RAJALAKSHMI ENGINEERING COLLEGE
An AUTONOMOUS Institution
Affiliated to ANNA UNIVERSITY, Chennai

25 years OF ACADEMIC EXCELLENCE

Department of Computer Science and Engineering

MIGRATION FORECASTING : A MACHINE LEARNING APPROACH

Mrs. M.Divya M.E.,
Department of CSE

Prashanth P
220701513

Problem Statement and Motivation

Problem Statement:

Accurate prediction of migration activities such as arrivals, departures, and net migration is critical for national policy formulation, economic planning, and resource management. This project focuses on building a machine learning model that can classify migration events based on factors like country, citizenship, and migration value.

Problem Statement and Motivation

Motivation:

Migration significantly impacts the socio-economic structure of a country. In regions like New Zealand, understanding migration patterns helps in optimizing infrastructure, social services, and employment strategies. Traditional statistical tools often struggle with high-dimensional or categorical data. By leveraging machine learning, particularly Support Vector Machines (SVM), this project introduces a more efficient and scalable approach to migration trend prediction.

Existing System

The current approach to analyzing and forecasting migration trends is largely based on traditional statistical techniques such as regression analysis, time-series forecasting, and historical trend evaluation. These methods, while useful for understanding general patterns, have several limitations when applied to large and complex datasets. They often require manual intervention for feature selection, are not well-suited for handling categorical or high-dimensional data, and struggle to adapt to rapid changes in migration dynamics. Moreover, these models assume linear relationships and may overlook non-obvious patterns present in the data. As a result, there is a growing need for more flexible and accurate systems capable of learning from the data itself — a gap that machine learning models are well-positioned to fill.

Objectives

The primary objective of this project is to develop a machine learning-based model that can accurately classify migration events as arrivals, departures, or net migration using historical data. This involves preprocessing raw migration data, converting categorical variables into numerical formats, handling missing values, and selecting meaningful features for prediction. By implementing and training a Support Vector Machine (SVM) model, the project aims to improve the accuracy and efficiency of migration classification compared to traditional methods.

Abstract

This project presents a machine learning-based approach to classify migration trends using historical data from New Zealand. The dataset includes information such as country, citizenship, migration value, and migration type (arrival, departure, or net). After cleaning and preprocessing the data, including handling missing values and encoding categorical variables, a Support Vector Machine (SVM) model is trained to predict the type of migration event.By leveraging machine learning, the system can identify complex patterns in migration behavior, offering valuable insights for policy planning and demographic analysis.

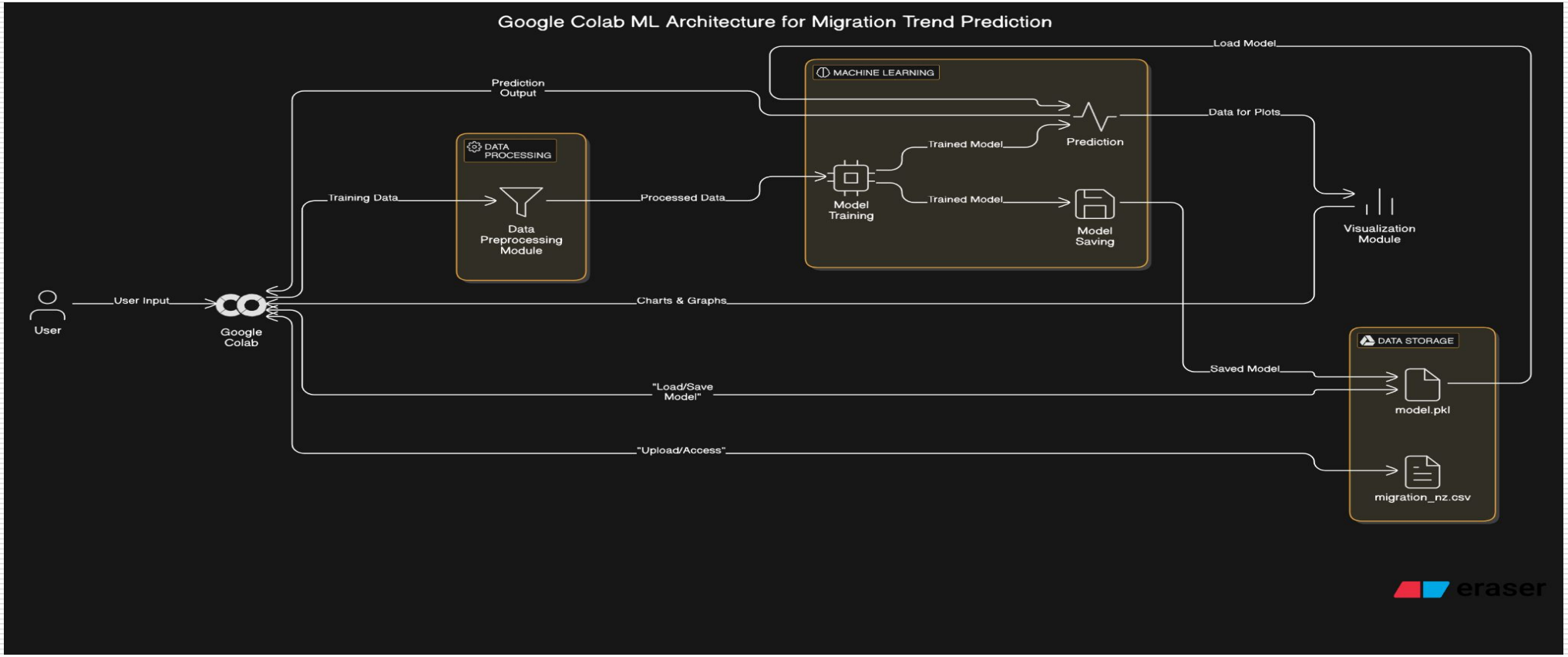
Proposed System

The proposed system uses a machine learning approach to classify migration data into arrivals, departures, and net migration. It focuses on data preprocessing and employs a Support Vector Machine (SVM) model for accurate classification.

