**LOAN APPLICATION ANALYSIS**

**2nd REVIEW**

**Team Members:**

**2010030362\_Rushitha**

**2010030550\_Kavya**

**Guided By:**

**Dr.K.Sreenivasa Rao**

**Table of Contents:**

* Introduction
* Project area
* Literature Survey
* Problem Statement
* Python Integration
* Conclusion

**Introduction**

Distribution of the loans is the core business part of almost every bank. The main portion of the bank’s assets directly came from the profit earned from the loans distributed by the banks. The main objective of the Project is to compare the Loan Prediction Models made implemented using various algorithms and choose the best one out of them that can shorten the loan approval time and decrease the risk associated with it. It is done by predicting if the loan can be given to that person since various parameters like credit score, income, age, marital status, gender, etc. The prediction model not only helps the applicant but also helps the bank by minimizing the risk and reducing the number of defaulters.

**Project area**

Data visualization provides information to communicate with complex data relationships and data-driven insights in a way that is easy to understand. Kiva loans dataset mainly focuses on visualizations to quickly summarize loan data by graphs. The aim of this project is to provide an easy way to choose deserving applicants. It checks the applicant’s data and provides whether his/her loan can be approved or not as a result.

**Literature Survey**

**Literature Survey-01**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S.No:** | **Authors** | **Title** | **Publishing** | **Techniques & Dataset** | **Pros** | **Cons** |
| 1. | Sheikh Mohammad Ahmad, Amit Kumar Goel, and Tapas Kumar. | An Approach for Prediction of Loan Approval using Machine Learning Algorithm | International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 490-494. IEEE, 2020. | Logistic regression,  Bank customers dataset | Logistic regression is used for the description of data and to explain the relationship between a single binary variable and single or multiple nominal, ordinal, and ration level variables which are independent in nature. | Binning cannot be used for categorical columns, coz the labels that have low frequencies might be affected from the Robustness of statistical models in a negative manner. |

**Literature Survey-02**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S.No:** | **Authors** | **Title** | **Publishing** | **Techniques & Dataset** | **Pros** | **Cons** |
| 2. | P. Maheshwari, and CH V. Narayana. | Predictions of Loan Defaulter - A Data Science  Perspective | International Conference on Computing, Communication and Security (ICCCS), pp. 1-4. IEEE, 2020. | Principal Component Analysis (PCA).  Lending club loan dataset | Principal Component Analysis (PCA).  Lending club loan dataset | K-Nearest Neighbor (KNN): The KNN algorithm takes more processing time for lagers datasets. |

**Literature Survey-03**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S.No:** | **Authors** | **Title** | **Publishing** | **Techniques & Dataset** | **Pros** | **Cons** |
| 3. | Patel, Bhoomi, Harshal Patil, Jovita Hembram, and Shree Jaswal. | Loan Default Forecasting using Data Mining | In 2020 International Conference for Emerging Technology (INCET), pp. 1-4. IEEE, 2020. | Gradient Boosting  Loan dataset | We can infer that Gradient Boosting and CatBoost Classifier is doing prediction well for the loan dataset. | Gradient boosting is a process consisting of multiple models.  It is unusual to discard a variable as the interpretation of the variables is not straight. On the other hand, it is an accepted practice to eliminate variables while fitting logistic regression, even if it minimizes the overall model accuracy and prediction power. |

**Literature Survey-04**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S.No:** | **Authors** | **Title** | **Publishing** | **Techniques & Dataset** | **Pros** | **Cons** |
| 4. | Chang, Yung-Chia, Kuei-Hu Chang, and Yi-Xin Lin. | Establishment of Business Loan Default Prediction Model by Integrating Survival Analysis with Logistic Regression | Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu 300, Taiwan b, 2020. | Logistic regression  Banking Dataset - Marketing Targets | Survival analysis is mainly for implementing in-depth discussions on the correlation between the survival time of a sample group and each variable. | The most significant difference between general linear regression and logistic regression lies in the “processible data attributes”; logistic regression could be applied for processing binary data and predicting its odds ratio for the occurrence of an event, regardless of whether the predictor variable is a categorical variable or continuous variable. |

