CREDIT CARD FRAUD DETECTION SYSTEM

CS 584: Machine Learning:

Group no: 51

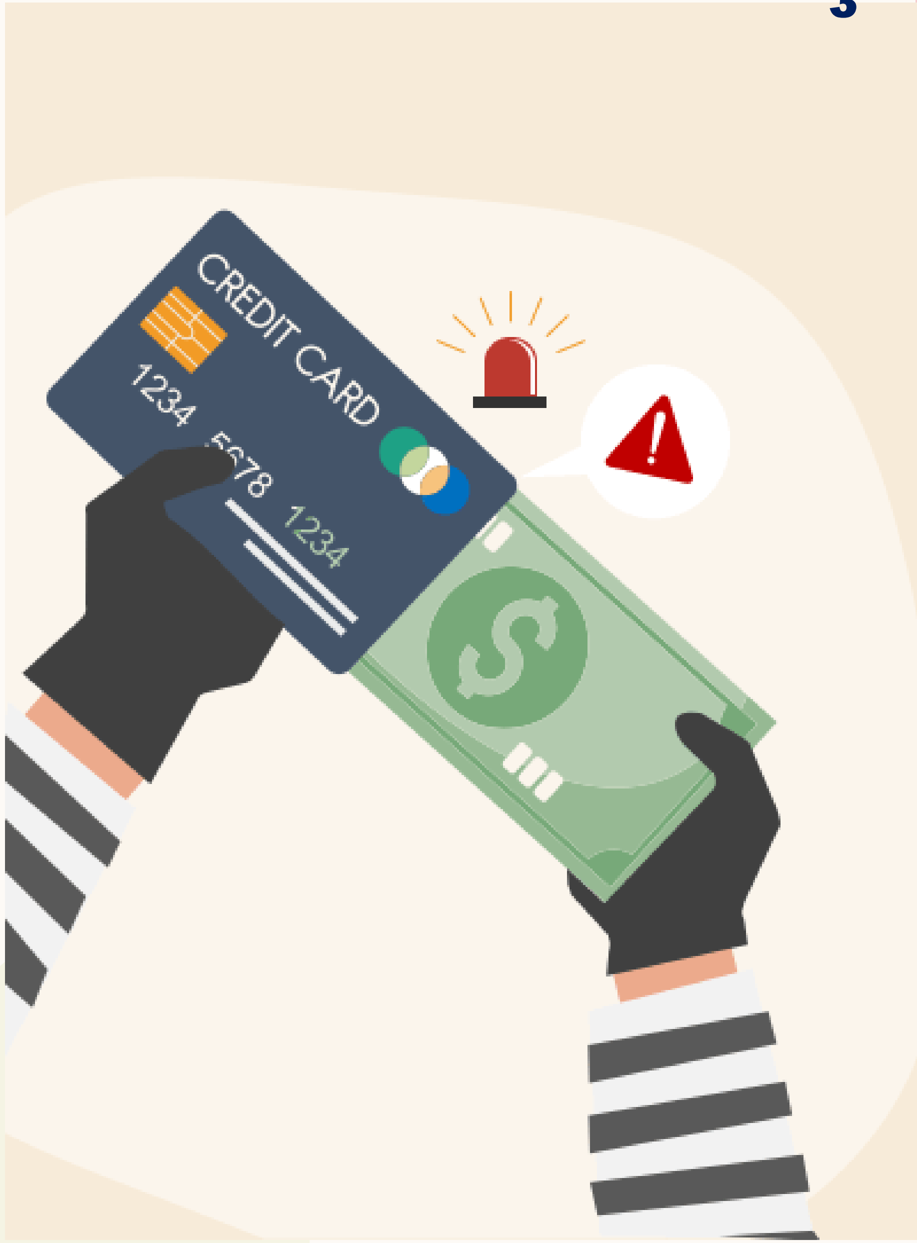
Team Members:
Manavkumar Patel (A20543097)
Nitesha Paatil (A20544932)
Prof: Yan Yan

INTRODUCTION

- Surge in online transactions has resulted to an increase in credit card fraud
- Effective fraud detection is essential to shield clients from monetary losses and unauthorized transactions
- Every year, credit card fraud costs billions and affects both consumers and businesses
- Machine learning algorithms can provide solutions to combat credit card frauds by analyzing large datasets to detect fraud patterns
- Unauthorized purchases are avoided, and default risks are decreased by early fraud detection
- This highlights the critical need to invent trustworthy or reliable techniques to lower credit card fraud in online purchases

PROBLEM STATEMENT

Due to the increase in e-commerce, credit card fraud, making the development of effective fraud detection systems necessary to reduce losses for both financial institutions and consumers.



