SPAIN

ELECTRICITY

SHORTFALL

WHAT WE DO…

- We are a consulting company and we specialize in providing strategic business solutions.

- We provide business insights through data analytics and machine learning models.

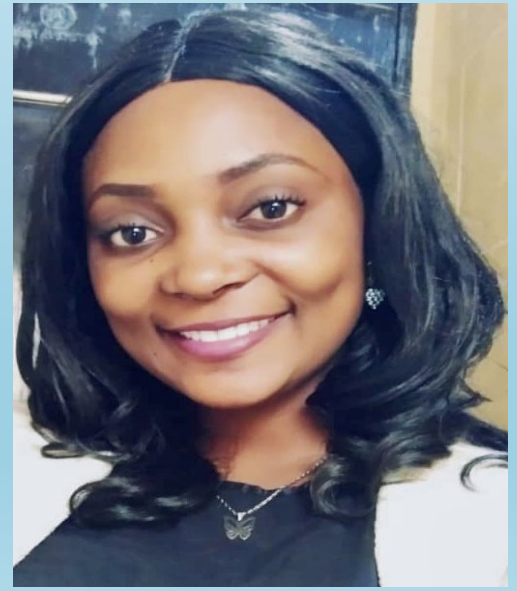
MEET THE TEAM…

Team Lead
Mr. Oladotun Jonibola

Ms. Diana Okeyo

Mr. Raphael Mbonu

Mr. Mijan Amos



Ms. Gloria Thompson

PROBLEM STATEMENT

- The population of Spain has been increasing creating a shortfall in fossil fuels and renewable energy consumption.

- The government is looking into investing in a renewable energy infrastructure.

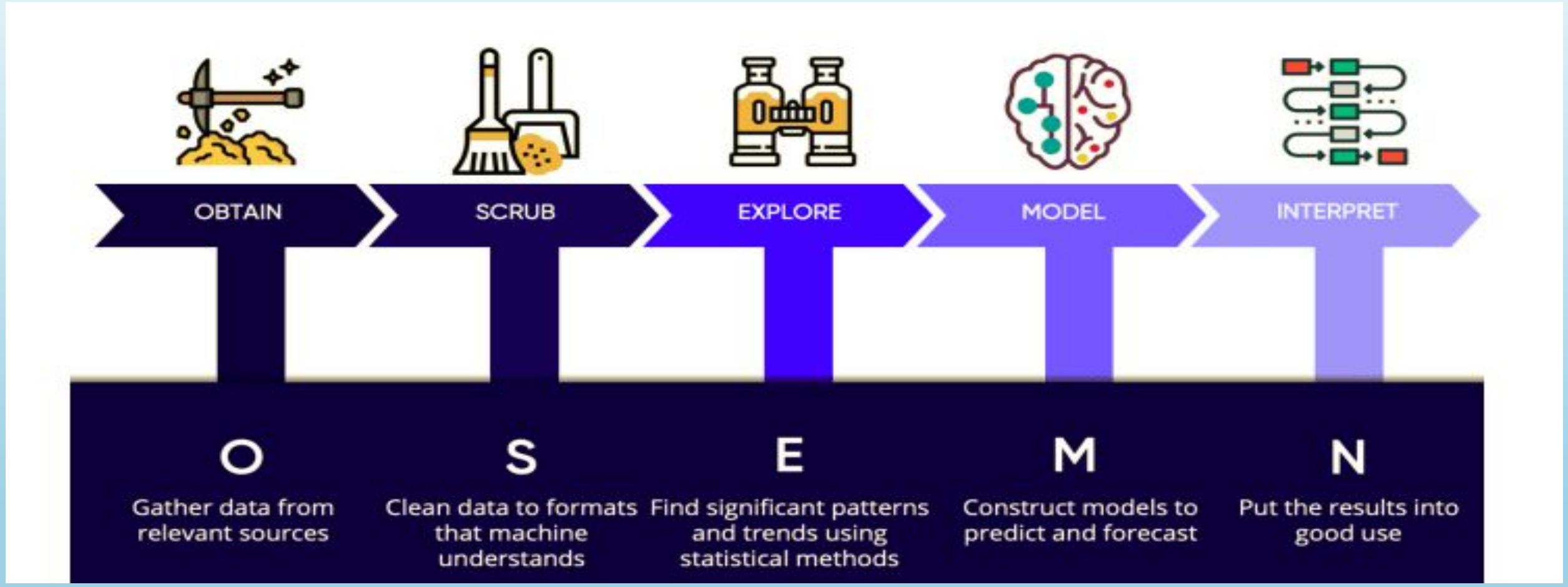
OUR MISSION

- Build an API to predict energy shortfall to help the government of Spain infrastructure investment.

