Define Problem / Problem Understanding

Introduction

LOANS are the major requirement of the modern world. By this only, Banks get a major part of the total profit. It is beneficial for students to manage their education and living expenses, and for people to buy any kind of luxury like houses, cars, etc.

But when it comes to deciding whether the applicant's profile is relevant to be granted with loan or not. Banks have to look after many aspects.

Specify the Business Problem

A loan is a sum of money that is borrowed and repaid over a period of time, typically with interest. There are various types of loans available to individuals and businesses, such as personal loans, mortgages, auto loans, student loans, business loans and many more. They are offered by banks, credit unions, and other financial institutions, and the terms of the loan, such as interest rate, repayment period, and fees, vary depending on the lender and the type of loan. A personal loan is a type of unsecured loan that can be used for a variety of expenses such as home repairs, medical expenses, debt consolidation, and more. The loan amount, interest rate, and repayment period vary depending on the lender and the borrower's credit worthiness. To qualify for a personal loan, borrowers typically need to provide proof of income and have a good credit score. Predicting personal loan approval using machine learning analyses a borrower's financial data and credit history to

determine the likelihood of loan approval. This can help financial institutions to make more informed decisions about which loan applications to approve and which to deny.

Business Requirement

In a Simple Term, Company wants to make automate the Loan Eligibility Process in a real time scenario related to customer's detail provided while applying application for home loan forms.

Banks Broadly Check:

• CIBIL Score and Report: It is one of the most important factor that affects your loan approval. A good credit score and report is a positive indicator of your credit health.

• Employment Status: Apart from a good credit history, lenders also check for your steady income and employment status. • Account Details: Suit filed or written off cases are carefully examined by lenders. • Payment History: Lenders check for any default on payments or amount overdue cases, which might project a negative overview of your overall report. • EMI to Income Ratio: Banks also consider the proportion of your existing loans when compared to your salary at the time of loan application. Your chances of loan approval get reduced if your total EMI's exceed your monthly salary by 50%. The credit lending business of the banking industry has seen

rapid progress and extreme competitions from numerous credit startups. At the same time, the increase in credit application and consumption has also led to the increase in losses resulting from bad credits. Credit loan refers to credits provided by banks or financial institutions in general to individuals/consumers which is payable at an agreed date with/without a specified interest. Credit loans are usually provided for many purposes some of which are personal use, educational purposes, medical purposes, travelling and business purposes. Through numerous modern predictive modeling, financial institutions can get insights into applicant's behavior, consumption patterns, default predictors and characteristics. Numerous studies have been conducted in order to identify the important factors that can affect the loan repayment, these studies are important as they help to maximize profit for banks.

The Model Decisions for Loan Prediction:

1. Support Vector Machine for Loan Prediction using Machine Learning

Support Vector Machine (SVM) is a supervised machine learning algorithm that generates a hyperplane (a decision boundary) to separate classes even in a high-dimensional vector space. It can capture different non-linear relationships between the features and the target variable. It decides a class for a sample based on the sign of w[T]+b. In the equation, w (weights) represents the negative and positive hyperplane margin, and b is the bias. SVM is particularly useful in loan prediction because this task usually has several features that need to be considered for the final decision

2. XGBoost for Loan Prediction using Machine Learning

"[Boosting](#)" is a method that combines individual models in an ensemble manner to gain higher performance gain. AdaBoost and Stochastic Gradient Boosting are the most popular boosting algorithms where higher weights are assigned to wrong classified instances during training. At the same time, SGB adds randomness as an integral part of training. Extreme Gradient Boost (XGBoost) is an improvement over Gradient Boost and is very popular in binary [classification algorithms](#). The decision trees are built in parallel in XGBoost than in series, giving it an edge over normal Decision Trees and Boosting algorithms.

3. Random Forest for Loan Prediction using Machine Learning

The random forest algorithm improves the flexibility and decision-making capacity of individual trees. It is another machine learning algorithm incorporating the ensemble learning theorem as its foundation, combining results from various decision trees to optimize training. In some use cases of loan and credit risk prediction, some features are more important than the rest or, more specifically, some features whose removal would improve the overall performance. Since we know the fundamentals of decision trees and how they choose features based on information gain, random forests would incorporate these benefits to give superior performance,

