



A model for the Spain electricity shortfall

EXPLORE NM4

Outline

- Introduction
- Problem statement
- Data science process
- Model exploration
- Deployment
- Conclusion



Introduction

Energy

- Essential part of our lives today.
- Crucial to every country's economy.
- Produced from either renewable and non-renewable sources.
-



Fossil fuels vs renewable energy

Fossil fuels (39%)

Renewable energy (40%)

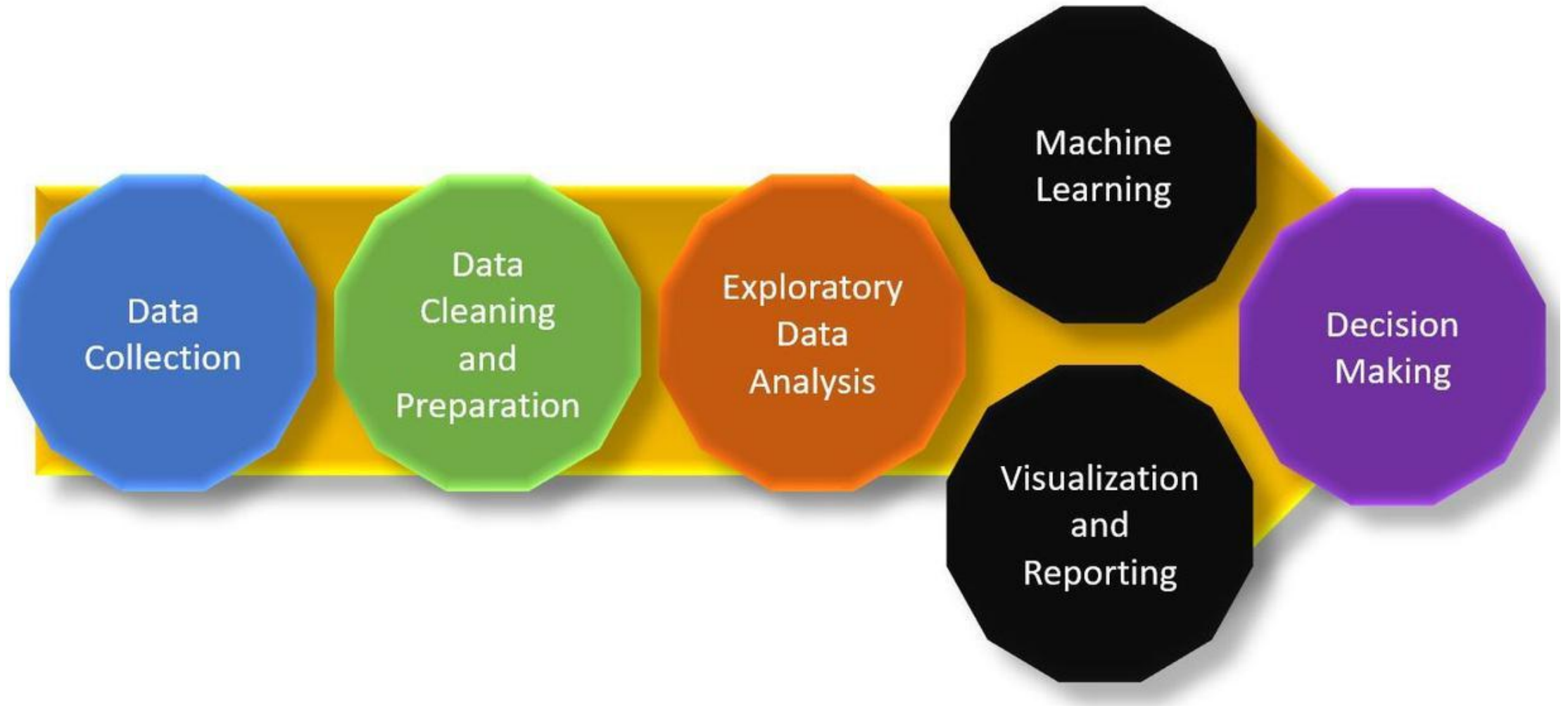


Problem Statement

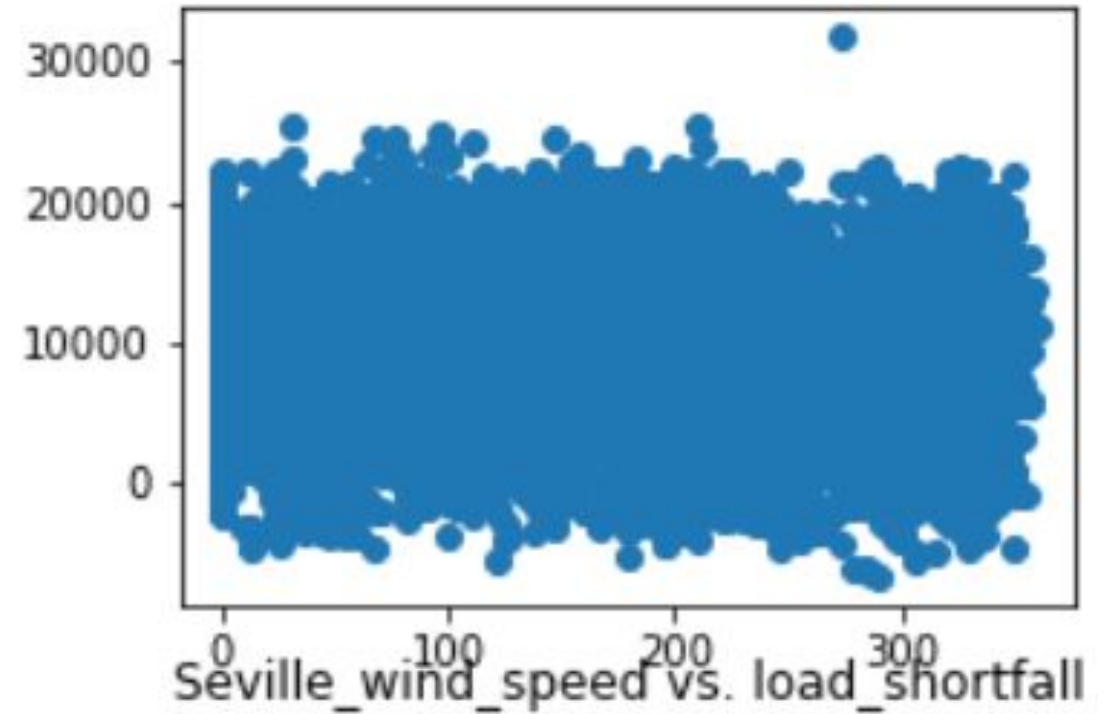
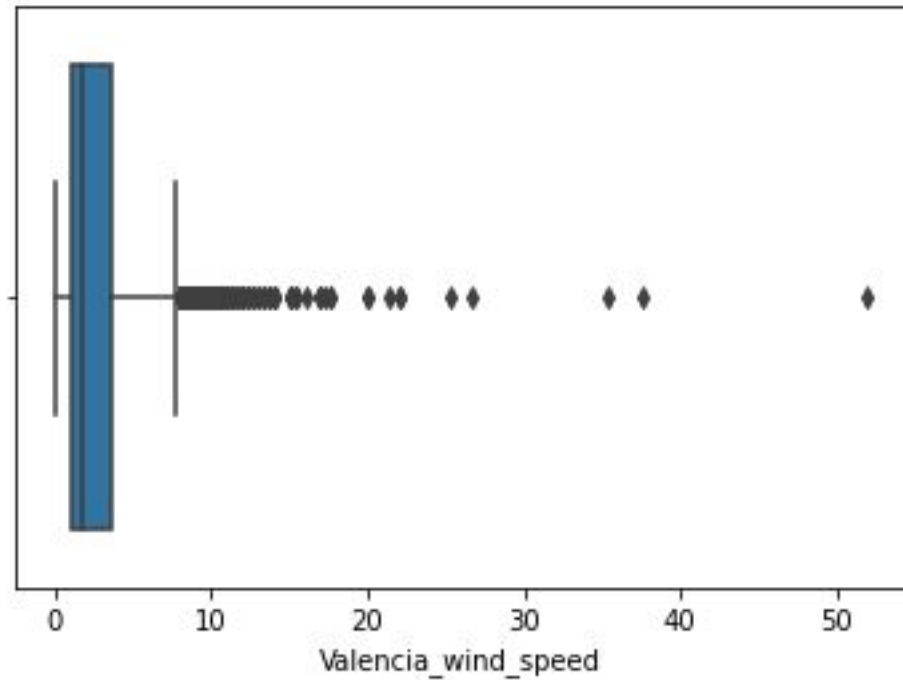
To model the shortfall between energy generated by fossil and renewable energy.