OVERVIEW

1. EXPLORATORY DATA ANALYSIS
2. DATA ENGINEERING
3. MODEL BUILDING
4. MODEL DEPLOYMENT

INTRODUCTION



EXPLORATORY DATA ANALYSIS (EDA)

EXPLORATORY DATA ANALYSIS (EDA)

A total of 8763 rows and 49 features.

There are three categorical variables, denoted as object dtype with remaining 46 features being numeric.

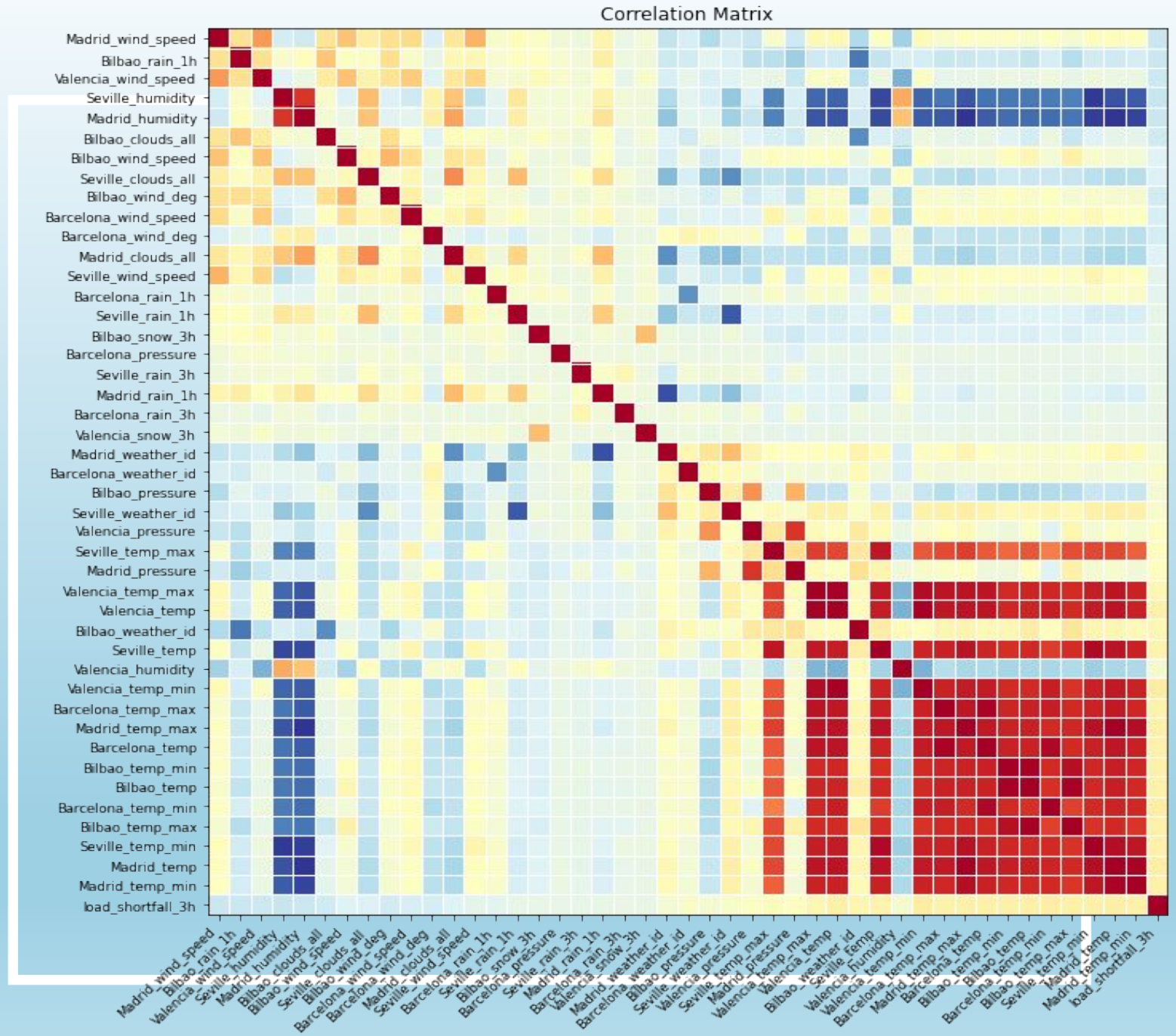
Some values were no recorded/empty. Specifically a column (Valencia Pressure) had 2068 empty cells.

