"Solar Radiation on Agricultural Soil Conditions Using Data Science and AI Models"

Author: Zonghong Yu*

*DSAN5550 ** Georgetown University: Data Science & Analytics *04/25/2024

GitHub: https://github.com/zy236yuz5/512-project-group-05

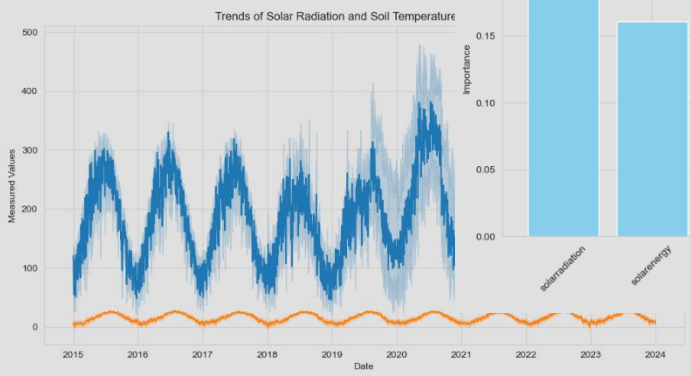
Introduction

Elections are the foundation of any democratic system, and their outcomes have a significant impact on the governance and future trajectory of a nation.

Concerns and issues regarding elections have emerged over the years, including voter suppres security, and money in politics

It's important to understand what factors election results

Factors that can influence election results ir age, food, race, marriage, education, and vc



Analysis & Methodology

- **Dataset Gathered:**

US Census and Election Results (2000-2020) from Kaggle. Fundamental variables include individuals' annual income annual total family income, age, gender, marital status race, citizenship status, language spoken at home education level, and employment status at the individua level.

- **Dataset Cleaned:**

Dataset columns are renamed such that reader will be easier to understand the features. Missing values are dropped due to vast time range, it will not be feasible to replace them.

- **Statistical Models:**

Uncover patterns and relationships within the data, make predictions about various outcomes, and ultimately gain a deeper understanding.

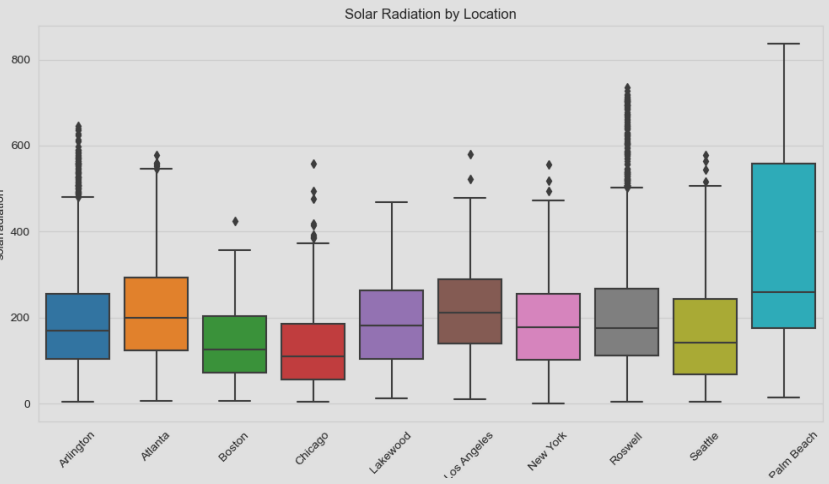
Logistic Regression & LDA & QDA Random Forest

KNN Clustering GBM & SVM

XGBoost LASSO

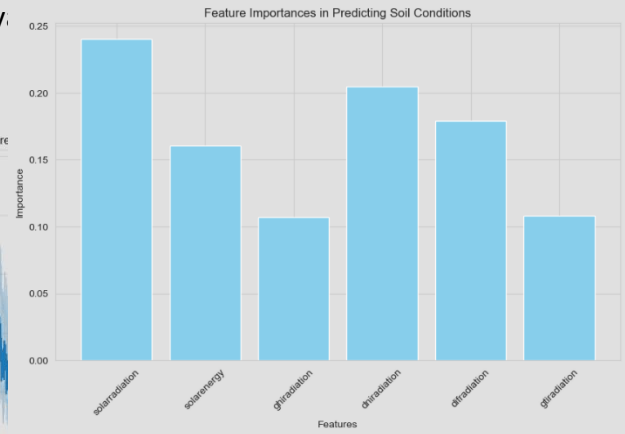
EDA

Some fundamental variables pattern and correlation: We can see that the majority electors have average age from 40 to 50.