Highlights:

- Uses SVM for multi-class classification.
- Converts categorical data using factorization.
- Handles missing values with median imputation.
- Splits data into training and testing sets.

System Architecture



List of Modules

1. Pandas

- For data loading, exploration, and preprocessing.
- Reads CSV files and handles DataFrame operations.

2. Scikit-learn (sklearn)

- `train_test_split`: Splits dataset into training and test sets.
- `svm`: Implements Support Vector Machine model for classification.

List of Modules

3. Seaborn

- High-level data visualization library.
- Used for plotting statistical graphics.

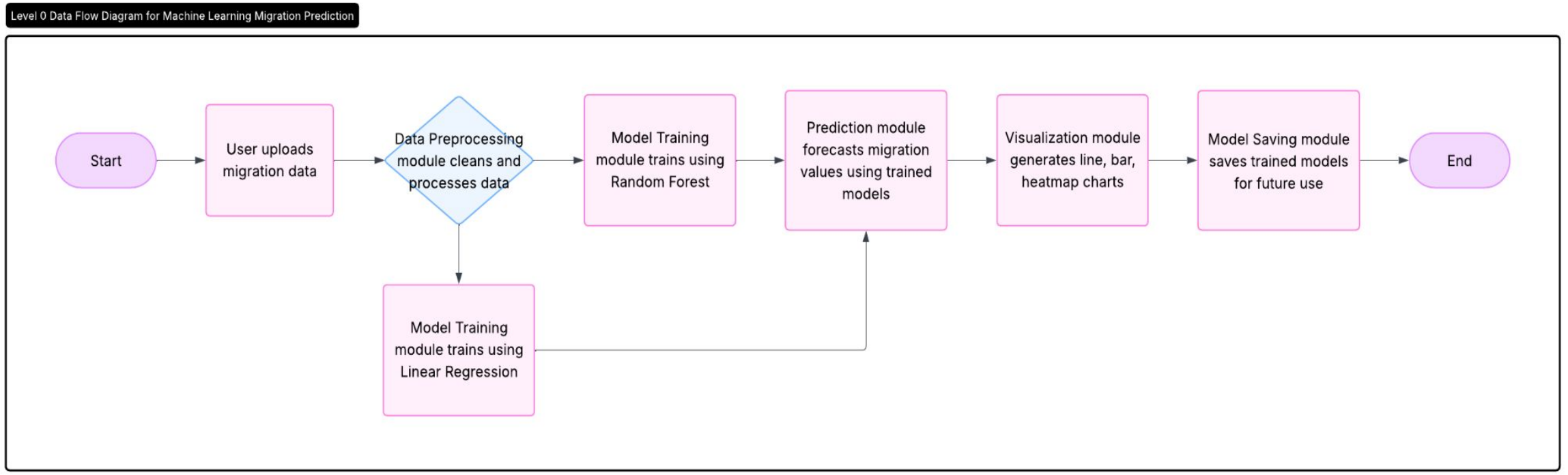
4. Matplotlib.pyplot

- Low-level plotting library.
- Supports detailed and customized visualizations.

5. Google Colab Drive

- Mounts Google Drive in Colab to access files like datasets.

Functional Description for each modules with DFD



Implementation & Results of Module

Implementation Steps

1. **Preprocessing**
    - CSV loaded and cleaned using `pandas`.
    - Categorical labels converted into numerical values.
2. **Data Visualization**
    - **Heatmap**: Shows low correlation between `Measure`, `Year`, `Value`, `CountryID`, and `CitID`.
    - **Bar Chart**: Displays yearly migration trends from 1979 to 2016.

Results of Module

Highest Migration Year:

2016, with approximately **990,000** migrations.

Consistent Growth Period:

From **2000 to 2016**, values steadily increase from ~650,000 to ~990,000.

Lowest Dip:

Around **1984–1985**, with values below **350,000**.

Sharp Rises:

After **2013**, growth accelerates significantly.

Conclusion & Future Work

In this project, we developed a basic machine learning model to predict migration trends using historical data for New Zealand. By cleaning the data, encoding key features, and applying Random Forest and Linear Regression, we achieved reasonable predictive accuracy. Visual analysis supported the model's findings, revealing trends over the years. In the future, we plan to include more variables, explore advanced models like LSTM for time-series prediction, and develop a scalable system for country-specific insights.

References

Scikit-learn – Pedregosa et al., *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research, 2011.
https://scikit-learn.org

Pandas – McKinney, Wes. *Data Structures for Statistical Computing in Python*, 2010.
https://pandas.pydata.org

Data Source – New Zealand Migration Statistics (from the user-provided dataset `migration_nz.csv`)

Google Colab – Google Research, *Colaboratory: Machine Learning Notebooks*, 2017.
https://colab.research.google.com

Thank You