Time column as well as other categorical variables need to be changed to the right format.

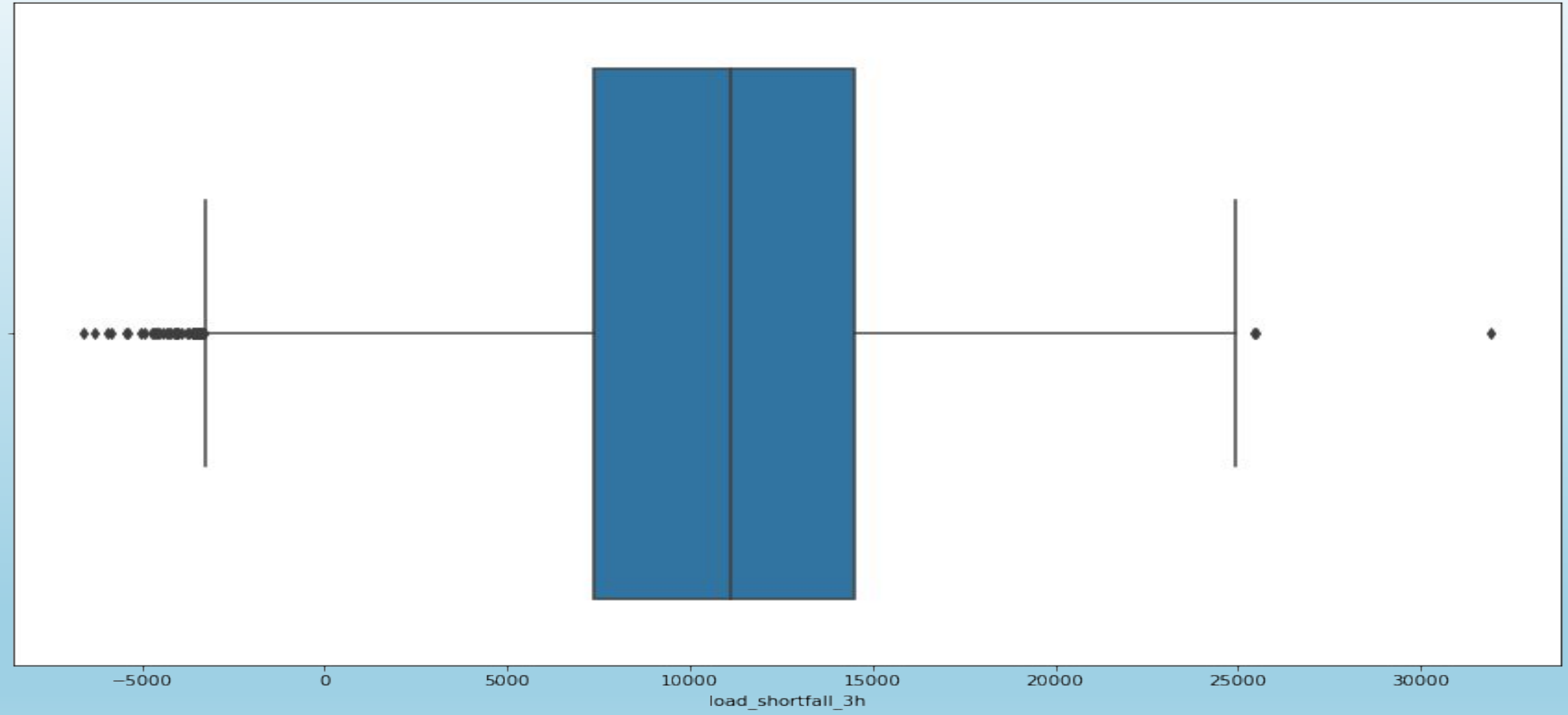
Some features that have zero values shows error during recording or probably that event did not occur at the time of reporting.