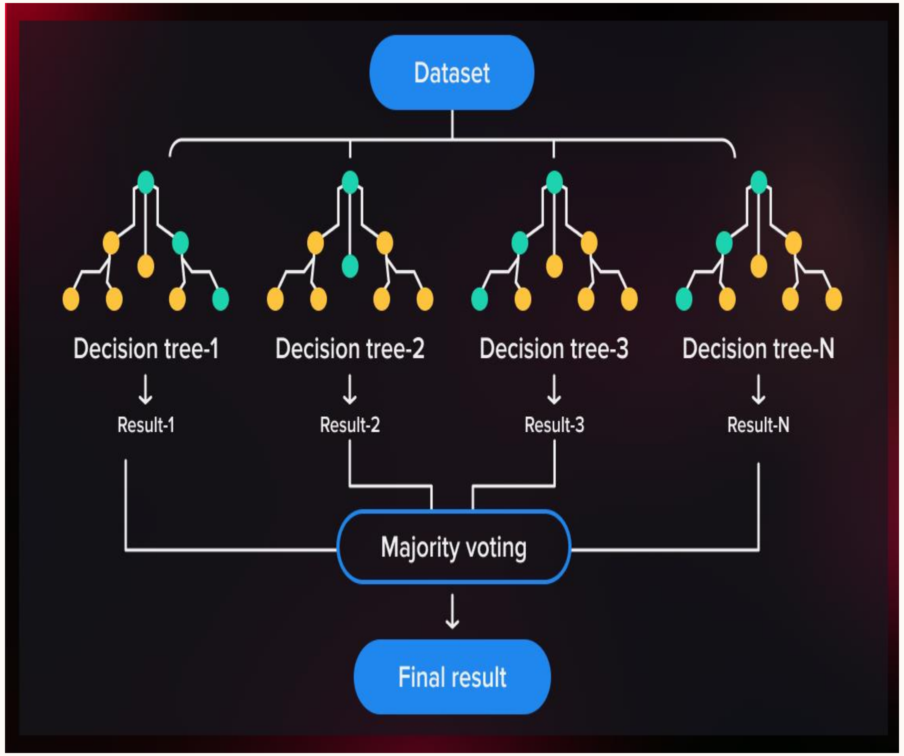
ALGORITHMS USED

- Implemented multiple machine learning algorithms for credit card fraud detection:

1. Logistic Regression
2. AdaBoost Classifier
3. Random Forest
4. Decision Tree
5. Linear Discriminant Analysis
6. Naive Bayes Classifier
7. XGBoost Classifier

- Evaluated the performance of each algorithm
- Random Forest was selected as the most accurate model for the given application

RANDOM FOREST



- Random Forest is an ensemble learning method that constructs multiple decision trees during training
- Each tree is built using a random subset of features and a bootstrap sample of the training data
- Final prediction is made by averaging the predictions of all the individual trees (for regression) or using the mode of the classes (for classification)
- Known for its ability to reduce overfitting and improve generalization, making it a popular choice for various machine learning tasks
- Versatile, handles large datasets, maintains accuracy with missing data, and estimates feature importance

FRAUD DETECTION SYSTEM

- Uses data analysis, machine learning, and pattern recognition to identify and prevent fraudulent activities
- These systems are crucial for industries such as banking, insurance, e-commerce, and healthcare to protect against financial losses and maintain trust with customers
- By analyzing transaction data and user behavior, fraud detection systems can detect anomalies and potential fraud
- Continuous adaptation and learning enable these systems to evolve and improve their effectiveness in detecting new fraud patterns and technologies