Evaluation metrics and Results

Few of the features shows strong correlation to each other. In addition, the response feature is the winner, the categorical v



...

n see that the cluster group 3, and 4 are balanced

le results show that the top 10 feature that contribute

cluster group 3 only contains 9 observations.

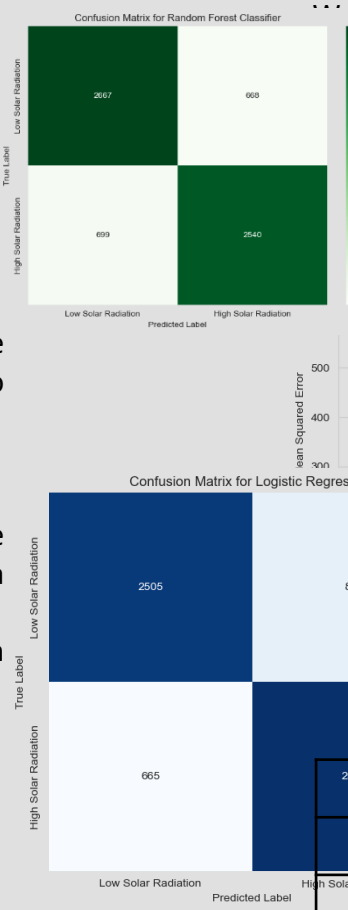
ostly to election results are marriage, race, citizenship,

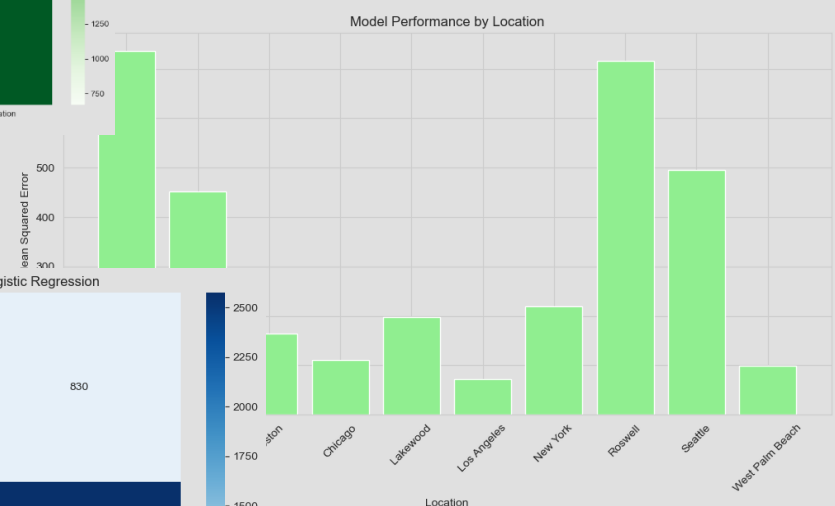
we can see that the cluster one group's average

ortgage payments, doctoral degree, and employment.

ng party are significantly higher than the other three

r group.





Conclusion

- We have found that the key features that affect election outcome are actually "Never Married" & "Race Info".
- We find that both supervised and unsupervised method are giving [illegible] for predicting our target variable.
- There are differences among Republican and Democrat Winners: Education, Age, and some other features
- We have found that our dataset contains many features with strong correlation, and we removed them.
- Random Forest are having a better performance than XGBoost, which could because of our data set does not contains many features after data cleaning. Features such as foodstamps, family

Reference

Nguyen, M. (2021). US Census for Election Predictions 2000-2020. Retrieved April 23, 2023, from https://www.kaggle.com/datasets/manavnguyen/census-for-election-predictions-20002020?select=county_census_and_election_result.csv