FEATURES USED IN PREDICTING THE LOAD_SHORTFALL_3H IN SPAIN

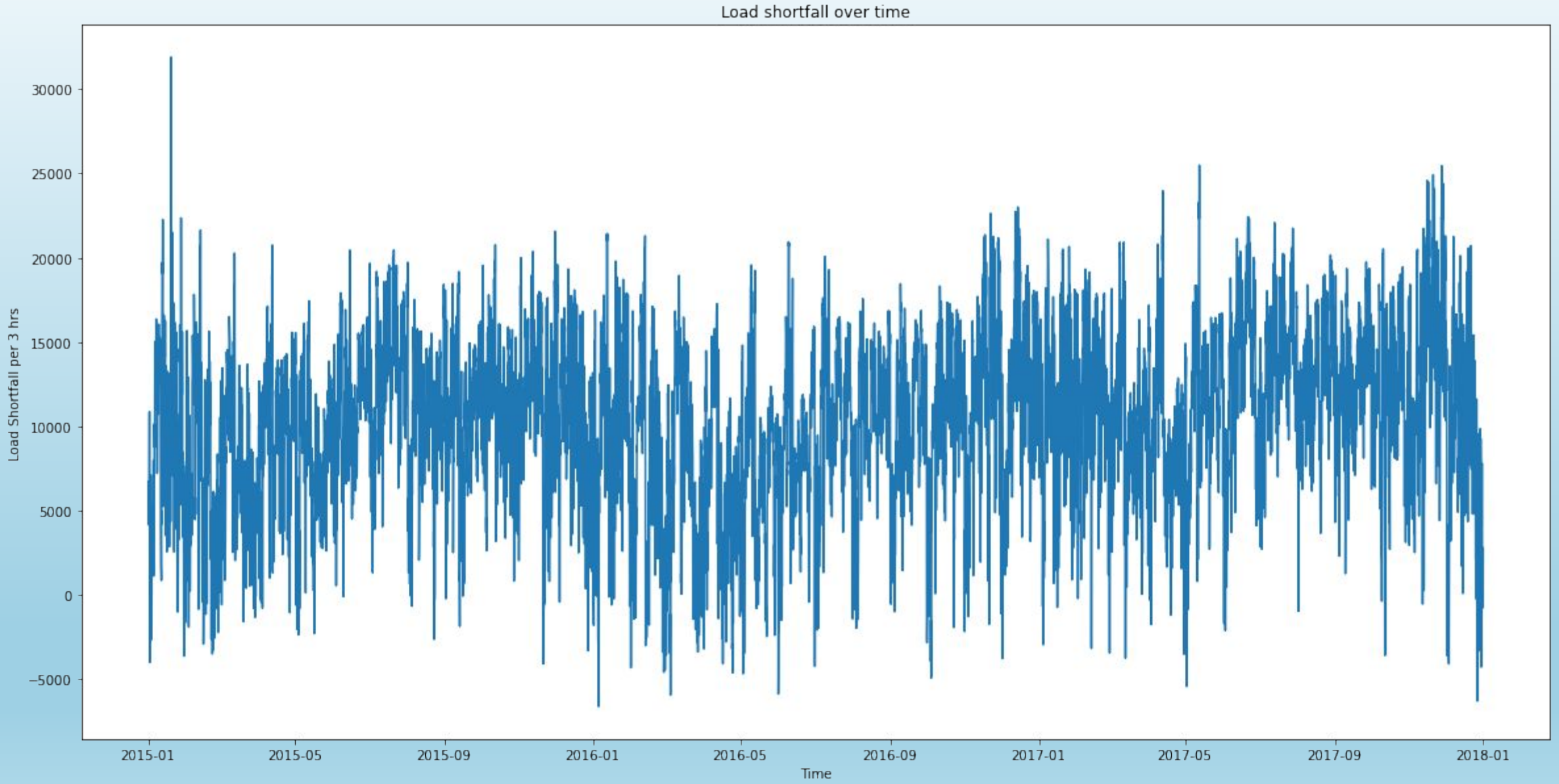
- The bolder region in the image from the non diagonal parts depicts features that are closely related to one another