FEATURE EXTRACTION

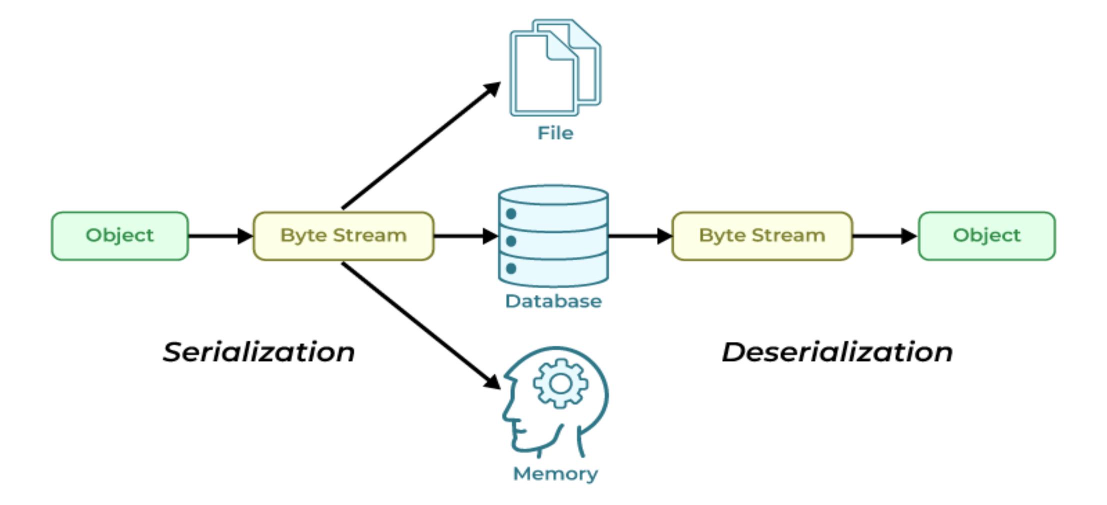
A crucial step in data analysis which transforms raw data into a more suitable format for further processing

Correlation matrix:

- Helps in feature extraction as it provides observations into the relationships between variables in the dataset
- It helps detect redundant or irrelevant features that can be removed
- Simplifies the model by selecting the most predictive features
- Improves model performance and efficiency by reducing complexity
- Helps detect multicollinearity, where features are highly correlated, which can adversely affect the model's performance
- Guides feature selection process for a more accurate and streamlined model

PICKLE MODEL

- Pickle is a Python module used for serializing and deserializing Python objects
- It allows objects to be converted into a byte stream for storage or transmission
- Pickle can be used to save machine learning models for later use without retraining
- It provides a convenient way to store complex data structures in Python
- Pickle is widely used in Python for saving and loading data objects in various applications



FLASK

- Flask is a lightweight web application framework for Python
- It allows you to quickly build web applications with Python
- Flask is easy to learn and use, making it ideal for beginners and small projects
- It provides tools and libraries for tasks such as URL routing, template rendering, and handling HTTP requests
- Flask is widely used for building web APIs, websites, and web applications

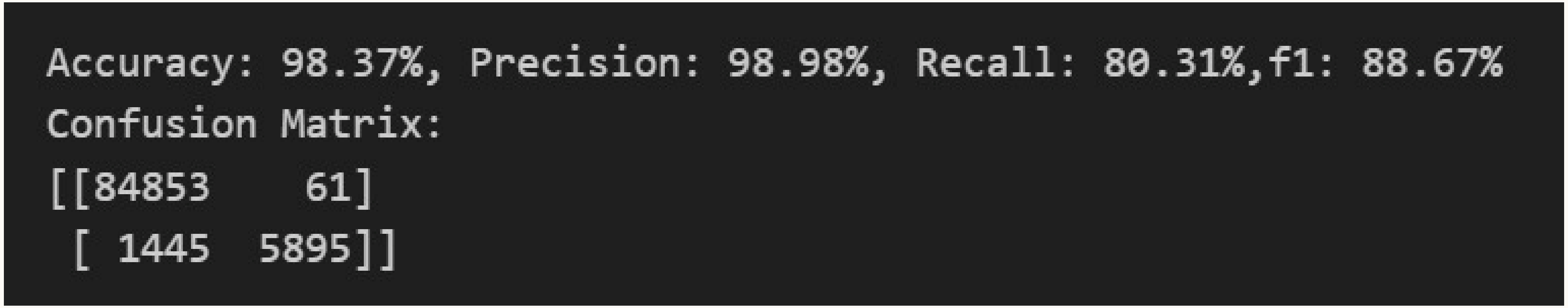
USER INTERFACE (UI)

- Provides insights into credit card defaulters based on their respective attributes
- Allows users to explore and analyze data related to credit card defaultees, such as demographics, transaction history, and payment behavior
- Presents this information in an easy-to-understand format, using charts, graphs, and tables to visualize the data
- Users can interact with the UI to filter and sort the data, gaining valuable insights into the characteristics of credit card defaultees