Data science process



Exploratory data analysis



Observations From EDA

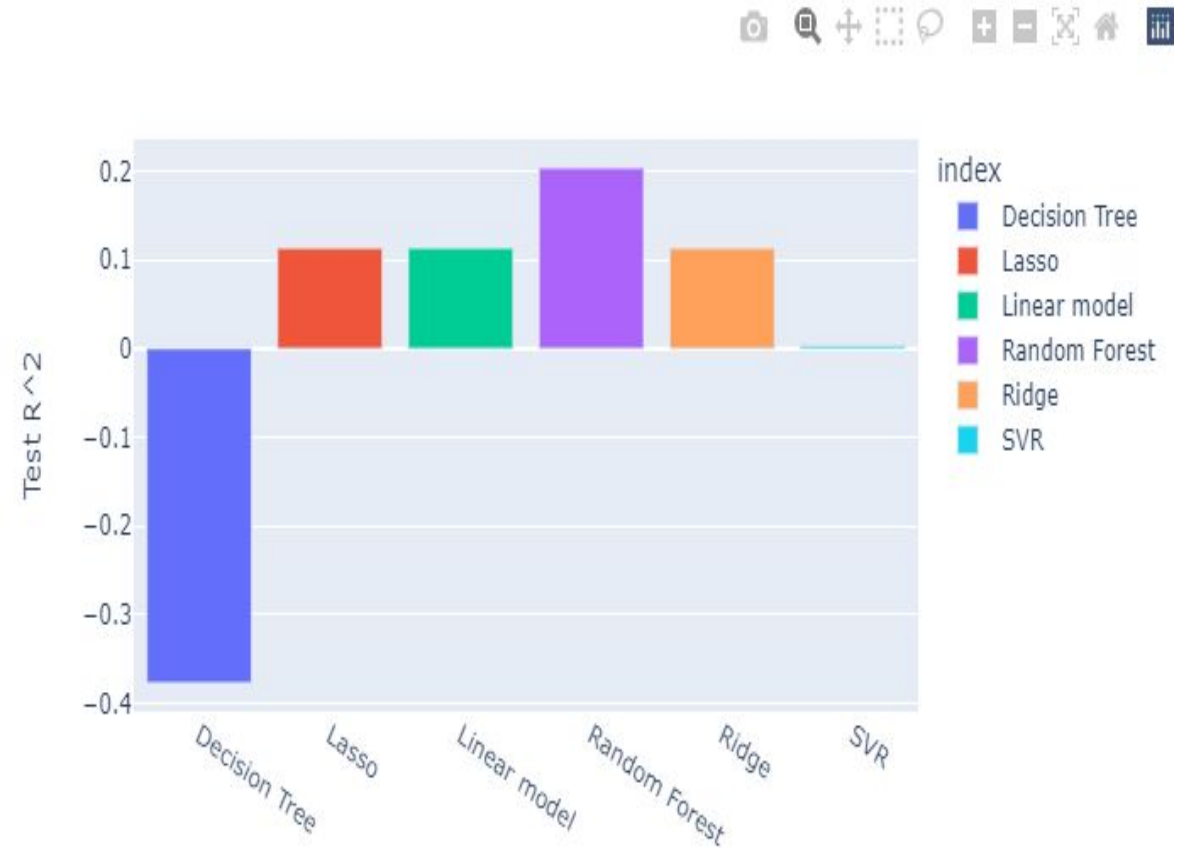
- The EDA revealed missing data that were filled with the mean value and the median
- The skewness and kurtosis of the features revealed that there were strong presence of outliers.
- Features that had no significance to our work like '.Unnamed ' were dropped.
- Categorical data such as Seville_pressure and Valencia_wind_degree were converted to numerical data.