LOAD SHORTFALL DATA DISTRIBUTION



LOAD SHORTFALL AGAINST TIME



FEATURE ENGINEERING

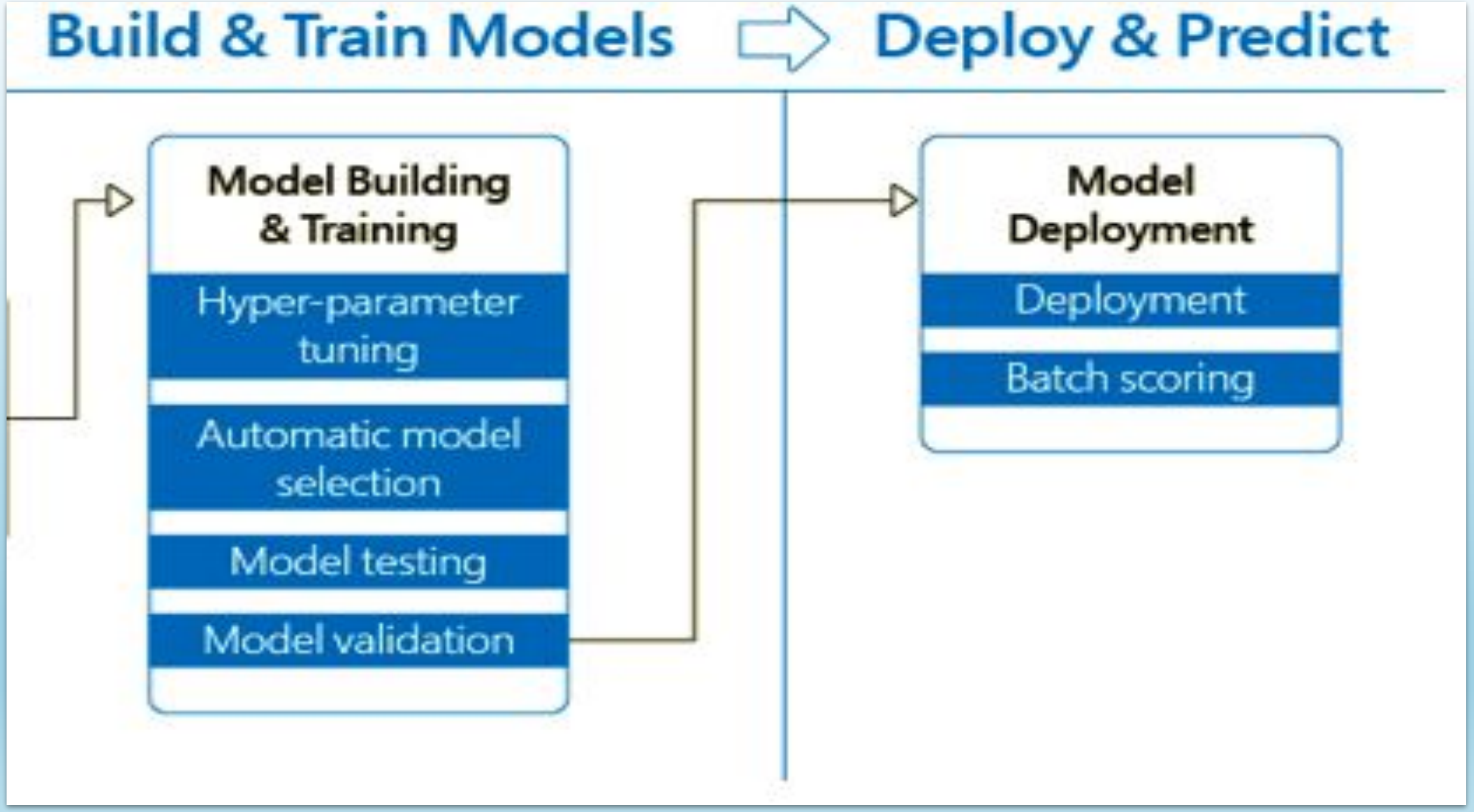
Our favorite heuristics for feature engineering. These can boost model performance like crazy.

