Project Phase #3

Fall - CSE 587

Teammate 1: Harikumar Reddy Vengi Reddy (50518453)
Teammate 2: Deekshitha Devalla (50546172)

LOAN APPLICANT DATA FOR CREDIT RISK ANALYSIS

Problem Statement:

In the context of financial lending, the goal is to develop predictive models that evaluate an applicant's creditworthiness in respect to loans.
The purpose of this task is to predict the probability of loan default using the dataset containing applicant characteristics such as age, income, home ownership status, years in job, intended purpose for loan, loan amount, interest rate, historical time of borrower history, and records of missed payments or this would aim at giving lenders a credible tool that helps them make safer choices, reduce default risks and enhance assessment processes in general.

Dataset Description:

The dataset contains all relevant information regarding applicants of loans and their attributes.
Features are age, annual income, home ownership, employment length (in years), loan intent, loan grade, loan amount, loan interest rate, loan status, loan percent income, default history, credit history length.
To check the credit eligibility, a web application is created using flask, this reflects all the work done in phase 1 and phase 2 (EDA, Training with different machine learning algorithms) by automating the process with the pre-processed data set.
In addition to the previous phases, this web application allow user to enter all the column (feature) details by themselves and can check the credit eligibility whether it is rejected or approved.

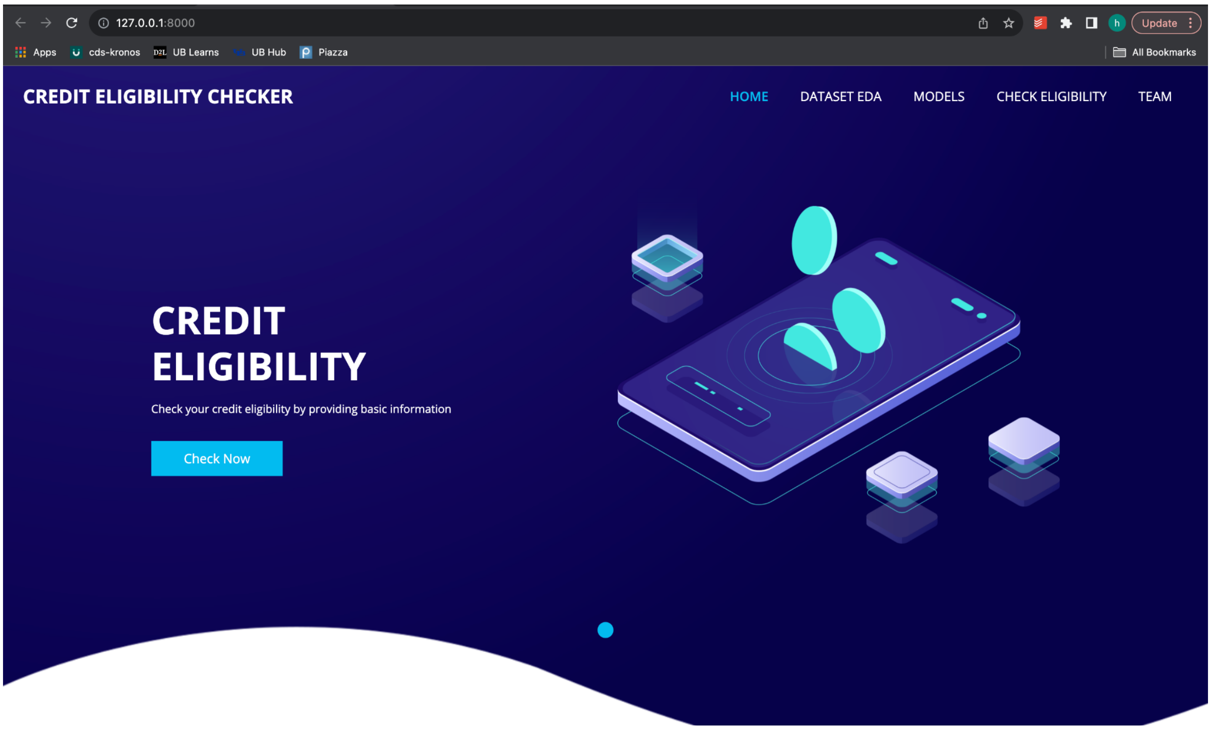
In all the three phases combined, the project includes data, pre-processing, exploratory data analysis, modelling, and creating Web application using python which are needed to predict the credit eligibility for the user inputs.

