost algorithm results. However, we are currently obtaining a high-interest rate for each customer via prediction as a result of the sample data acquired. Another part of this work is loan application automation through character recognition. In loan application automation, the research focuses on both Sinhala handwritten characters and number recognition in an equitable manner to obtain more accuracy. In the Sinhala details, the recognition form recognizes the name of the applicant from the filled form. Then, number recognition modules recognized the application number and NIC details. After that, all the details recognized by the system will display in the data entry and save as a CSV file as evidence. For this work, we used a sliding window method for character recognition and Tesseract OCR technologies. According to the findings, identified that using Tesseract OCR is more accurate for character recognition than the sliding window method.

The proposed methods under this study are the system that can be implemented for the predicted interest rate separately for the individual customer and recognize the customer details from the handwritten application form. For this study, we needed a bulk of data for both the interest prediction part and the character recognition part. The ABC bank is given their past data and their loan application form without filling data. The first step of the flow is to identify the loan repayment determinants relevant to the data set in the context of Sri Lanka and then use those determinant factors to perform binary classification to predict whether the loan applicant will be able to repay the loan. Then interpret approved repayment data and macroeconomic factors by applying linear regression models to predict the interest rate for repayment borrowers. The next step is to proceed with the loan application form. In the Sinhala details, the recognition form recognized the name of the applicant from the filled form by using the CNN methodology. With the limited data, accuracy was less compared to other research works. Also, here used the sliding window technique to separate the characters from the given text. That was not a success compared to object detection. Then number recognition modules recognized the application number and NIC details. After that, all the details recognized by the system will display in the data entry and save as a CSV file as evidence. The development of this system uses data mining techniques. Image processing and convolutional neural network, among other techniques.

V. CONCLUSION

This study seeks to find the determinants of credit default in microfinance institutions with the help of Information Technology. The proposed system is mainly concerned with predicting the interest rate for lending money and assuring the data accuracy of the process. For the interest prediction process, evaluate the customer applicability to request a loan and evaluate the interest rate that the system supports for the loan repayment process which does not negatively affect both customers and the bank. When considering the data accuracy, because of the unclear handwriting data entry operator sometimes may fail to update the bank database with accurate data. By identifying the basic details, the bank can give the loan to the customer and continue, but in case something happens related to the Judiciary, the bank has no accurate evidence to prove that the transaction happened via their bank. So in the past few years, microfinance institutions had a lot of bad experiences in proving the data and proving that.

So that this proposed system can provide the processed accurate data which customers write on the loan application form to maintain the accuracy of the process. Loan repayment was predicted using the XGBoost algorithm, and the results were 91 per cent accurate. GridSearchCV and SMOTE techniques were made to improve model performance. XGB produced the best results when compared to the other classifications. The accuracy scores for Gradient Boosting, Logistic Regression, and CatBoost Classifier are all less than 91 per cent. This model will be useful for predicting loan repayment and reducing default cases in the microfinance industry. Predicting the interest rate by identifying the impact of macroeconomic factors is used to do a comparison between regression models to predict the interest rate for non-default borrowers. To obtain a better solution, it was compared regression models by checking the Mean Absolute Error of each model.

It can be chosen the model that has the lowest error. If the model has the lowest mean absolute error, that is the best model for the prediction. In character recognition, here only focus on non-overlapping characters in the field and only one field is recognized. The main input to the program is the scanned image of the loan application form. So, the quality of the image has a higher impact on the process. Because depending on the quality of the image, pre-processing steps may cause some problems, and because of that incorrectly identifying and did not predict the correct results. Another thing is that because of the similarity of the characters, the prediction failed. When it comes to number recognition using tesseract OCR can get high accuracy of character recognition.

REFERENCE

Ayush, P. and Singh Chauhan, S. (2016) 'A Literature Survey on Handwritten Character Recognition', *in. (IJCSIT) International Journal of Computer Science and Information Technologies.*

Bansal, M. et al. (2021) 'Transfer learning for image classification using VGG19: Caltech-101 image data set', *Journal of Ambient Intelligence and Humanized Computing.* doi: 10.1007/s12652-021-03488-z.

Chamikara, M. A. P. et al. (2014) 'Fuzzy Neural Hybrid Method for Sinhala Character Recognition'.

Chaudhuri Arindam et al. (2016) 'Optical character recognition systems for different languages with soft computing'. *New York, NY: Springer Berlin Heidelberg*.