DATA SCIENCE PRIMER

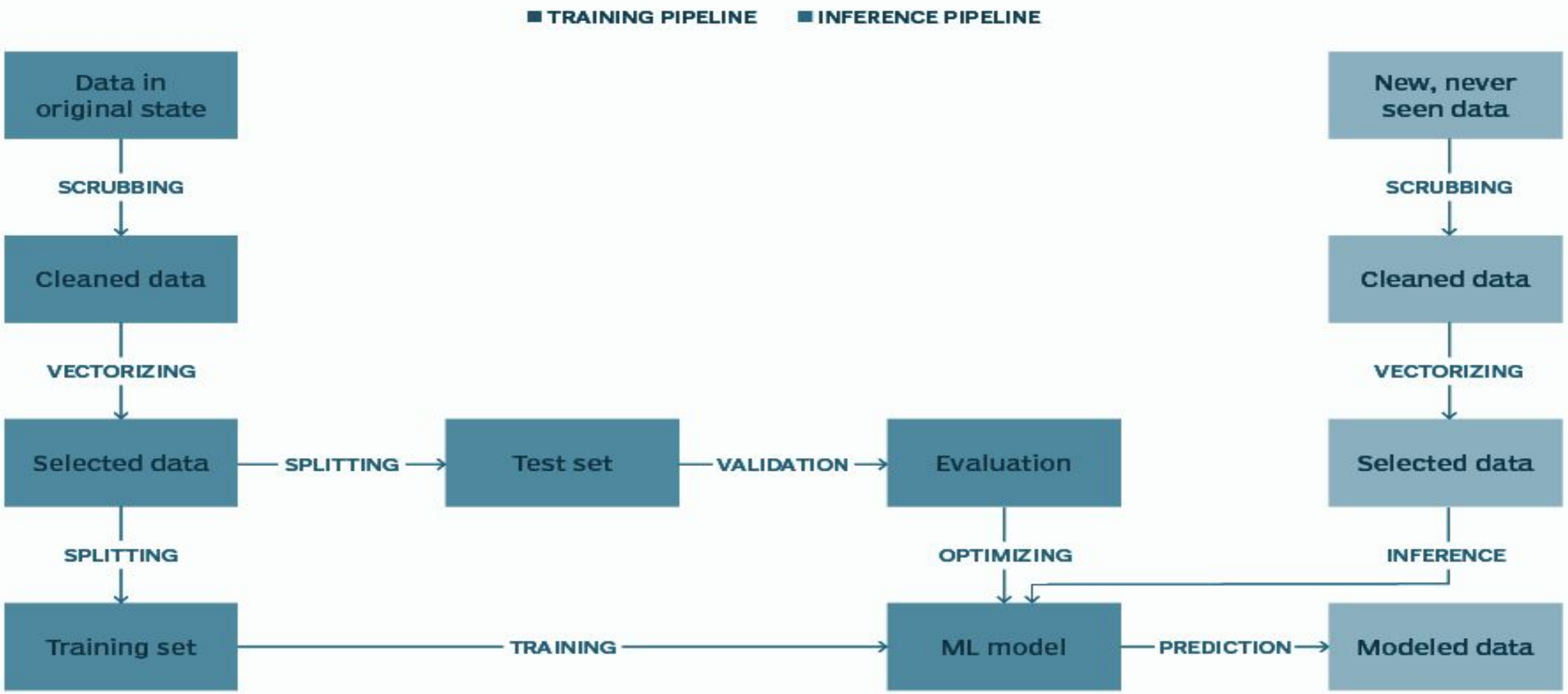
by EliteDataScience

PROCESSES INVOLVED

- The dataset had missing values
- Replaced the missing values with **mode** for categorical and **mean** for numerical data
- ML models built predicted the load_shortfall against time


