

# **Smart Lender-Applicant Credibility Prediction for Loan Approval**

## **1. INTRODUCTION :**

The primary source of revenue for the banking industry and source of financial risk for banks is a loan. The interest collected on loans disbursed directly accounts for significant amounts of a bank's assets. The ability of the borrower to repay the loan within the allotted period is one of the major hazards associated with the lending of loans. It's known as "credit risk." A candidate's credit score served as the basis for determining whether or not to approve a loan on them. Thus, the purpose of this study is to describe the use of various machine learning approaches that effectively identify who to lend money to and assist banks in identifying loan defaulters for significantly lower credit risk.

The fact that our banking system offers a wide range of goods, a bank's credit line is its primary source of income. As a result, they are able to profit from interest on the loans they credit. Lenders always seek to lower their credit risk since commercial loans have historically represented a significant portion of the banking sector. The function that banks play in the

modern market economy is substantial. Loans, or whether clients repay or don't return them, significantly impact a bank's profitability. Before granting loans to borrowers, banks must determine if they are good (non-defaulters) or bad (defaulters).

The creditworthiness of the borrowers is one of the most critical issues in commercial loan financing.

The probability that borrowers may default on their loan commitments is referred to as credit risk. For any bank or institution, determining whether a borrower will be good or bad is a very difficult process. The banking system employs a manual procedure to determine whether or not a borrower has defaulted. The manual method will undoubtedly be more precise and efficient, but it will not be able to handle a high volume of loan applications at once. When a situation like this arises, it will take a very long time to make decisions and a lot of labour will be needed.

## **2. OBJECTIVE :**

Loans account for a large portion of bank profits. For financial companies, the loan approval process is crucial. Because loan defaults are occurring more frequently and it is becoming more challenging for banking authorities to properly assess loan requests and address the dangers of people defaulting on loans, it is very difficult to forecast if clients will be able to pay back the loan. Numerous scholars have been focusing on loan approval system prediction in recent years. For vast amounts of data, the machine learning technique is highly helpful in predicting outcomes. Four algorithms, including Random Forest, Decision Tree, Naive Bayes, and Logistic Regression, are employed in this study to forecast whether or not clients would be approved for loans. The same information will be analysed for all four methods, and the most accurate algorithm will be chosen to deploy the model. From this point forward, we create a machine learning-based bank loan prediction system that chooses the qualified applicants for loan approval on its own.

### 3. LITERATURE SURVEY:

#### **Paper 1:An Approach For Prediction Of Loan Approval**

**Publication year:**May-June 2021

**Author name :**Ms. Kathe Rutika Pramod

**Journal name:** International Journal of Creative Research Thoughts (IJCRT)

**Summary:** In our banking system banks have many products to sell but main source of income of any banks is on its credit line they can earn from interest of those loans which they credits a bank's profit or a loss depends to a large extent on loans the customers are paying back the loan defaulting by predicting the loan defaulters the bank can reduce its non performing assets. the maximization of profits, it is essential to study the nature of the different methods and their comparison.

**Methodology used:** The prediction model which is constructed using three different training algorithms to train a supervised two layer feedforward network. The results show that the training algorithm improves the design of loan default prediction model. and we use of machine learning.

#### **Paper 2:Loan Prediction System using Machine Learning**

**Publication year:** March-April 2021

**Author name:** Ramrao adik,yash patil

**Journal name:** International Conference on Advances in Computing and Communication (ICACC)

**Summary:** As the needs of people are increasing, the demand for loans in banks is also frequently getting higher every day. Banks typically process an applicant's loan after screening and verifying the applicant's eligibility, which is a difficult and time-consuming process. In some cases, some applicants default and bank lose capital. The Machine Learning Approach is ideal for reducing human effort and effective decision making in the loan approval process by implementing machine.

**Methodology used:** The algorithm which will be used for data modelling is Logistic Regression using stratified k-folds cross-validation and Random Forest.