Working Instructions:

1. Framework: Flask
2. Language: Python
3. Requirements: We need to do *pip install -r requirements.txt* to install all the requirements. Requirements.txt file is in the base location.

4. To start the application: Run *python3 app.py* in the terminal in the base path to start the web server in port 8000.

5. Open the application: Use http://127.0.0.1:8000 to open the web application once the server is up and running.



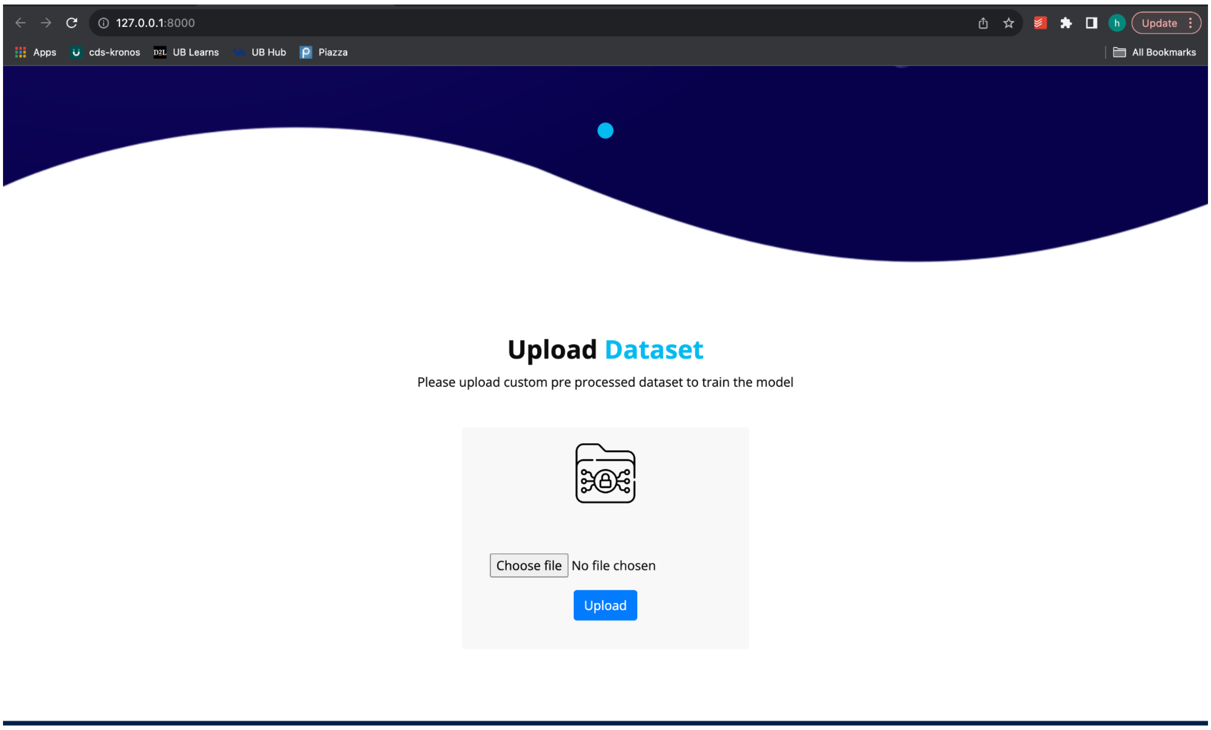