Build & Train Models
Deploy & Predict
Model Building & Training
Hyper-parameter tuning
Automatic model selection
Model testing
Model validation
Model Deployment
Deployment
Batch scoring

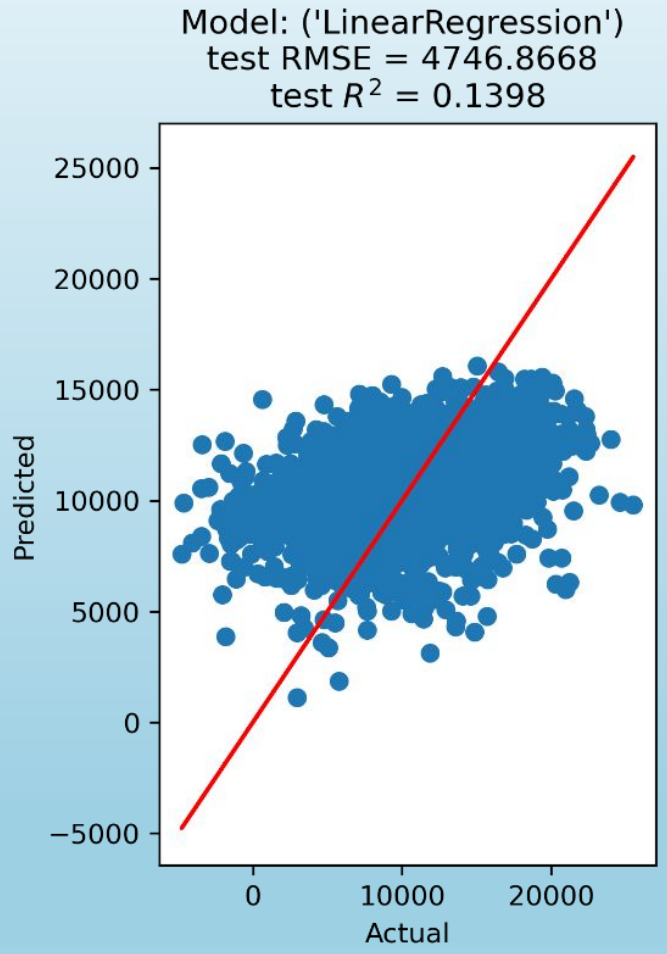
DATA MODELLING