#### **Paper 3:Loan Credibility Prediction System Based on Decision Tree Algorithm**

**Publication year:**09 September 2015

**Author name:** Sivasree M S, Rekha Sunny T

**Journal name:** International Journal of Engineering Research & Technology (IJERT)

**Summary:** Data mining techniques are becoming very popular now-a-days because of the wide availability of huge quantity of data and the need for transforming such data into knowledge. Techniques of data mining are implemented in various domains such as retail industry, telecommunication industry, biological data analysis, intrusion detection and other scientific applications. Data mining techniques can also be used in the banking industry which help them compete in the market well equipped. In this paper we introduce an effective prediction model for the bankers that help them predict the credible customers who have applied for loan. Data Mining Algorithm is applied to predict the attributes relevant for credibility.

**Methodology used:** Decision Tree, Credit Risk Assessment, Classification, Prediction, Attribute Selection.

#### **Paper 4: Loan Approval Predictions**

**Publication year:** 04 | Apr 2022

**Author name:** M. A. Sheikh, A. K. Goel and T. Kumar

**Journal name:** International Research Journal of Engineering and Technology (IRJET)

**Summary:** Today a lot of people/companies are applying for bank loans. The core business part of every bank is the distribution of loans. The main objective of the banking sector is to give their assets in safe hands. But the banks or the financial companies take a very long time for the verification and validation process and even after going through such a regress process there is no surety that whether the applicant chosen is deserving or not. To solve this problem, we have developed a system in which we can predict whether the applicant chosen will be a deserving applicant for approving the loan or not. The system predicts on the basis of the model that has been trained using machine learning algorithms.

#### **Paper 5: Loan default prediction of Chinese P2P market: a machine learning methodology**

**Publication year:** 21 September 2021

**Author name :** Junhui Xu, Zekai Lu & Ying Xie

**Journal name:** scientific reports

**Summary:** Repayment failures of borrowers have greatly affected the sustainable development of the peer-to-peer (P2P) lending industry. The latest literature reveals that existing risk evaluation systems may ignore important signals and risk factors affecting P2P repayment. In our study, we applied four machine learning methods (random forest (RF), extreme gradient boosting tree (XGB), gradient boosting model (GBM), and neural network (NN)) to predict important factors affecting repayment by utilizing data from Renrendai.com in China from

Thursday, January 1, 2015, to Tuesday, June 30, 2015

**Paper 6:Rethinking SME default prediction: a systematic literature review and future perspectives**

**Publication year:** 29 January 2021

**Author name:** Ciampi, F., Giannozzi, A., Marzi, G.

**Journal name :** Scientometrics

**Summary:** Over the last dozen years, the topic of small and medium enterprise (SME) default prediction has developed into a relevant research domain that has grown for important reasons exponentially across multiple disciplines, including finance, management, accounting, and statistics.

**Paper 7:Analysis of Loan Availability using Machine Learning Techniques**

**Publication year:** September 2021

**Author name :**Sharayu Dosalwar, Ketki Kinkar, Rahul Sannat, Dr Nitin Pise

**Journal name:** International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

**Summary:** In the banking system, banks have a variety of products to provide, but credit lines are their primary source of revenue. As a result, they will profit from the interest earned on the loans they make.Loans, or whether customers repay or default on their loans, affect a bank's profit or loss. The bank's Non-Performing Assets will be reduced by forecasting loan defaulters. As a result, further investigation into this occurrence is essential. Because precise forecasts are essential for benefit maximisation, it's crucial to analyse and compare the various methodologies.

**Paper 8:Analysis and Comparison of Loan Sanction Prediction Model Using Python**

**Publication year:**Jun 2018

**Author name:** SRISHTI SRIVASTAVA, AYUSH GARG, ARPIT SEHGAL & ASHOK KUMAR

**Journal name:**International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR)

**Summary:**Sanctioning of loan to borrowers form the most vital part of every bank's business, as most of its assets come from the profit gained in the loan distribution process. Therefore, it is essential for banks to estimate whether the customer is right or not i.e., his ability to default or not in the coming future. By doing so, the bank would know that its assets are in safe hands

### **Paper 9: Applications of Machine Learning In Loan Prediction Systems**

**Publication year:**22 May, 2021

**Author name:**Dr. Kavita Khadse

**Journal name:**Chetana's Ramprasad Khandelwal Institute Of Management & Research, Affiliated To Mumbai University, Mumbai, India

**Summary:**In India, The Quantity Of Individuals Applying For The Advances Gets Expanded For Different Reasons Lately. The Bank Representatives Can't Dissect Or Foresee If The Client Can Make Restitution Of The Sum (Great Client Or Terrible Client) For The Given Loan Fee. The Point Of This Paper Is To Discover The Idea Of The Customer Applying For The Individual Advance.