CREDIT ELIGIBILITY CHECKER

CREDIT ELIGIBILITY

Check your credit eligibility by providing basic information

Check Now

One can upload the data set here,



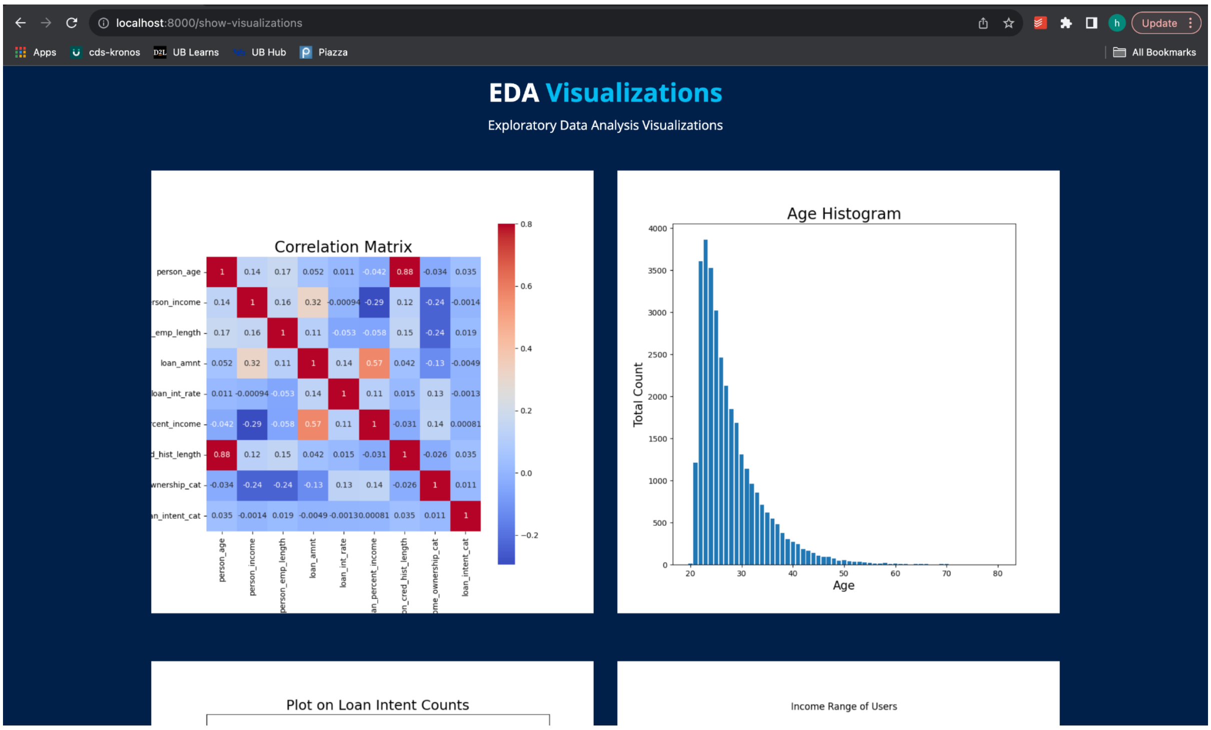
Upload Dataset

Please upload custom pre processed dataset to train the model

Choose file No file chosen

Upload

After uploading the data set, the code in the backend runs by saving the uploaded file which is in the form of CSV and gives the visualisation for the better understanding of the data set.