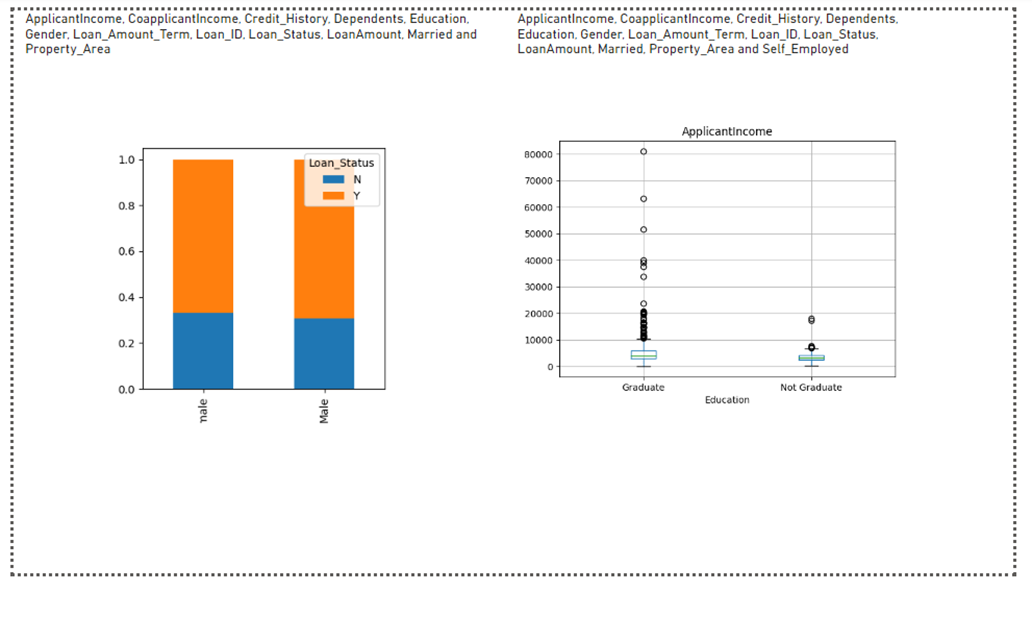
**Literature Survey-05**

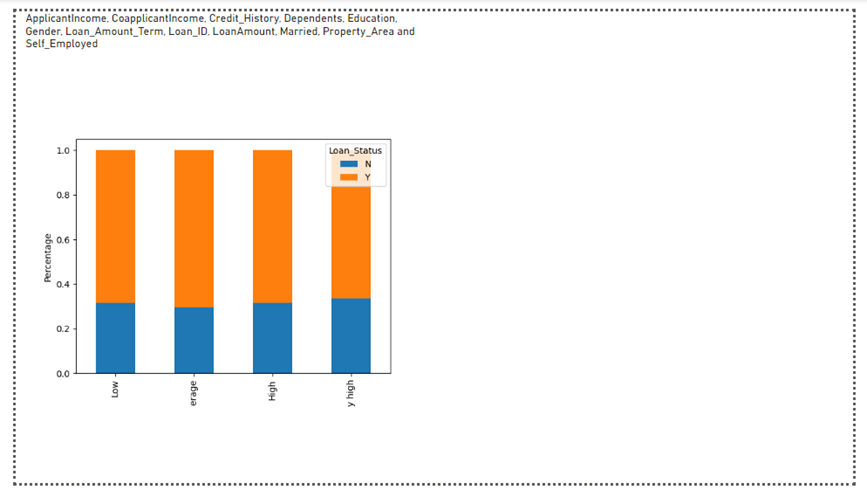
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S.No:** | **Authors** | **Title** | **Publishing** | **Techniques & Dataset** | **Pros** | **Cons** |
| 5. | Alaradi, Mohamed, and Sawsan Hilal. | Tree-Based Methods for Loan Approval | International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), pp. 1-6. IEEE, 2020. | Synthetic Minority Oversampling Technique.  Loan prediction dataset | Random forests technique has the advantage of de-correlating  the constructed decision trees and hence reducing the variance when averaging the trees. | It performed less efferently in classifying the approved category with higher rates for false negatives which translates into the problem statement of this work as lost opportunity to grant a loan. |

**Problem Statement**

Nowadays, several people apply for loans for various reasons. However, there are several cases where people do not repay the amount which causes huge financial loss. Today many banks / financial companies approve loans after a regress process of verification and validation but still, there is no surety whether the chosen applicant is the right applicant out of all applicants. We try to build a system that can classify the loaners among all, and it would prevent huge financial loss. The whole process of feature validation is automated by the machine learning technique.

**Python Integration using power bi**





**Conclusion**

The main objective of the Project is to build a Loan Prediction Model that can shorten the loan approval time and decrease the risk associated with it. It is done by predicting if the loan can be given to that person since various parameters like credit score, income, age, marital status, gender, etc. The prediction model not only helps the applicant but also helps the bank by minimizing the risk. The loan application analysis system can assist the banks in making the best judgement on whether to approve or deny a customer's loan request.