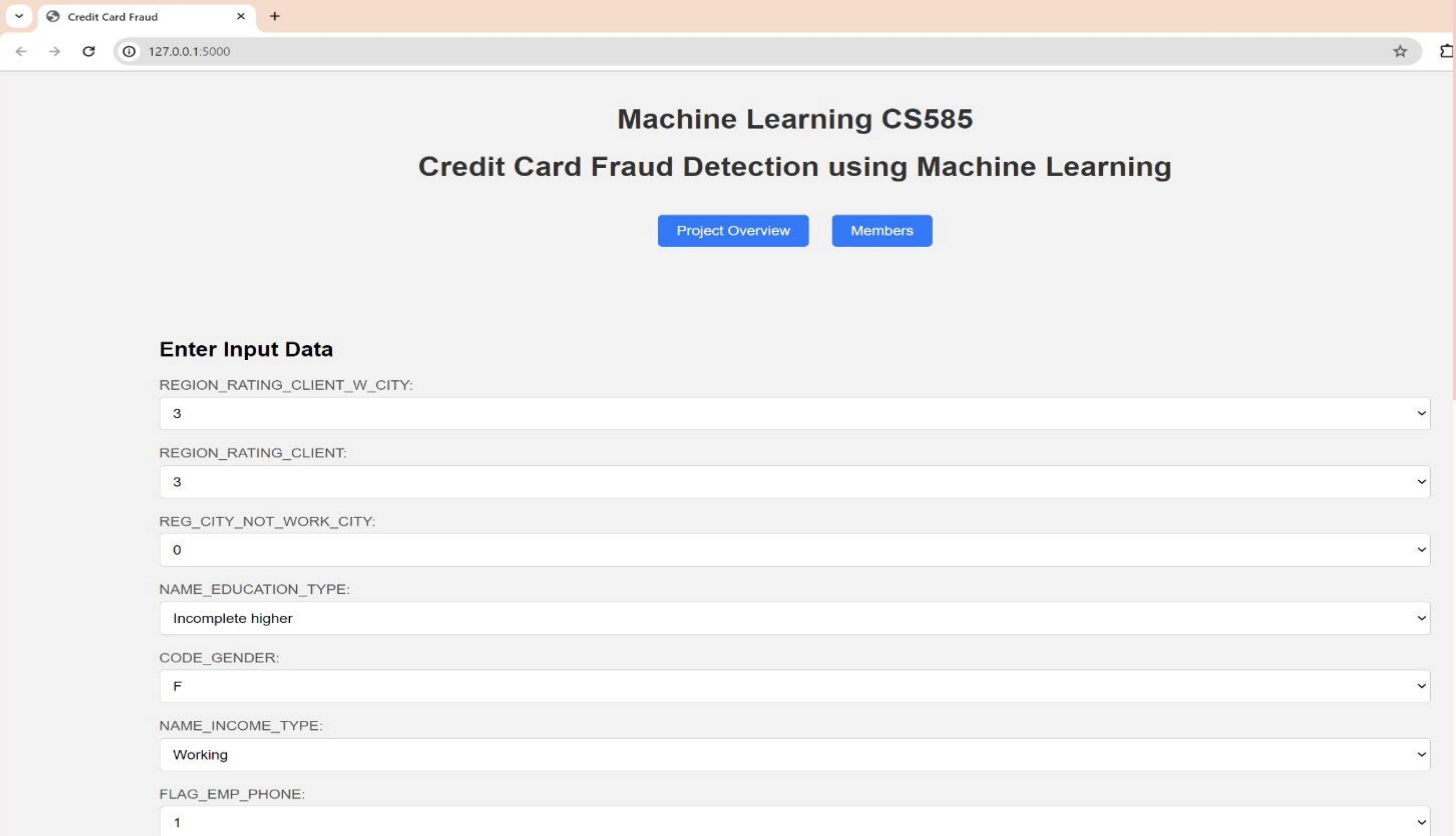
DEMONSTRATION

RESULTS

- After analyzing various algorithms, the credit card fraud detection model achieves remarkable accuracy and precision, surpassing 90%
- The Random Forest algorithm stands out as the most accurate option for this application, ensuring robust performance.



```
Accuracy: 98.37%, Precision: 98.98%, Recall: 80.31%,f1: 88.67%
Confusion Matrix:
[[84853    61]
 [ 1445  5895]]
```



