**LOAN APPLICATION ANALYSIS (TEAM NO : 13)**

**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. |