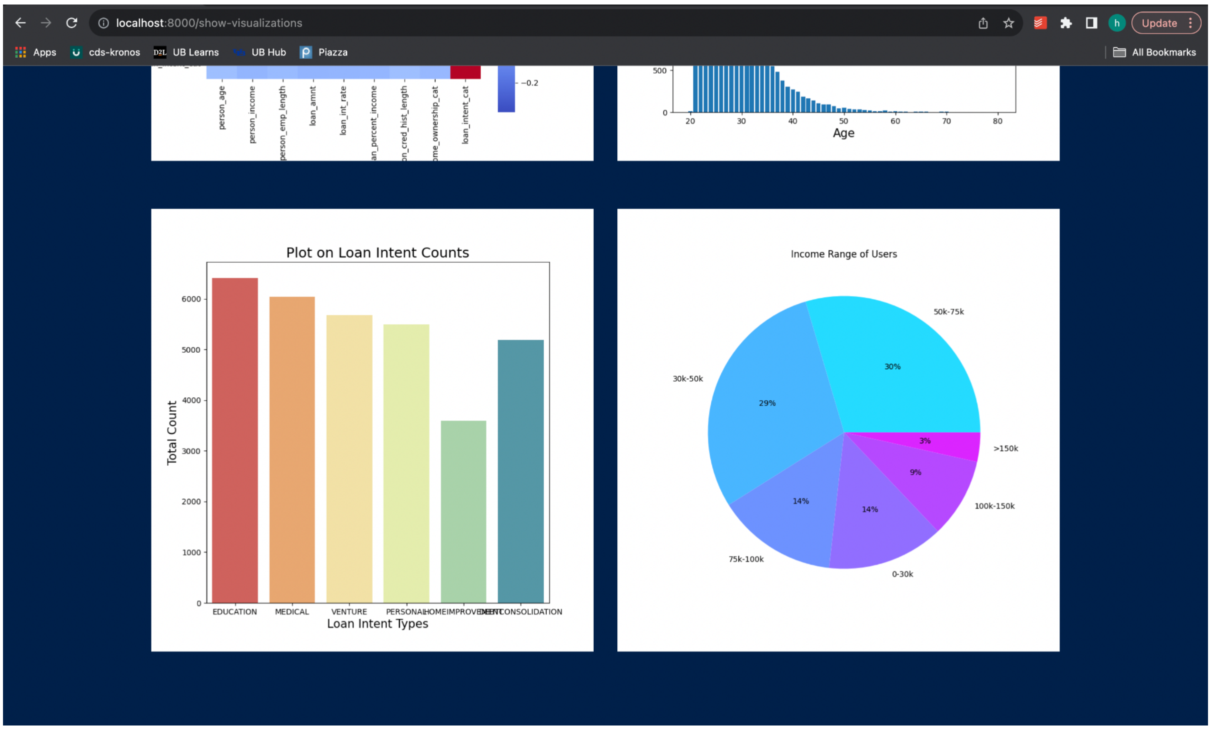


EDA Visualizations

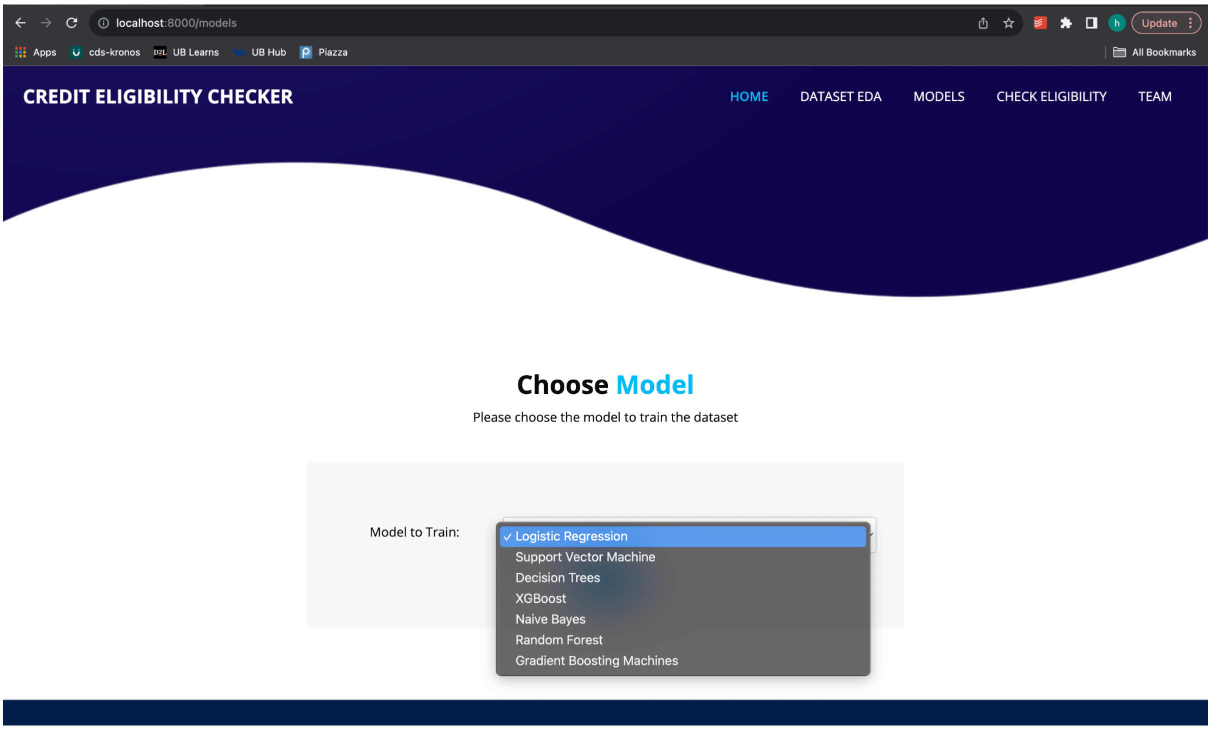
Exploratory Data Analysis Visualizations

