LOAN REPAYMENT PREDICTION

NAME	REG NO
LADE VIVEK	20BCE0432
LAKSHYA RELA	20BCE2831
KAGANA KARTHIKEYA	20BCE0490

ABSTRACT

In the recent times it is becoming very difficult to determine whether a person can repay the loan if given to him by bank or money vendors on time without any delay so we are going to predict whether a person can repay the given loan in time using the historic data of the given person. We are going to first explore the data and extract the important features from the data and then we are going to clean the data and process the data and build our model and then our model can predict whether a person can repay their loan in time taking the historic data of the person as an input. We are using python to build this project and we are using Jupyter notebook as our platform for executing our project. So, by using our model we solve the problem or doubt of whether to give this person the loan required by him this can be used by the bank and money vendors as tool to predict their customers

INTRODUCTION

Loan repayment prediction using Artificial Neural Networks (ANN) is a machine learning technique that involves using historical data to predict whether or not a borrower will repay their loan on time. This project utilizes ANN, a type of deep learning algorithm that is inspired by the structure and function of the human brain, to make these predictions.

The ANN is trained on a dataset that contains information about borrowers, such as their credit score, income, and employment history, as well as information about the loans, such as the loan amount, interest rate, and term. The historical data is used to train the ANN to recognize patterns and relationships between the borrower and loan characteristics and the loan repayment behavior.

INTRODUCTION

Once the ANN is trained, it can be used to predict loan repayment behavior for new borrowers based on their characteristics and loan details. The accuracy of the prediction will depend on the quality and quantity of the historical data used to train the ANN.

The ultimate goal of this project is to use ANN to make accurate and reliable loan repayment predictions, which can be used by banks and other financial institutions to make informed decisions about lending to borrowers. This can help reduce the risk of default and improve overall loan portfolio performance

LITERATURE REVIEW

Paper Name	Author(s)	Year	Methodology	Limitations
"Loan repayment prediction using clustering and SVM"	N. P. Singh, P. Garg	2021	K-means clustering, SVM	Limited data size, lack of explanation for feature selection
"Loan repayment prediction using machine learning and econometric models"	M. M. Rahman, M. A. Hoque, M. A. Islam	2021	Decision tree, random forest, autoregressive integrated moving average (ARIMA)	Limited data availability, lack of comparison with other models
"Loan repayment prediction using hybrid feature selection and machine learning models"	Y. Wang, M. J. Lin	2021	Hybrid model combining decision tree and SVM, feature selection	Limited data size, lack of external validation and less accuracy due to hybrid feature extarction
"Loan repayment prediction using deep feature selection and ensemble learning"	C. Liu, Y. Liu, L. Zheng	2021	Convolutional neural network (CNN), random forest, XGBoost, feature selection	Limited sample size, lack of external validation and high computing power
"Loan repayment prediction using feature selection and decision tree"	Y. Yao, L. Li, X. Liu	2021	Decision tree, feature selection	Limited data size, lack of external validation

LITERATURE REVIEW

Paper Name	Author(s)	Year	Methodology	Limitations
"Loan repayment prediction using a hybrid model of ARIMA and decision tree"	Z. Li, X. Guo	2019	ARIMA, decision tree	Limited data size, lack of external validation and high computing power needed
"Loan repayment prediction using hybrid feature selection and ensemble learning"	Y. Liu, C. Liu, X. Wang	2020	Hybrid model combining decision tree and KNN, feature selection	Limited sample size, lack of external validation and cannot be used in real life due to ensemble limitations

LITERATURE REVIEW

Paper Name	Author(s)	Year	Methodology	Limitations
"Loan default prediction using machine learning algorithms: A comparative study"	S. Bhattacharya, D. Dey, M. Islam	2017	Logistic regression, SVM, decision tree, random forest, ANN	Limited sample size, lack of external validation and no feature extraction done.
"Loan repayment prediction using deep learning and random forest models"	C. C. Lim, A. S. M. Sajeev	2018	Deep belief network, random forest	Limited data size, lack of external validation and very slow compared to other algorithms present
"Loan repayment prediction using time series forecasting models"	S. Kumar, M. S. Rana	2019	ARIMA, Holt-Winters, ANN	Limited data availability, lack of external validation and there is high chance of predicting wrong results with time series.

LITERATURE REVIEW SUMMARY

From the above papers we came to know that many have tried to implement using traditional statistical and machine learning techniques like logistic regression , decision trees , svm , .. etc and most of them are limited to very small dataset and there is no significant feature extraction of the data and some did not have the required accuracy so we are going to do feature extraction and use Artificial Neural Networks to create the model so that we will get higher accuracy and better results than the traditional machine learning algorithms.

LIST OF MODULES

1. Exploratory Data Analysis
2. Data PreProcessing
3. Creating the Model
4. Evaluating the Model

PROPOSED ARCHITECTURE

We are going to take a dataset and first we will explore the data using various visualizations and then we will analyze the data and see for any duplicate data or missing data and try to fill the data and then we will deal with the categorical columns and then we are going to create the model using artificial neural networks and then we are going to predict whether the person can repay their loan by passing the persons historic data to the model and then we will evaluate the model using various metrics like accuracy..etc.