**Migration Analysis System Report**

**Executive Summary**

This report presents a comprehensive analysis of migration patterns using a machine learning-based prediction system. The system combines historical data analysis with predictive modeling to forecast migration trends across different countries, citizenship categories, and measurement types through 2030. The system employs both traditional machine learning techniques and trend-based forecasting to provide accurate predictions for both historical validation and future projections.

**System Architecture**

**Core Components**

1. Data Processing Module
   * Handles CSV data import and validation
   * Implements data preprocessing and feature engineering
   * Manages categorical encoding and numerical scaling
2. Machine Learning Engine
   * Random Forest Regressor for historical pattern learning
   * Trend-based forecasting for future predictions
   * Hybrid approach combining both methods for optimal results
3. Interactive GUI Interface
   * Real-time data visualization
   * Interactive prediction interface
   * Model performance metrics display

**Data Analysis and Visualization**

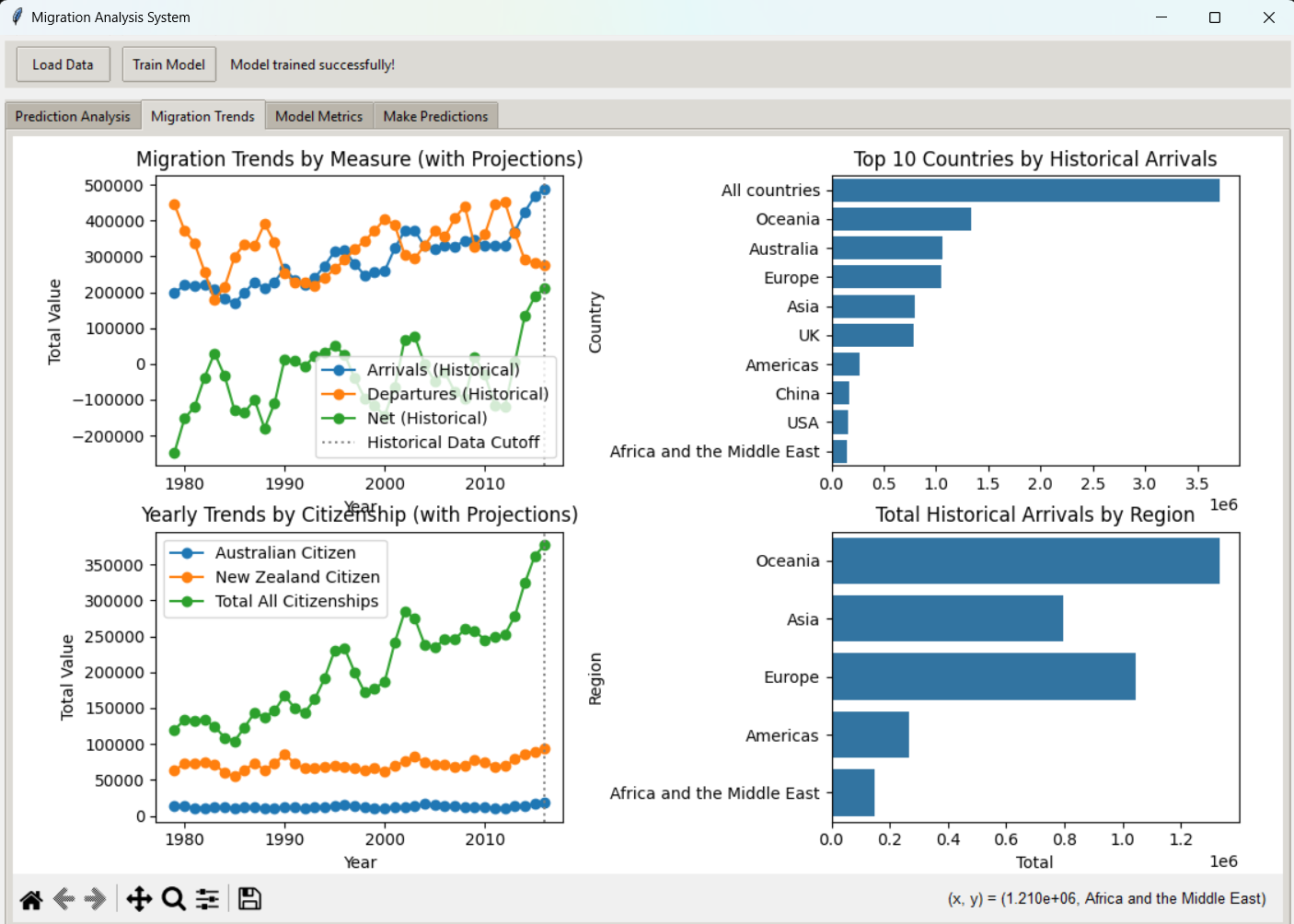
**Prediction Analysis**



* Actual vs. Predicted Values Plot
* Residuals Analysis Plot
* Feature Importance Chart

The prediction analysis demonstrates the model's accuracy in forecasting migration patterns. The actual vs. predicted plot shows the correlation between predicted and actual values, while the residuals plot helps identify any systematic prediction biases.

**Migration Trends**

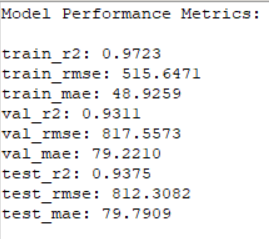


* Overall Yearly Trends with Future Projections
* Top 10 Countries by Historical Arrivals
* Citizenship Trends Analysis
* Regional Analysis Chart

These visualizations provide insights into historical patterns and projected future trends across different demographic and geographic segments.