Correlation Matrix

	person_age	person_income	person_emp_length	loan_amnt	loan_int_rate	an_percent_income	on_cred_hist_length	me_ownership_cat	loan_intent_cat
person_age	1	0.14	0.17	0.052	0.011	-0.042	0.88	-0.034	0.035
rson_income	0.14	1	0.16	0.32	-0.00094	-0.29	0.12	-0.24	-0.0014
_emp_length	0.17	0.16	1	0.11	-0.053	-0.058	0.15	-0.24	0.019
loan_amnt	0.052	0.32	0.11	1	0.14	0.57	0.042	-0.13	-0.0049
loan_int_rate	0.011	-0.00094	-0.053	0.14	1	0.11	0.015	0.13	-0.0013
rcent_income	-0.042	-0.29	-0.058	0.57	0.11	1	-0.031	0.14	0.00081
d_hist_length	0.88	0.12	0.15	0.042	0.015	-0.031	1	-0.026	0.035
wnership_cat	-0.034	-0.24	-0.24	-0.13	0.13	0.14	-0.026	1	0.011
an_intent_cat	0.035	-0.0014	0.019	-0.0049	-0.0013	0.00081	0.035	0.011	1



The web application allows the user to select a machine learning model to train on the uploaded credit risk dataset, with options including Logistic Regression, Support Vector Machines, Random Forest, Naive Bayes, and others. User on clicking the models button will be redirected into this page as follows



