

Energy Distribution in Latin America

BY: MARIA PINZON



Problem Statement

Analyze energy distribution in Latin America to analyze areas of inequality that can aid in indicating regions of need for investment

Data

- “Latin America and Caribbean – Utility Benchmark Database”
- Data collected from World Bank Group
- Raw data: 4074 rows & 31 columns

Data Wrangling

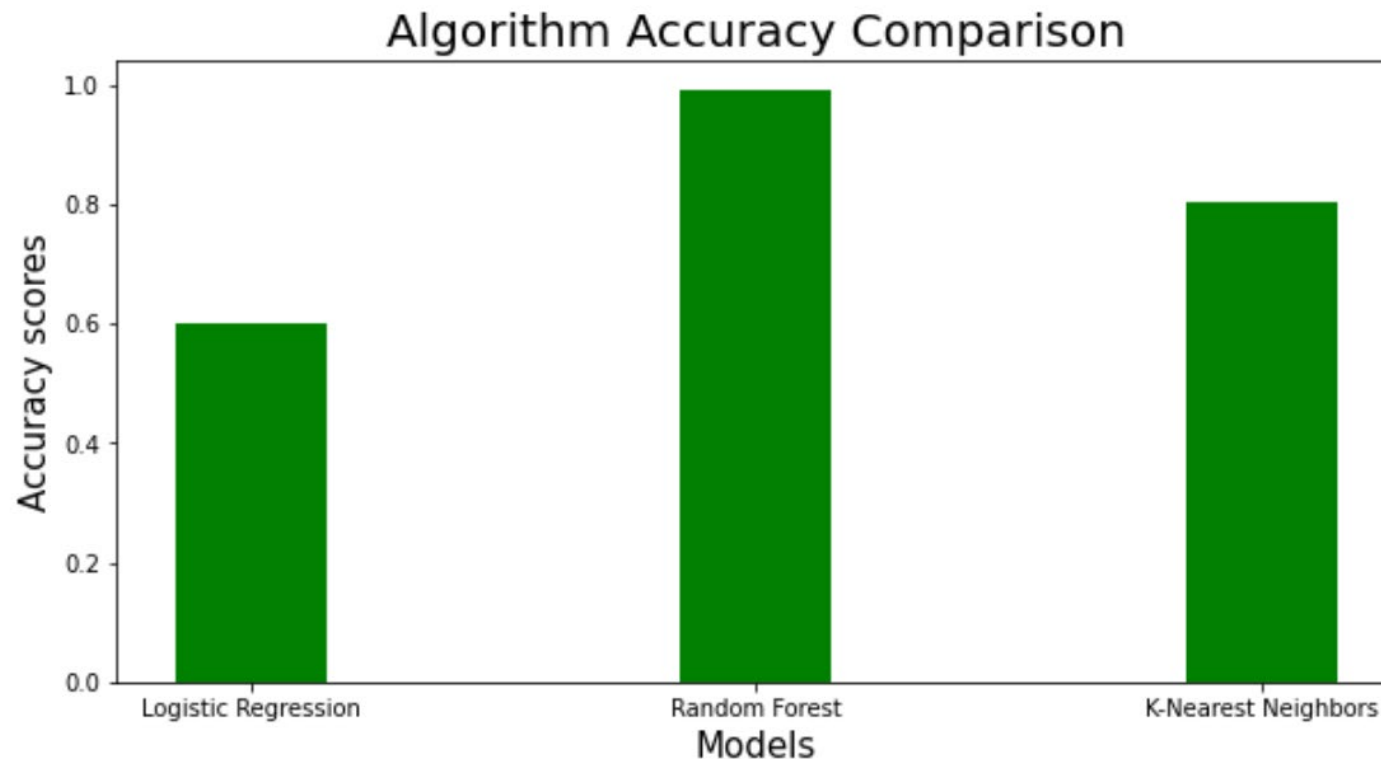
- Column names were extensive and needed to be condensed
- Multiple columns contained a large amount null values – columns with over 80% of null values were dropped
- Data was grouped by utility and null values were filled with mean