Machine Learning CS585

Credit Card Fraud Detection using Machine Learning

Enter Input Data

REGION_RATING_CLIENT_W_CITY: 3

REGION_RATING_CLIENT: 3

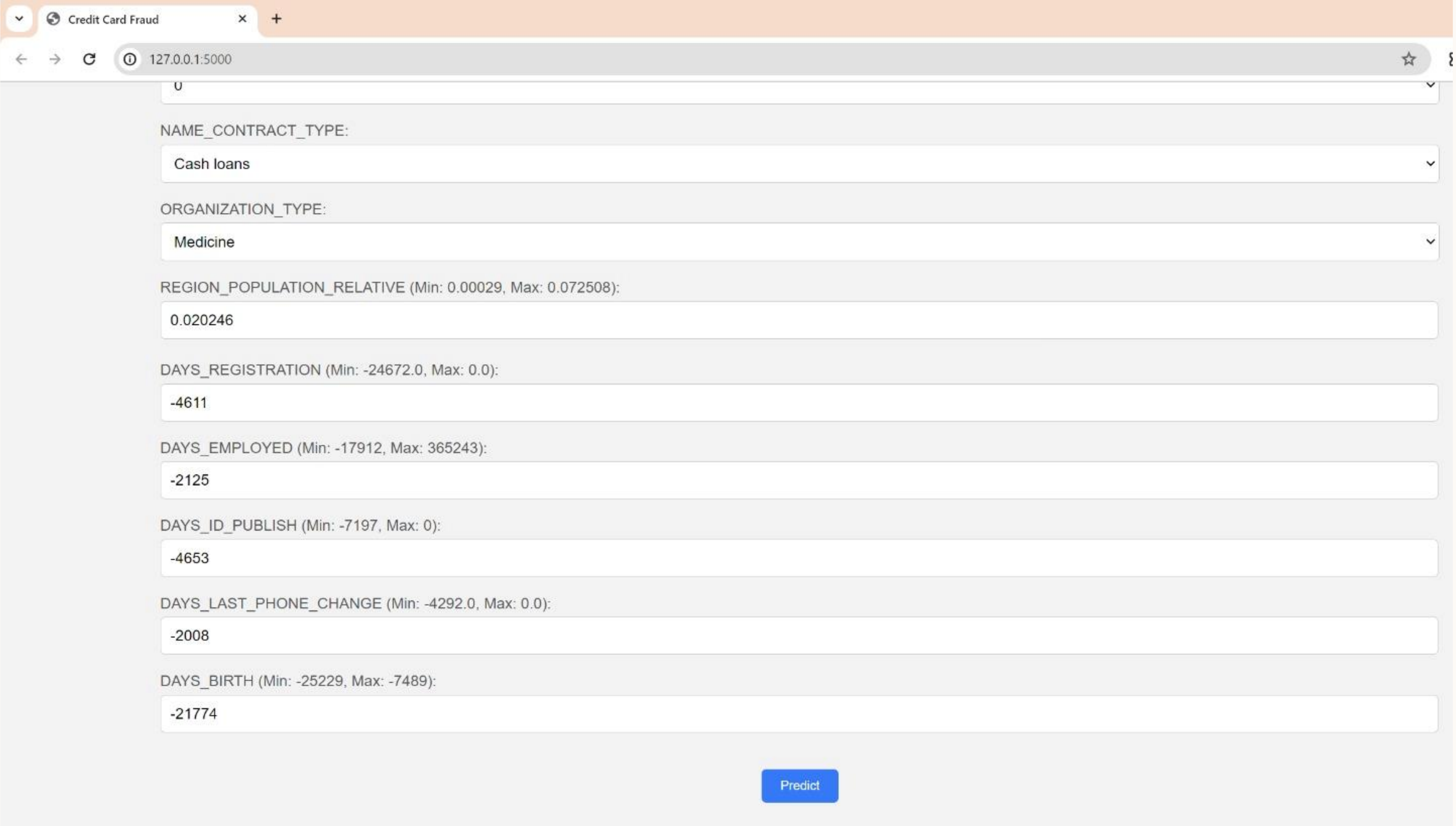
REG_CITY_NOT_WORK_CITY: 0

NAME_EDUCATION_TYPE: Incomplete higher

CODE_GENDER: F

NAME_INCOME_TYPE: Working

FLAG_EMP_PHONE: 1



0

NAME_CONTRACT_TYPE:
Cash loans

ORGANIZATION_TYPE:
Medicine

REGION_POPULATION_RELATIVE (Min: 0.00029, Max: 0.072508):