Literature Survey

A recent development of machine learning techniques and data mining has led to an interest of implementing these techniques in various fields [17]. The banking sector is no exclusion and the increasing requirements towards financial institutions to have robust risk management has led to an interest of developing current methods of risk estimation. Potentially, the implementation of machine learning techniques could lead to better quantification of the financial risks that banks are exposed to. Within the credit risk area, there has been a continuous development of the Basel accords, which provides frameworks for supervisory standards and risk management techniques as a guideline for banks to manage and quantify their risks. From Basel II, two approaches are presented for quantifying the minimum capital requirement such as the standardized approach and the internal ratings based approach (IRB) [16]. There are different risk measures banks consider in order to estimate the potential loss they may carry in future. One of these measures is the expected loss (EL) a bank would carry in case of a defaulted customer. One of the components involved in ELestimation is the probability if a certain customer will default or not. Customers in default means that they did not meet their contractual obligations and potentially might not be able to repay their loans [18]. Thus, there is an interest of acquiring a model that can predict defaulted customers. A technique that is widely used for estimating the probability of client default is Logistic Regression [19]. In this thesis, a set of machine learning methods will be investigated and studied in order to test if they can challenge the traditionally applied techniques. A prediction is a statement about what someone thinks will happen in the future. People make predictions all the time. Some are very serious and are based on scientific calculations,

but many are just guesses. Prediction helps us in many things to guess what will happen after some time or after a year or after ten years. Predictive analytics is a branch of advanced analytics that uses many techniques from data mining, statistics, modeling, machine learning, and artificial intelligence to analyze current data to make predictions. "Adyan Nur Alfiyatin, Hilman Taufiq [14] and their friends work on the house price prediction. They use regression analysis and Particle Swarm Optimization (PSO) to predict house price". One other similar work on the Mohamed El Mohadab, Belaid Bouikhalene [15] and Said Safi to predict the rank for scientific research paper using supervised learning. Kumar Arun, Garg Ishan and Kaur Sanmeet [13] work on bank loan prediction on how to bank approve a loan. They proposed a model with the help of SVM and Neural networks like machine learning algorithms. This literature review helps us carry out our work and propose a reliable bank loan prediction model. Manjeet et al (2018) [24] there are seven types of variables that may influence consumer loan default; consumer's annual income, debt-income ratio, occupation, home ownership, work duration and whether or not consumer possesses IJARSCT ISSN (Online) 2581-9429 International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) Volume 5, Issue 1, May 2021 Copyright to IJARSCT DOI: 10.48175/IJARSCT-1168 451 www.ijarsct.co.in Impact Factor: 4.819 a saving/checking account. In a work by Steenackers [26] and Goovaerts, the key factors that may influence loan default are borrower's age, location, resident/work duration, owner of phone, monthly income, loan duration, whether or not applicant works in a public sector, house ownership and loan numbers. Another study by Ali Bangher pour [27] on a large dataset within the period of 2001-2006 indicated that loan age was the most important factor when predicting loan default while market loan-to-value was the most effective factor for mortgage loan applications. In addition to identifying factors that may influence loaned fault, there is also a need to build robust and effective machine learning models

that can help capture important patterns in credit data. The choice of model so great importance as the chosen model plays a crucial role in determining accuracy, precision and efficiency of a prediction system. Numerous models have been used for loan default prediction and although there is no one optimal model, some models definitely do better than others. In 2019, Vimala and Sharmili [1] proposed a loan prediction model using and Support Vector Machines(SVM)methods. Naïve Bayes, an independent speculation approach, encompasses probability theory regarding the data classification. On the other hand, SVM uses statistical learning model for classification of predictions. Dataset from UCI repository with 21 attributes was adopted to evaluate the proposed method. Experimentations concluded that, rather than individual performances of classifiers (NB and SVM), the integration of NB and SVM resulted in an efficient classification of loan prediction. In 2019, Jency, Sumathi and Shiva Sri [2] proposed an Exploratory Data Analysis(EDA) regarding the loan prediction procedure based on the client's nature and their requirements. The major factors concentrated during the data analysis were annual income versus loan purpose, customer 's trust, loan tenure versus delinquent months, loan tenure versus credit category, loan tenure versus number of years in the current job, and chances for loan repayment versus the house ownership. Finally, the outcome of the present work was to infer the constraints on the customer who are applying for the loan followed by the prediction regarding the repayment. Further, results showed that, the customers were interested more on availing short-tenure loans rather than long-tenure loans. In 2019, Supriya, Pavani, Saisushma, Vimala Kumari and Vikas [3] presented a ML based loan prediction model. Themodulesin the present approach were data collection and pre-processing, applying the ML models, training followed by testing the data. During the pre-processing stage, the detection and removal of outliers and imputation removal processing were carried out. In the present

method, SVM, DT, KNN and gradient boosting models were employed to predict the possibilities of current status regarding the loan approval process. The conventional 80:20 rule was adopted to split the dataset into training and testing processes. Experimentation concluded that, DT has significantly higher loan prediction accuracy than the other models. In 2017, Goyal and Kaur [4] presented a loan prediction model using several Machine Learning (ML) algorithms. The dataset with features, namely, gender, marital status, education, number of dependents, employment status, income, co applicant's income, loan amount, loan tenure, credit history, existing loan status, and property area, are used for determining the loan eligibility regarding the loan sanctioning process. Various ML models adopted in the present method includes, Linear model, Decision Tree (DT), Neural Network (NN), Random Forest (RF), SVM, Extreme learning machines, Model tree, Multivariate Adaptive Regression Splines, Bagged Cart Model, NB and TGA. When evaluated these models using Environment in five runs, TGA resulted in better loan forecasting performance than the other methods. In 2016, Aboobyda Jafar Hamid and Tarig Mohammed Ahmed [5] presented a loan risk prediction model based on the data mining techniques, such as Decision Tree (J48), Naïve Bayes (NB) and BayseNet approaches. The procedure followed was training set preparation, building the model, Applying the model and finally. Evaluating the accuracy. This approach was implemented using Weka Tool and considered a dataset with eight attributes, namely, gender, job, age, credit amount, credit history, purpose, housing, and class. Evaluating these models on the dataset, experimental results concluded that, J48 based loan prediction approach resulted in better accuracy than the other methods. In 2016, Kacheria, Shivakumar, Sawkar and Gupta [6] suggested a loan sanctioning prediction procedure based on NB approach integrated with K-Nearest Neighbor (KNN) and binning algorithms. The seven parameters considered were income, age, profession, existing loan with its tenure, amount and approval status.