Importance of observations

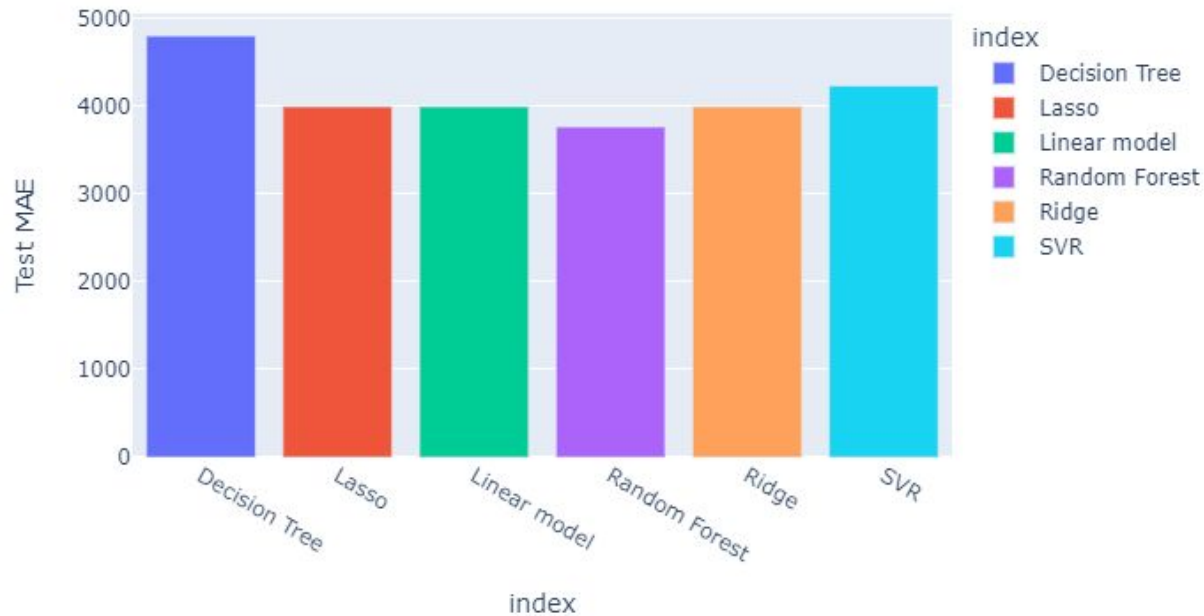
It is vital that some of the anomalies seen in the data (missing value, presence of solid outliers, and others) are engineered so that we can get the best model possible

The model

- Random Forests Algorithm is great with high dimensional data.
- It is faster to train than decision trees because we are working only on a subset of features in this model.
- It can easily work with hundreds of features.