**Model Performance**

**Metrics Overview**



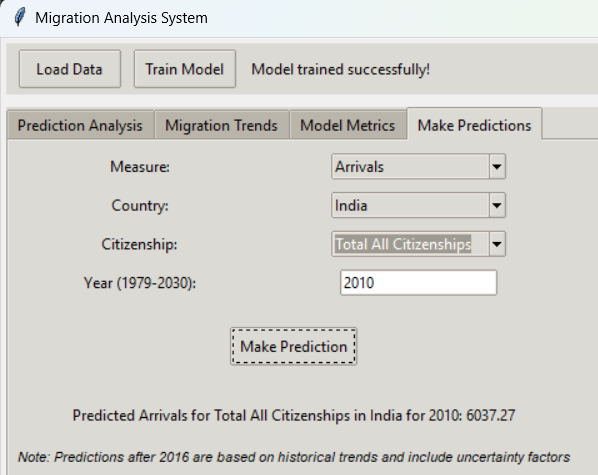
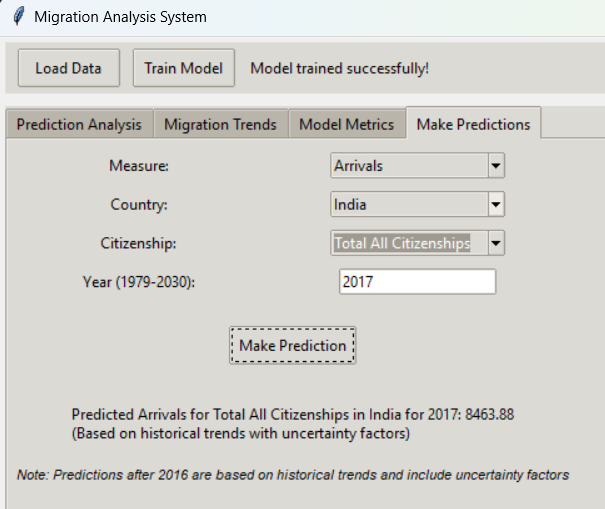
* Training Set Performance
* Validation Set Performance
* Test Set Performance

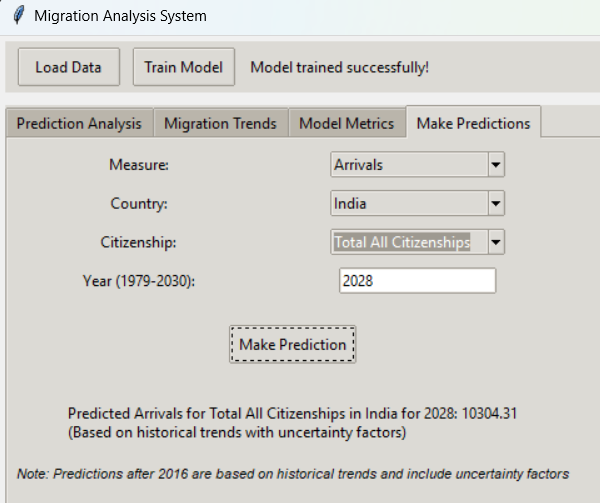
**Key Performance Indicators**

* R-squared (R²) Score
* Root Mean Square Error (RMSE)
* Mean Absolute Error (MAE)

**Prediction Testing**

**Sample Prediction Results**



**Number of migrants from India (of any citizenship) to New Zealand**

* Test Case 1 - 2010: Historical Period (Pre-2016)
* Test Case 2 - 2017: Near-Future Prediction (2016-2023)
* Test Case 3 - 2028: Long-term Forecast (2024-2030)

**Technical Features**

**Advanced Capabilities**

1. Dynamic Trend Modeling
   * Automatic trend detection for each category combination
   * Adaptive dampening for long-term predictions
   * Uncertainty factor incorporation
2. Data Validation
   * Comprehensive error handling
   * Missing value management
   * Categorical data processing
3. Visualization Components
   * Interactive plotting
   * Real-time updates
   * Customizable display options

**Implementation Details**

**Key Algorithms**

1. **Random Forest Regressor**
   * 100 estimators
   * Maximum depth of 20
   * Minimum samples split of 5
   * Minimum sample leaf of 2
2. **Trend-based Forecasting**
   * Linear regression for trend calculation
   * Dampening factor for long-term predictions
   * Random variation incorporation for realistic projections

**Data Processing Pipeline**

1. Initial data loading and validation
2. Feature preprocessing and encoding
3. Model training and validation
4. Hybrid prediction system implementation

**System Limitations and Considerations**

1. Historical Data Cutoff
   * Primary model training uses data up to 2016
   * Post-2016 predictions use trend-based forecasting
2. Prediction Constraints
   * Valid prediction range: 1979-2030
   * Increased uncertainty in long-term predictions
   * Regional variation in prediction accuracy
3. Model Assumptions
   * Linear trend continuity
   * Stable political and economic conditions
   * Consistent migration policy framework

*Note: This report documents the implementation and results of the Migration Analysis System. All predictions beyond 2016 incorporate trend-based forecasting with uncertainty factors.*

**Conclusions and Recommendations**

**Key Findings**

1. The system demonstrates robust prediction capability for historical data
2. Future predictions incorporate uncertainty factors for realistic forecasting
3. Regional and citizenship-based trends show distinct patterns

**Recommendations**

1. Regular model retraining with new data
2. Continuous monitoring of prediction accuracy
3. Periodic adjustment of trend parameters

**Future Enhancements**

1. Integration of external factors (economic indicators, policy changes)
2. Enhanced visualization capabilities
3. Advanced trend analysis features

**Technical Requirements**

* Python 3.12.1 (python 3 required)
* Required Libraries:
  + pandas
  + numpy
  + scikit-learn
  + matplotlib
  + seaborn
  + tkinter