Ebrahimzadeh, R. and Jampour, M. (no date) 'Efficient Handwritten Digit Recognition based on Histogram of Oriented Gradients and SVM'.

Goncu, A. (2019) 'Prediction of exchange rates with machine learning', *in Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing - AIIPCC '19. the International Conference, Sanya, China: ACM Press*, pp. 1–5. doi: 10.1145/3371425.3371448.

Gudde Jote, G. (2018) 'Determinants of Loan Repayment: The Case of Microfinance Institutions in Gedeo Zone, SNNPRS, Ethiopia', *Universal Journal of Accounting and Finance,* 6(3), pp. 108–122. doi: 10.13189/ujaf.2018.060303.

Hewavitharana, S., Fernando, H. C. and Kodikara, N. D. (2002) 'Off-Line Sinhala Handwriting Recognition Using Hidden Markov Models', *in. ICVGIP 2002, Proceedings of the Third Indian Conference on Computer Vision, Graphics & Image Processing, Ahmadabad, India.*

Ifhaam, M. F. A. and Jayalal, S. (2019) 'Sinhala Handwritten Postal Address Recognition for Postal Sorting', *in 2019 International Research Conference on Smart Computing and Systems Engineering (SCSE). 2019 International Research Conference on Smart Computing and Systems Engineering (SCSE), Colombo, Sri Lanka: IEEE*, pp. 134–141. doi: 10.23919/SCSE.2019.8842746.

Janda, K. and Zetek, P. (2014) 'Macroeconomic factors influencing interest rates of microfinance institutions in the Latin America and the Caribbean', *Czech Academy of Agricultural Sciences*, 60, pp. 159–173.

Jiang, W. and Zhang, L. (2020) 'Edge-SiamNet and Edge-TripleNet: New Deep Learning Models for Handwritten Numeral Recognition', *IEICE Transactions on Information and Systems,* E103.D(3), pp. 720–723. doi: 10.1587/transinf.2019EDL8199.

Kai Ding et al. (2007) 'A comparative study of gabor feature and gradient feature for handwritten chinese character recognition', *in 2007 International Conference on Wavelet Analysis and Pattern Recognition. International Conference on Wavelet Analysis and Pattern Recognition, ICWAPR '07, Beijing: IEEE,* pp. 1182–1186. doi: 10.1109/ICWAPR.2007.4421612.

Lakshmi, C. V., Jain, R. and Patvardhan, C. (2007) 'Handwritten Devnagari Numerals Recognition with Higher Accuracy', *in International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007). International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007),* Sivakasi, Tamil Nadu, India: IEEE, pp. 255–259. doi: 10.1109/ICCIMA.2007.443.

Madushanka, P. T. C., Bandara, R. and Ranathunga, L. (2017) 'Sinhala handwritten character recognition by using enhanced thinning and curvature histogram based method', *in 2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP),* Singapore: IEEE, pp. 46–50. doi: 10.1109/SIPROCESS.2017.8124503.

Munene, H. N. and Guyo, S. H. (2013) 'Factors Influencing Loan Repayment Default in Micro-Finance Institutions: The Experience of Imenti North District, Kenya', *International Journal of Applied Science and Technology,* 3 No 3.

Nanayakkara, G. and Stewart, J. (2015) 'Gender and other repayment determinants of microfinancing in Indonesia and Sri Lanka', *International Journal of Social Economics,* 42(4), pp. 322–339. doi: 10.1108/IJSE-10-2013-0216.

Priyankara, D. T. and Sumanasiri, E. A. G. (2019) 'Determinants of Microfinance Loan Default: An Empirical Investigation in Sri Lanka', *South Asian*

Journal of Social Studies and Economics, pp. 1–13. doi: 10.9734/sajsse/2019/v4i330127.

Saqib, N. et al. (2022) 'Convolutional-Neural-Network-Based Handwritten Character Recognition: An Approach with Massive Multisource Data', *Algorithms*, 15(4), p. 129. doi: 10.3390/a15040129.

Zhao, X., Chi, Z. and Feng, D. (2010) 'An improved algorithm for segmenting and recognizing connected handwritten characters', *in 2010 11th International Conference on Control Automation Robotics & Vision. Vision (ICARCV 2010)*, Singapore, Singapore: IEEE, pp. 1611–1615. doi: 10.1109/ICARCV.2010.5707382.