INTRODUCTION:

We present this documentation as a detailed account of our Machine Learning project focused on revolutionizing the loan approval process. This document encapsulates our exploration of a dataset containing 13 attributes and 4269 rows, which served as the basis for our research.

Our approach commenced with rigorous data preprocessing to ensure dataset integrity. We explored various Machine Learning algorithms, including KNN Classifier, Decision Tree Classifier, SVM, Naive Bayes Classifier, LGBM, and Convolutional Neural Networks (CNN). After meticulous testing, we combined CNN with LGBM to create a hybrid model, excelling in accuracy and efficiency.

This documentation outlines our technical methodologies and the integration of our hybrid model into an intuitive user interface developed with React and integrated with Flask. The interface allows users to input parameters, enabling real-time loan approval predictions. Visual representations, such as the ROC curve and a consolidated histogram, enhance understanding of model performance, aiding stakeholders in decision-making.

We invite you to explore the details of our project, gaining insights into our research methodologies, challenges, and outcomes. We appreciate your interest and trust that this documentation will provide a comprehensive understanding of our commitment to excellence in Machine Learning.

DATASET DESCRIPTION:

The loan approval dataset is a collection of financial records and associated information used to determine the eligibility of individuals or organizations for obtaining loans from a lending institution. It includes various factors such as cibil score, income, employment status, loan term, loan amount, assets value, and loan status. This dataset is commonly used in machine learning and data analysis to develop models and algorithms that predict the likelihood of loan approval based on the given features.

USED METHODOLOGY:

**1. Data Collection:** The study utilized a loan approval dataset comprising 13 attributes and 4269 rows. This dataset served as the foundation for our research, providing essential information for analysis and modelling.

**2. Data Preprocessing:** The dataset underwent rigorous preprocessing to ensure data integrity. This involved handling missing values, outlier detection, and normalization. Categorical variables were encoded, and feature scaling techniques were applied to standardize the data for uniformity.

**3. Algorithm Selection:** We explored various Machine Learning algorithms, including KNN Classifier, Decision Tree Classifier, Support Vector Machine (SVM), Naive Bayes Classifier, LGBM, and Convolutional Neural Networks (CNN). Each algorithm was implemented and evaluated separately to gauge its performance in predicting loan approvals.

**4. Hybrid Model Development:** A hybrid approach was adopted, combining CNN with LGBM. CNN was employed to extract intricate patterns from the data, and its output was utilized as features for LGBM, a gradient boosting framework. This hybrid model aimed to leverage the strengths of both techniques, enhancing prediction accuracy and efficiency.

**5. Model Evaluation:** To assess the performance of the models, standard metrics such as accuracy, precision, recall, and F1 score were calculated. Additionally, visualizations, including the Receiver Operating Characteristic (ROC) curve and a consolidated histogram, were generated to provide a comprehensive overview of the models' performance across different evaluation criteria.

**6. User Interface Development:** The hybrid model was integrated into an intuitive user interface developed using React for the frontend and Flask for the backend. This interface allowed users to input relevant parameters for loan approval prediction. The integration ensured real-time processing and seamless user experience.

**7. Visualization and Interpretation:** Visual aids, such as the ROC curve and histogram, were instrumental in interpreting complex model outputs. These graphical representations facilitated a better understanding of the models' strengths and weaknesses, aiding stakeholders in decision-making processes.

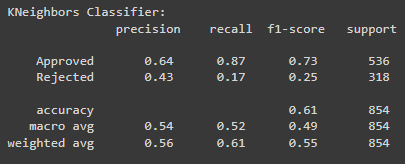
**8. Cross-Validation:** To mitigate overfitting and ensure the generalizability of the model, k-fold cross-validation was employed. The dataset was divided into subsets, and the hybrid model was trained and evaluated iteratively, providing a more robust assessment of its performance.

**9. Performance Comparison:** A comparative analysis was conducted to evaluate the hybrid model against individual Machine Learning algorithms. Performance metrics were systematically compared, and statistical tests were applied to identify significant differences, providing valuable insights into the superiority of the hybrid approach.