Model evaluation

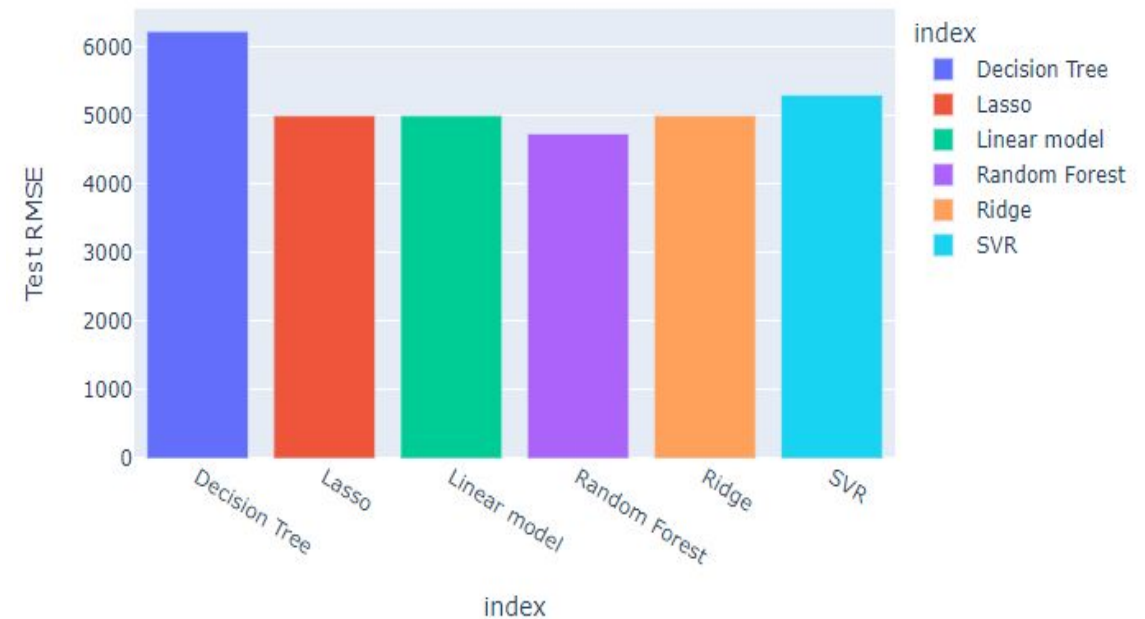


Test MAE

Decision Tree -	4798.474904
Lasso -	3992.059399
Linear model -	3992.055388
Random Forest -	3762.297770
Ridge -	3992.106017
SVR -	4229.349896

Test RMSE

Decision Tree -	6226.572335
Lasso -	4997.095055
Linear model -	4997.092645
Random Forest -	4735.751875
Ridge -	4997.138050
SVR -	5298.470881