Exploratory Data Analysis

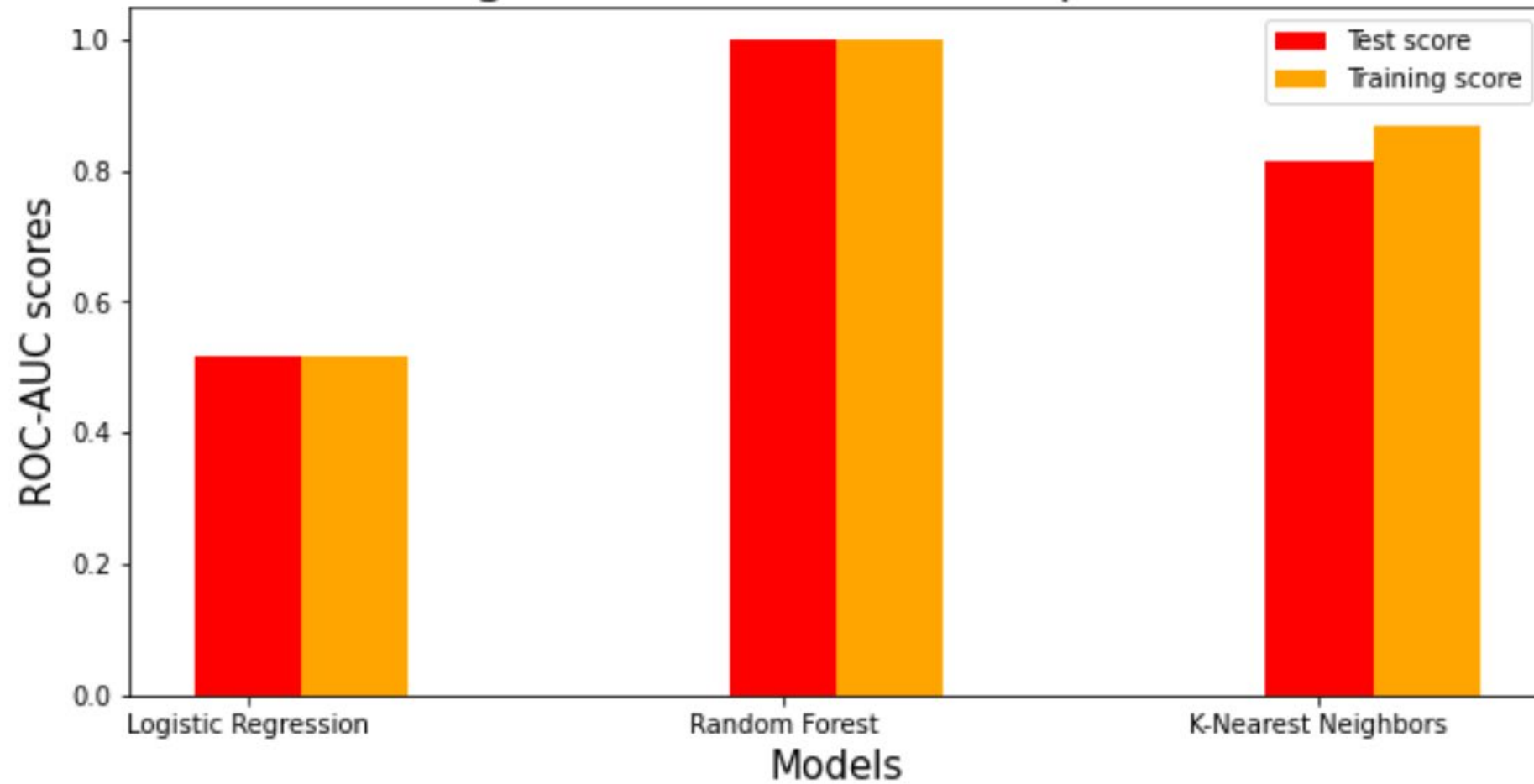
- After cleaning – data: 4073 rows & 24 columns
- Most data is numerical
- Grouped data by utility to show residential energy connections

Modeling (Selection and Comparison)

- Models analyzed: Logistic Regression, Random Forest, K-Nearest Neighbors



Algorithm ROC-AUC Comparison



Conclusion and Further Analysis

- Model chosen: Random Forests
- The analysis can be further analyzed by adding more current data and analyzing important features.