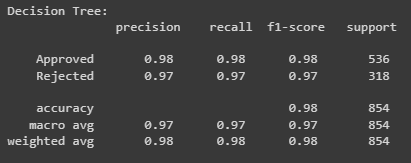
This methodological framework guided our research, enabling us to develop and evaluate a powerful hybrid model for loan approval predictions. The systematic approach ensured the reliability and effectiveness of our findings, contributing to the advancement of predictive modeling in the domain of loan approval processes.

RESULTS:

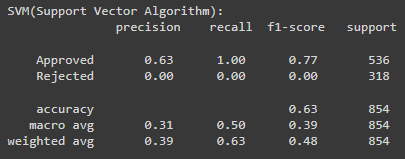
KNN CLASSIFIER:



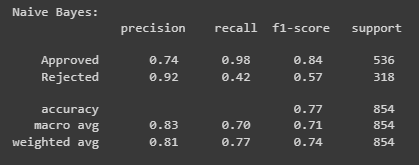
DECISION TREE CLASSIFIER:



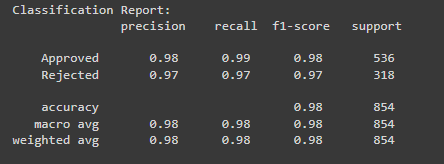
SVM:



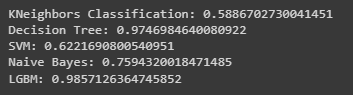
NAÏVE BAYES:



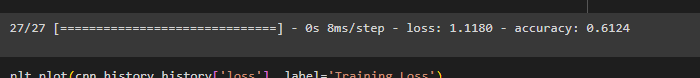
LGBM:



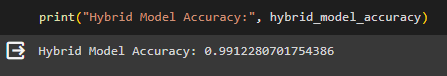
The results for all the algorithms in one place:



CNN Accuracy:



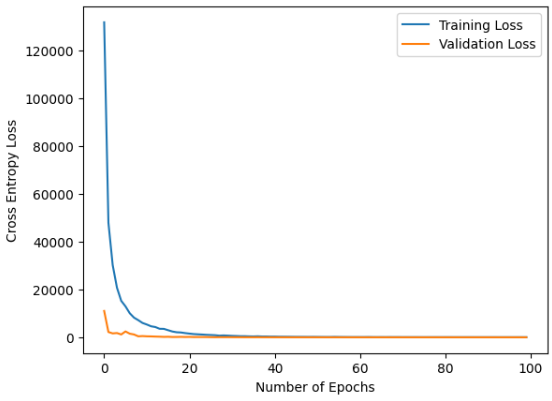
Hybrid Model Accuracy:



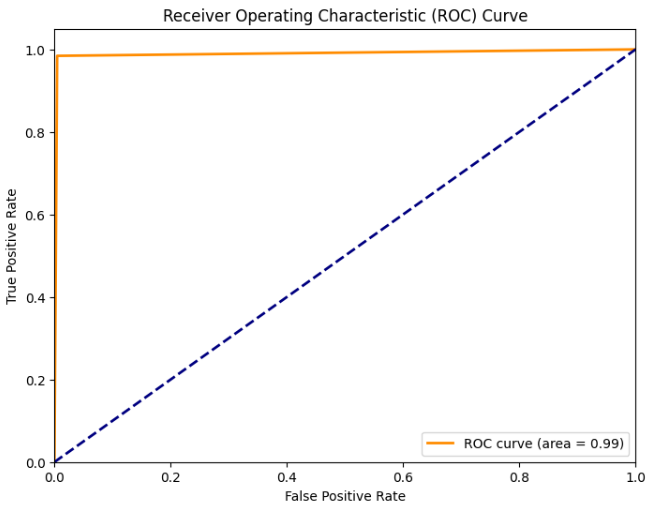
**Thus we know that the best accuracy is of the Hybrid model.**

CNN MODEL PLOTS:

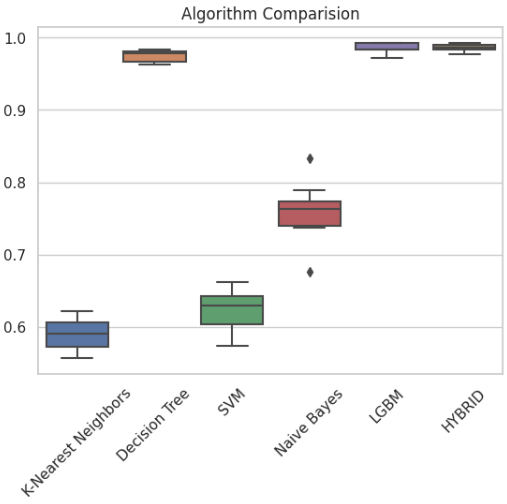
A comparison of the Cross Entropy Loss and the Number of Epochs:



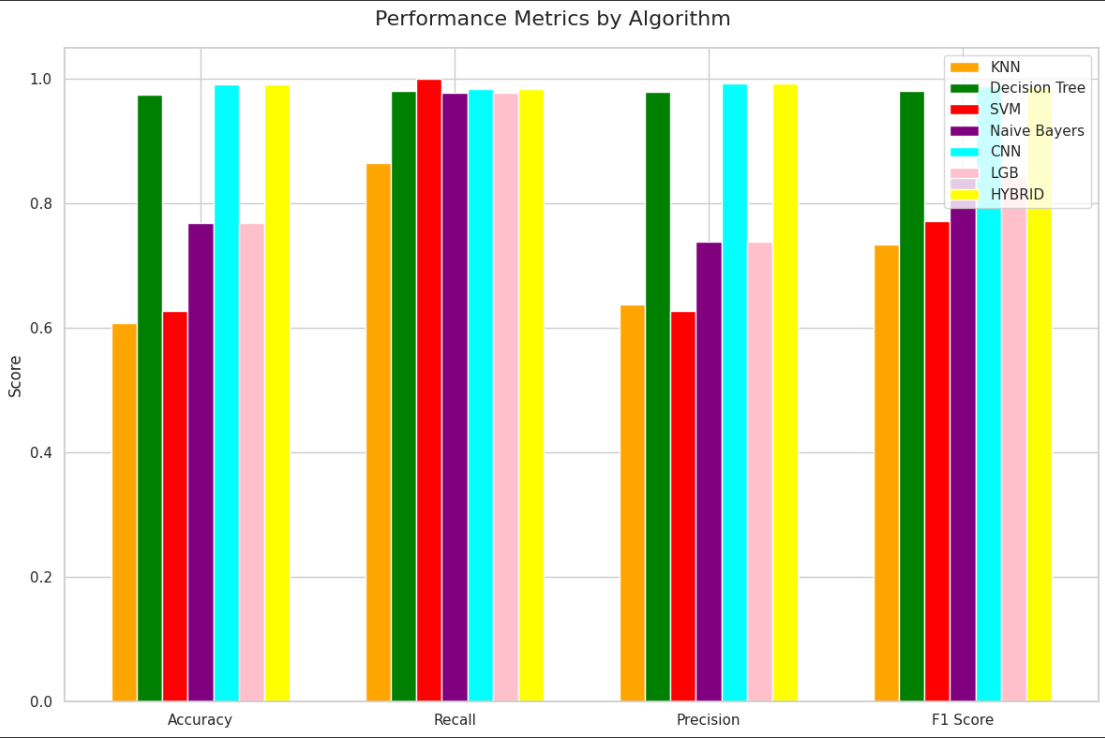
RECEIVER OPERATING CHARACTERISTIC (ROC) CURVE:



BOXPLOT-



HISTOGRAM-



The above graph shows the performance of each algorithm by their respective scores. As is evident from the graph, the Hybrid model outperforms the other algorithms in most aspects and has a fairly steady result. It is the appropriate algorithm for our prediction system.

OUR WEBSITE:

INDIVIDUAL ROLE:

Harsh Jain (102103432)

* Data Preprocessing
* KNN
* Decision Tree
* LGBM
* CNN Model
* Plotting
* Model integration using Flask in the React Website

Shreeya Chatterji (102103447)

* Finding proper dataset
* SVM
* Naïve Bayes Classifier
* Hybrid Model Creation
* Plotting
* Creating the frontend of the website using React JS