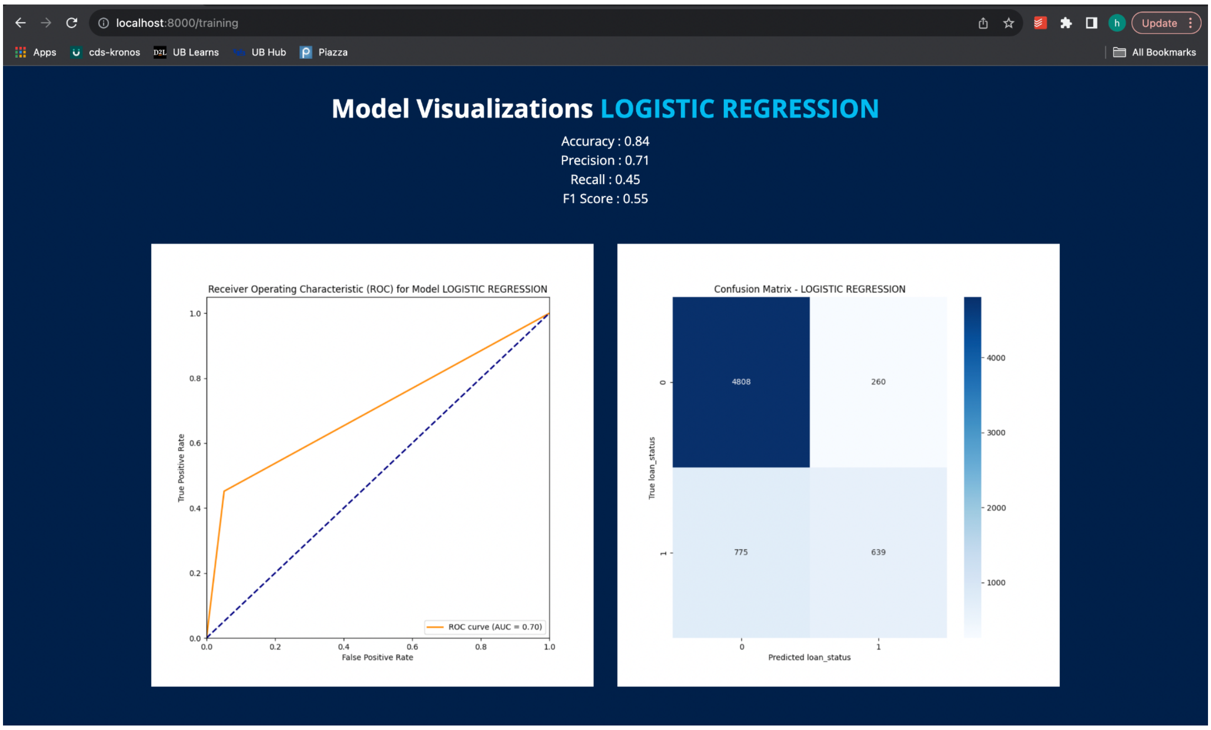
CREDIT ELIGIBILITY CHECKER

Choose Model

Please choose the model to train the dataset

Model to Train: Logistic Regression

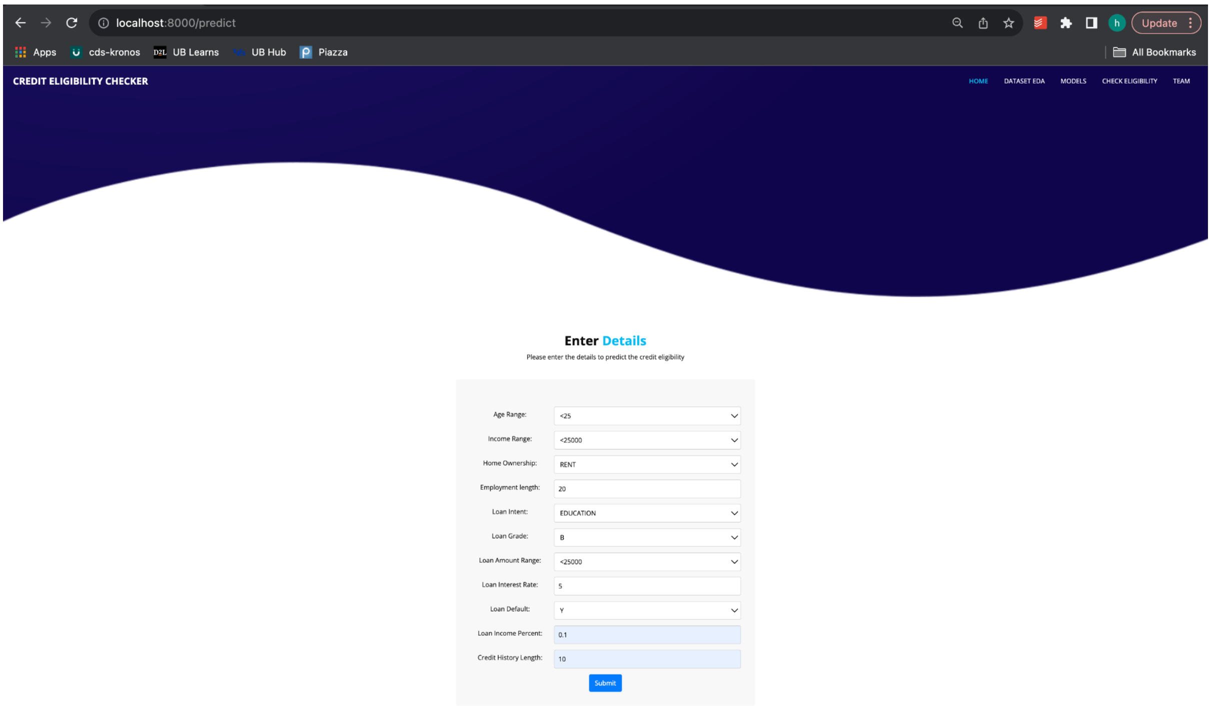
The code in the back end runs for the particular model and generates plots accordingly displaying all the evaluation metrics with ROC and confusion matrix plots.