The sub-processes include, Preprocessing (handling the missing values with KNN and data refinement using binning algorithm), Classification using IJARSCT ISSN (Online) 2581-9429 International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) Volume 5, Issue 1, May 2021 Copyright to IJARSCT DOI: 10.48175/IJARSCT-1168 452 www.ijarsct.co.in Impact Factor: 4.819 NB approach and Updating the dataset frequently results in appropriate improvement in the loan prediction process. Experimentation put-forth the conclusion that, integration of KNN and binning algorithm with NB resulted in improved prediction of loan sanctioning process. In 2016, Goyal and Kaur [7] suggested an ensemble technique based loan prediction procedure for the customers. The sub processes in the present method includes, data collection, filtering the data, feature extraction, applying the model, and finally analysis the results. The various loan prediction procedures implemented in the present method were Random Forest (RF), SVM and Tree model with Genetic Algorithm (TGA). The parameters considered for evaluating the models were accuracy, Gini Coefficient, Area Under Curve (AUC), Receiver Operating Curve (ROC), Kolmogorov - Smirnov (KS) Chart, Minimum Cost - Weighted Error Rate, Minimum Error Rate, and K-Fold Cross Validation parameters. Experimentation outcome concluded that the integration of three methods (RF, SVM and TGA) resulted in improved loan - prediction results rather than individual method 's prediction. In 2006, Sudhamathy [8] suggested a risk analysis method in sanctioning a loan for the customers using R package. The various modules include data selection, pre-processing, feature extraction and selection, building the model, prediction followed by the evaluation. The dataset used for evaluation in this method was adopted from UCI repository. To fine tune the prediction accuracy, the pre-processing operation includes the following sub-processes: detection, ranking and removal of outliers, removal of imputation, and balancing of dataset by proportional bifurcation regarding

testing and training process. Further, feature selection process improves the prediction accuracy. When evaluated, the DT model resulted in 94.3% prediction accuracy. The process of analyzing data from different perspectives and extracting useful knowledge from it. Tithe core of knowledge discovery process. The various steps involved in extracting knowledge from raw data. Different data mining techniques include classification, clustering, association rule mining, prediction and sequential patterns, neural networks, regression etc. Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Fraud detection and credit risk applications are particularly well suited to classification technique. This approach frequently employs Decision tree based classification Algorithm. In classification, a training set is used to build the model as the classifier which can classify the data items into its appropriate classes. A test set is used to validate the model.

Social & Business Impact

Social Impact:

A social impact loan provides growth capital to ventures that are delivering financially and operationally sustainable solutions to social or environmental challenges. But why would a corporation choose to *loan* philanthropic dollars to an impact organization instead of giving it in the form of a grant?

More and more corporations are realizing the increasing value that social responsibility actions and corporate giving practices are generating for their businesses. Today, aligning your business with a

mission yields better financial performance. Despite this, the amount of corporate grant and sponsorship dollars (the traditional forms of corporate giving) remains limited, largely because of the high costs and direct expenses associated with these methods.

The good news is that there are new lower-risk, lower-cost alternatives to grants and sponsorship. More and more corporations are diversifying their charitable giving practices in order to mobilize their financial resources that are not earmarked for granting. In addition to providing grants to nonprofits that are doing good, many corporations are now also lending money to for-profit social ventures and revenue-generating nonprofits that are delivering impact at scale.

Business Impact

Business loans are a form of commercial financing that gives companies access to borrowed funds that they can use for operational expenses, expansion, and more. Generally, business loans work one of two ways. First, the lender might provide a lump sum. Second, they may operate as a line of credit. Additionally, either option may be secured or unsecured.

Business loans do come with interest rates, though they're potentially lower than what people find with personal loans and similar alternatives. For example, the Small Business Administration (SBA) loan program offers rates between [base rate plus 2.25% to base rate plus 4.75%](). However, rates do vary by lender, so it's possible to see interest rates far above that mark.

In exchange for receiving the money, your business agrees to repay the principal along with interest and fees. The payment schedule can vary depending on the loan agreement. Some require monthly payments, while others have daily, weekly, or biweekly payment requirements.