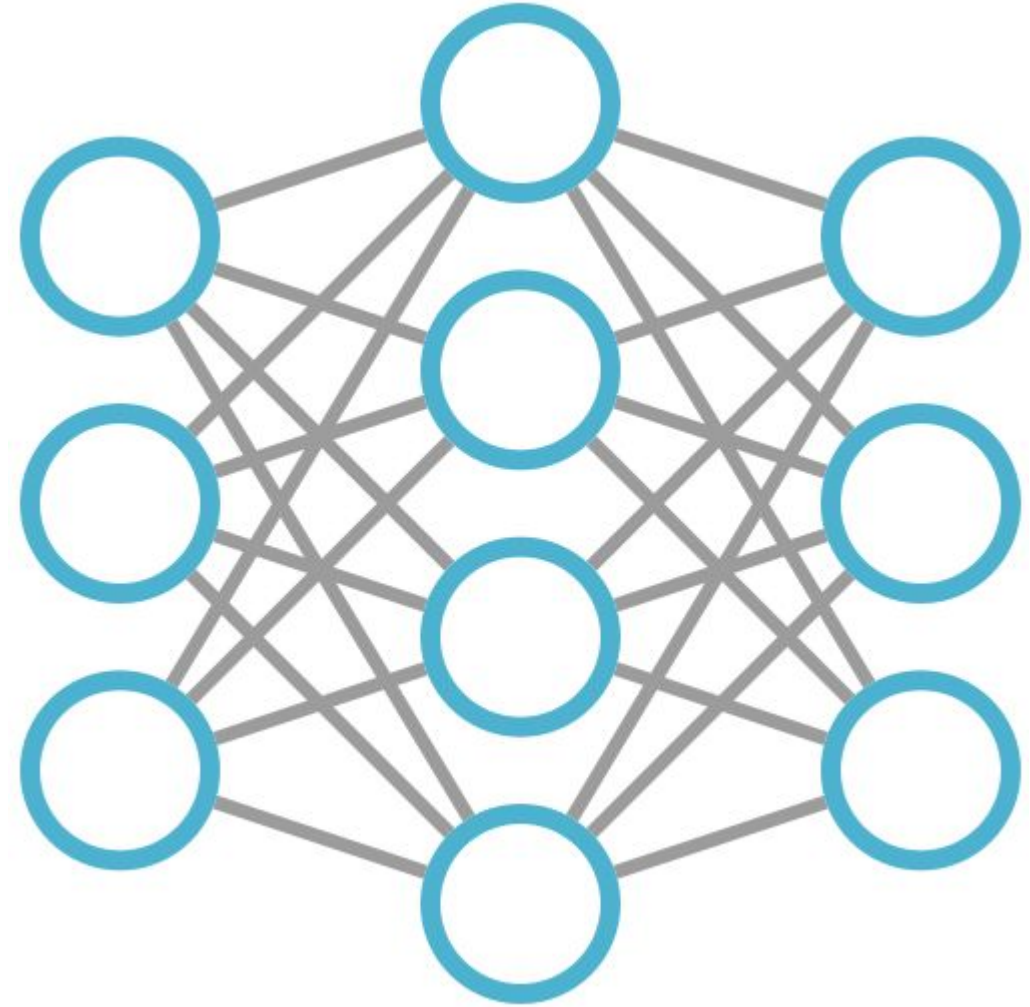
Model deployment

- **What we did?**

We made our model available to other systems within the organization , so that they can receive data and return their predictions

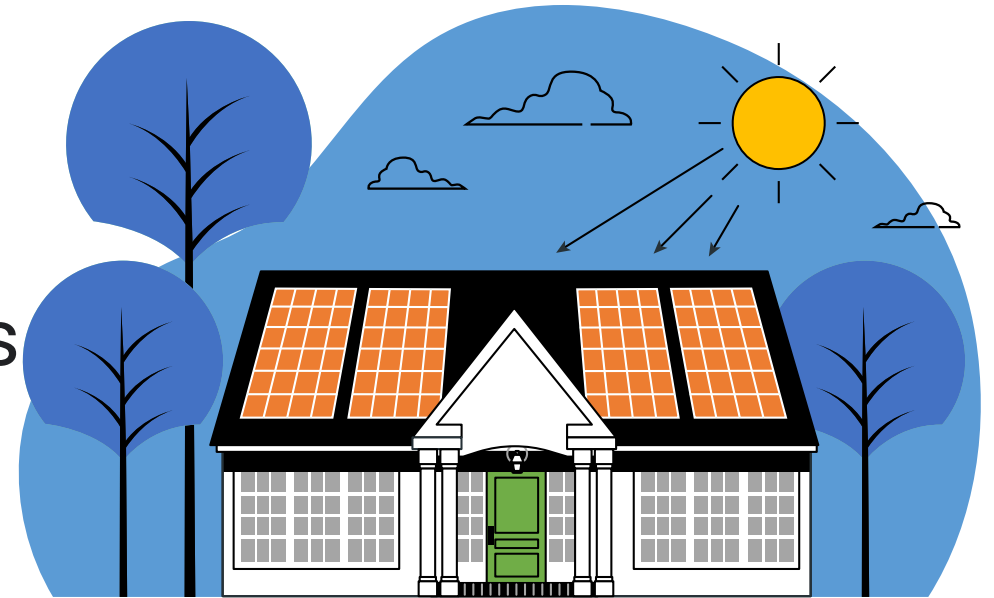
- **Why it's important?**

For practical decision-making.



Conclusion

- Developed the best model that will assist in predicting the 3 hourly load shortfall.
- Informed the Spanish Government of any trends and patterns of the country's renewable resources and fossil fuel energy generation.
- Whether it is viable to expand its' renewable energy resource infrastructure investments.



FUTURE WORK AND RECOMMENDATION

- Improve API to accept arbitrary data
- Evaluate other models like XGBoost for better result.

Thapelo
Mofokeng

Ayoola
Solanke

Humphery
Ojo

Samson
Oguntuwase

Theresah
Koomson

Abiola
Akinwale