Model Visualizations LOGISTIC REGRESSION

Accuracy : 0.84
Precision : 0.71
Recall : 0.45
F1 Score : 0.55

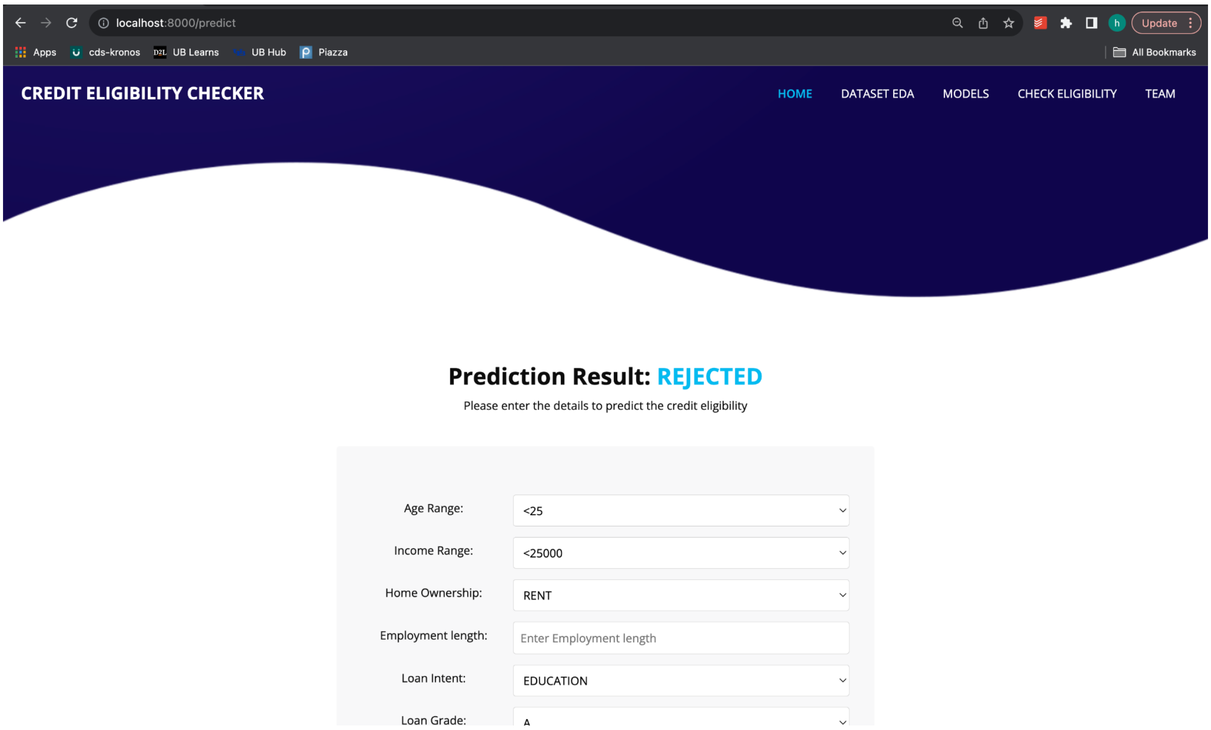
Users can click on check eligibility to check the credit eligibility status by entering all the inputs required. This would result in loan approved or rejected status.



Enter Details

Please enter the details to predict the credit eligibility

Age Range:	<25
Income Range:	<25000
Home Ownership:	RENT
Employment length:	20
Loan Intent:	EDUCATION
Loan Grade:	B
Loan Amount Range:	<25000
Loan Interest Rate:	5
Loan Default:	Y
Loan Income Percent:	0.1
Credit History Length:	10



Prediction Result: REJECTED

Please enter the details to predict the credit eligibility

Age Range:	<25
Income Range:	<25000
Home Ownership:	RENT
Employment length:	Enter Employment length
Loan Intent:	EDUCATION
Loan Grade:	A

Notes on phase 2 models:

In phase 2, the models that are implemented for logistic regression, support vector machines, random forest, decision trees, XGBoost, Naïve Bayes, Gradient Boosting Machines. Phase 2 report has the best tuned values upon selecting different hyperparameters like for SVM, both linear and RBF kernel functions were tried to optimize accuracy. Similarly, the number of trees was tuned for the Random Forest model based on out-of-bag errors. So finally, the best algorithm for credit risk analysis is the random for model as an evaluation metric depicts that it has a highest accuracy of 92% with a procession being 94% and recall of 67%. These metrics are crucial for the analysis as the balance that needs to correctly identify the defaults, while minimising false positives. The result shows the best balance of high precision and high recall, indicating its effectiveness at minimising false positives and false negatives. Key thresholds relevant to the business problem are determined - for example, setting an appropriate probability cut-off for identifying high-risk loans while maintaining recall. Such tuning considerations are included to contextualize the models for real-world usage.

Recommendations:

- From our product, users can learn many things as follows:
- Users, particularly lenders, can understand the risk associated with each loan application by examining the probability of default. This can guide their decision-making process on whether to approve or reject a loan application.

- By analysing the model's feature importance, users can learn which applicant characteristics (e.g., income, employment history) are most predictive of credit risk. This can help lenders focus on the most informative criteria when assessing applications.