0.020246

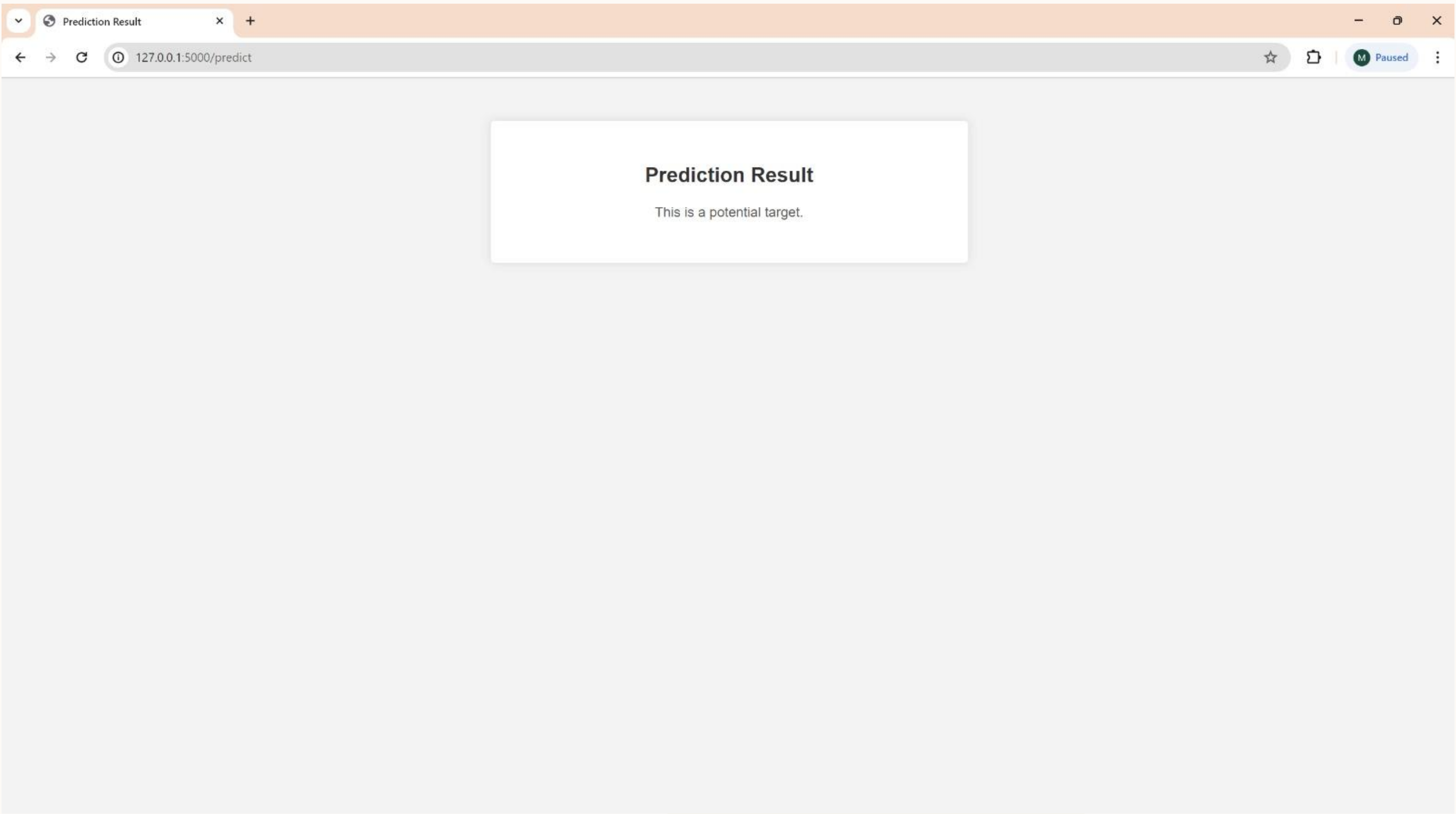
DAYS_REGISTRATION (Min: -24672.0, Max: 0.0):
-4611

DAYS_EMPLOYED (Min: -17912, Max: 365243):
-2125

DAYS_ID_PUBLISH (Min: -7197, Max: 0):
-4653

DAYS_LAST_PHONE_CHANGE (Min: -4292.0, Max: 0.0):
-2008

DAYS_BIRTH (Min: -25229, Max: -7489):
-21774



Prediction Result

This is a potential target.

THANK YOU

Manavkumar Patel

Nitesha Paatil