### **Paper 10: Crop-yield and Price Forecasting using Machine Learning**

**Publication year:** 2020

**Author name:** Sadiq A Mulla,Dr.S.A.Quadri.

**Journal name:** The International journal of analytical and experimental modal analysis.(IJAEMA)

**Summary:** Guaranteeing food profitability is a significant issue for the creating nations like India, where more than 33% of the individuals is live in neediness. Season crop yield estimations are inconceivably recognized as a significant contribution for exploring food accounting reports and yield deficiencies. Harvest yield estimation and evaluation is done worldwide on a territorial premise to empower high yield and cost estimation. To estimate cost there is no system in place to advice farmers what crops to grow. Hence, In this paper we attempt to predict crop price that a farmer can obtain from his land, by analysing patterns in past data. We have considered few rabi and kharif season crops like paddy, arhar, bajra, barley etc for our analysis. We make use of several data such as rainfall, temperature, market prices, area of land and past yield of a crop. In this project, we implement a supervised machine learning algorithm namely, Decision tree algorithm and analyse the data and predict for the new set of data. We also predict the price and the gain for next twelve months over the past twelve months and give the time series analysis of the same

### **Paper11: AN INVESTIGATION ON CROP YIELD PREDICTION USING MACHINE LEARNING**

**Publication year:**2021

**Author name:** Guna Sekhar Sajja.

**Journal name:** International Conference on Inventive Research in Computing Applications.

**Summary:**The existence of humans, agriculture is vitally crucial. For a big population of the globe, agriculture provides a living. It also provides the locals with a large number of work openings. Many farmers desire to use old-fashioned farming techniques, which provide poor income. Critical to the economy's long-term development and advancement are agriculture and the related industries. Decision making, crop selection and supporting systems for increased crop output are the primary problems for agricultural production. The prediction of agriculture depends on parameters such as temperature, soil fertility, amount of water, water quality and seasons, crop price, etc. Machine learning plays an important role in crop yield prediction on the basis of geography, climate details, and season. It helps farmers in growing most appropriate crop for their farm land. This paper presents a machine learning based framework for prediction of crop yield. For experimental set up, a data set is created for crop details.

Machine learning algorithms SVM, random forest and ID3 are used for investigation.

### **Paper12: Toward Automated Machine Learning-Based Hyperspectral Image Analysis in Crop Yield and Biomass Estimation**

**Publication year:**2022

**Author name:** Kai-Yun Li , Raul Sampaio de Lima , Niall G. Burnside , Ele Vahtmäe .

**Journal name:**Multi Disciplinary Digital Publishing Institute.(MDDPI)

**Summary:** The incorporation of autonomous computation and artificial intelligence (AI) technologies into smart agriculture concepts is becoming an expected scientific procedure. The airborne hyperspectral system with its vast area coverage, high spectral resolution, and varied narrow-band selection is an excellent tool for crop physiological characteristics and yield prediction. However, the extensive and redundant three-dimensional 3D cube data processing and computation have made the popularization of this tool a challenging task. This research integrated two important open-sourced systems (R and Python) combined with automated hyperspectral narrowband vegetation index calculation and the state-of-the-art AI-based automated machine learning (AutoML) technology to estimate yield and biomass, based on three categories (spring wheat, pea and oat mixture, and spring barley with red clover) with multifunctional cultivation practices in northern Europe and Estonia.

## **4.CONCLUSION:**

We have selected the machine learning technique to analyse the bank dataset in order to forecast the loan approval status of the applied client. We tested a number of machine learning algorithms to see which one would work the best on the dataset to get the most

accurate results. Using this strategy, we discovered that, aside from logistic regression, the other algorithms delivered accurate results in a suitable manner.

The remaining algorithms' accuracy ranged from 75% to 85%. The logistic regression, however, provided us with the highest level of accuracy (88.70%) after comparison of all the methods.

We also identified the key elements that have the greatest impact on the loan approval status. The performance accuracy of these most crucial characteristics applied to a few chosen algorithms is then compared to the case where all features were employed. The banks may use this model to determine what elements are crucial for the loan approval process. Based on their accuracy, the comparison research clearly identifies which algorithm will be the best and disregards the others.