- The insights gained from the model can inform the development or adjustment of credit policies and lending criteria to mitigate risk.

- This helps to Solve Problems via automating the credit scoring process can significantly reduce the time and resources spent on manual reviews, increasing efficiency.
- Using a machine learning model helps ensure that decisions are made consistently and are based on quantitative data rather than potentially biased human judgment.

- The model can be continuously updated with new data, allowing lenders to adapt to changing economic conditions and risk profiles over time.

Ideas to Extend the Project:

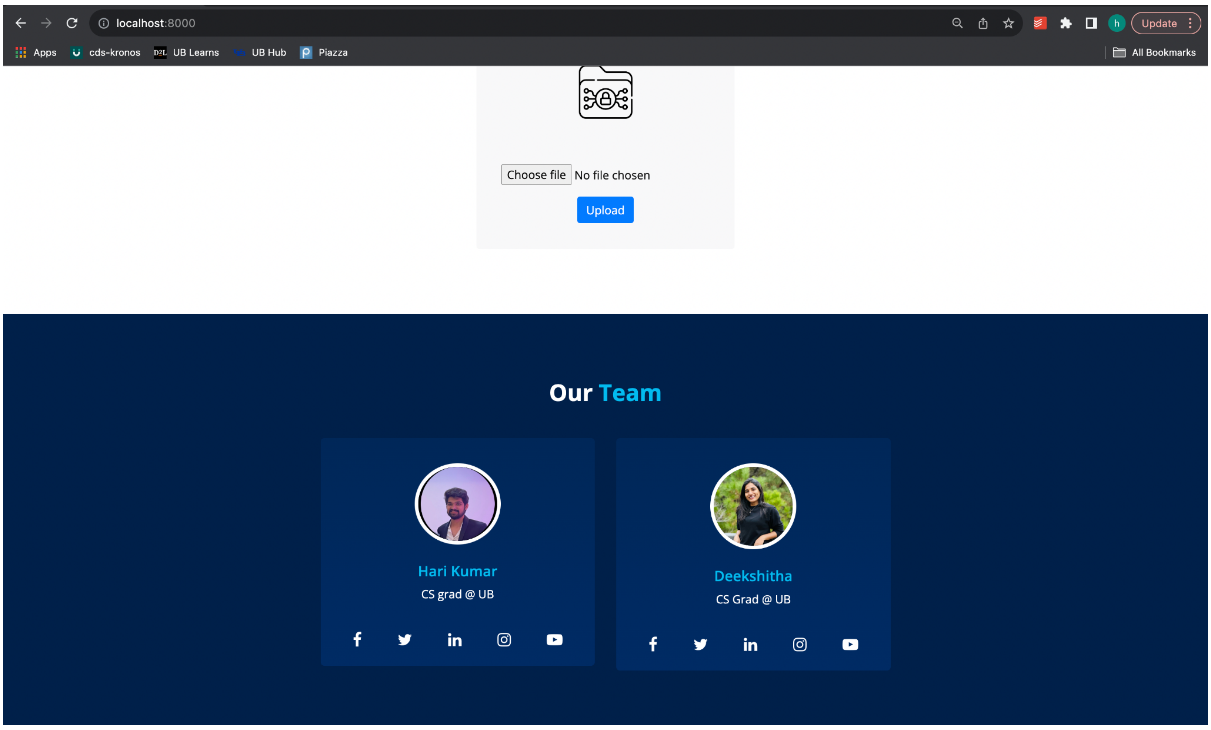
- Integrating real-time data, such as current credit score updates or real-time banking transactions, to provide up-to-date predictions.

- Exploring non-traditional data sources like utility payment histories, rental payments, or social media behaviour to enhance the model's predictive power.

- Developing tools to interpret the model's decisions for transparency, such as explainable AI techniques that can provide reasons for a specific prediction.

- Using the model's output to not only decide on credit eligibility but also to adjust loan terms and interest rates based on the level of risk.

We have also added a section team with team member details


localhost:8000
Apps
cds-kronos
UB Learns
UB Hub
Piazza
All Bookmarks
Update
Choose file
No file chosen
Upload
Our Team
Hari Kumar
CS grad @ UB
Deekshitha
CS Grad @ UB