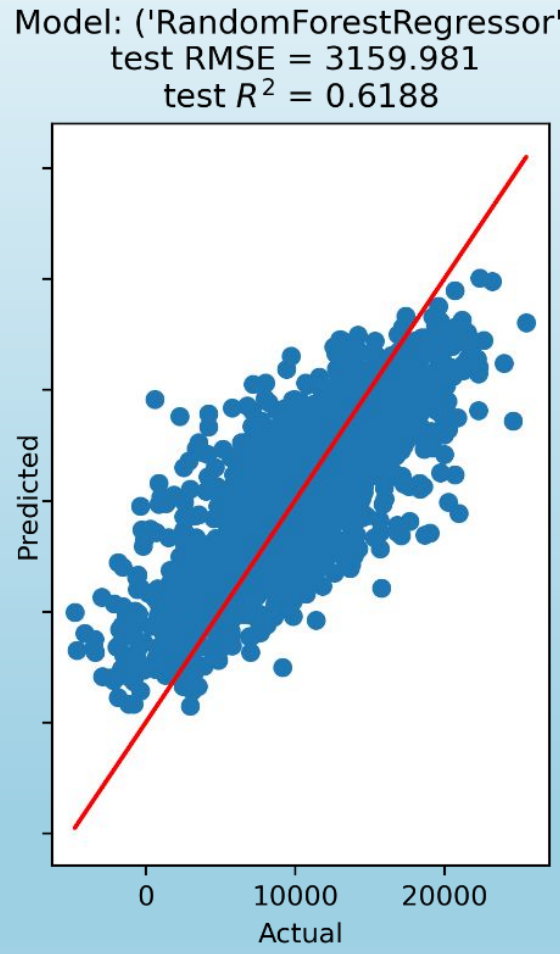


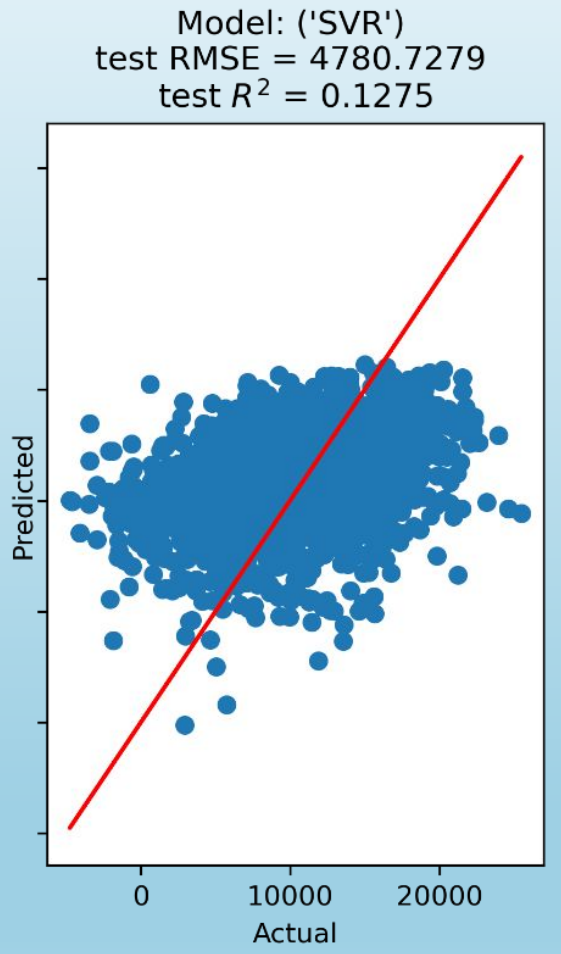
DATA MODELLING

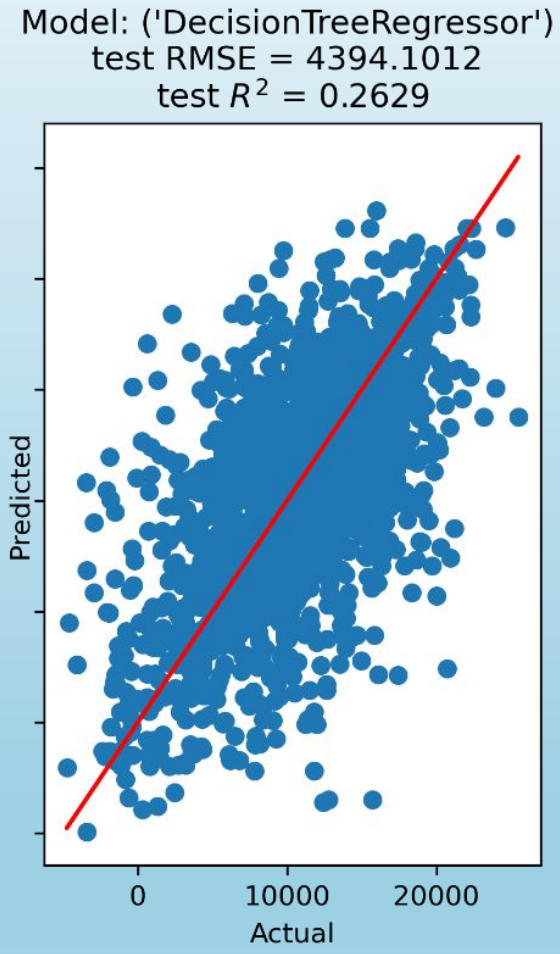
- Splitting of features into Train-Test set
- Model prediction
- Model performance

MODELS' RMSE PERFORMANCE

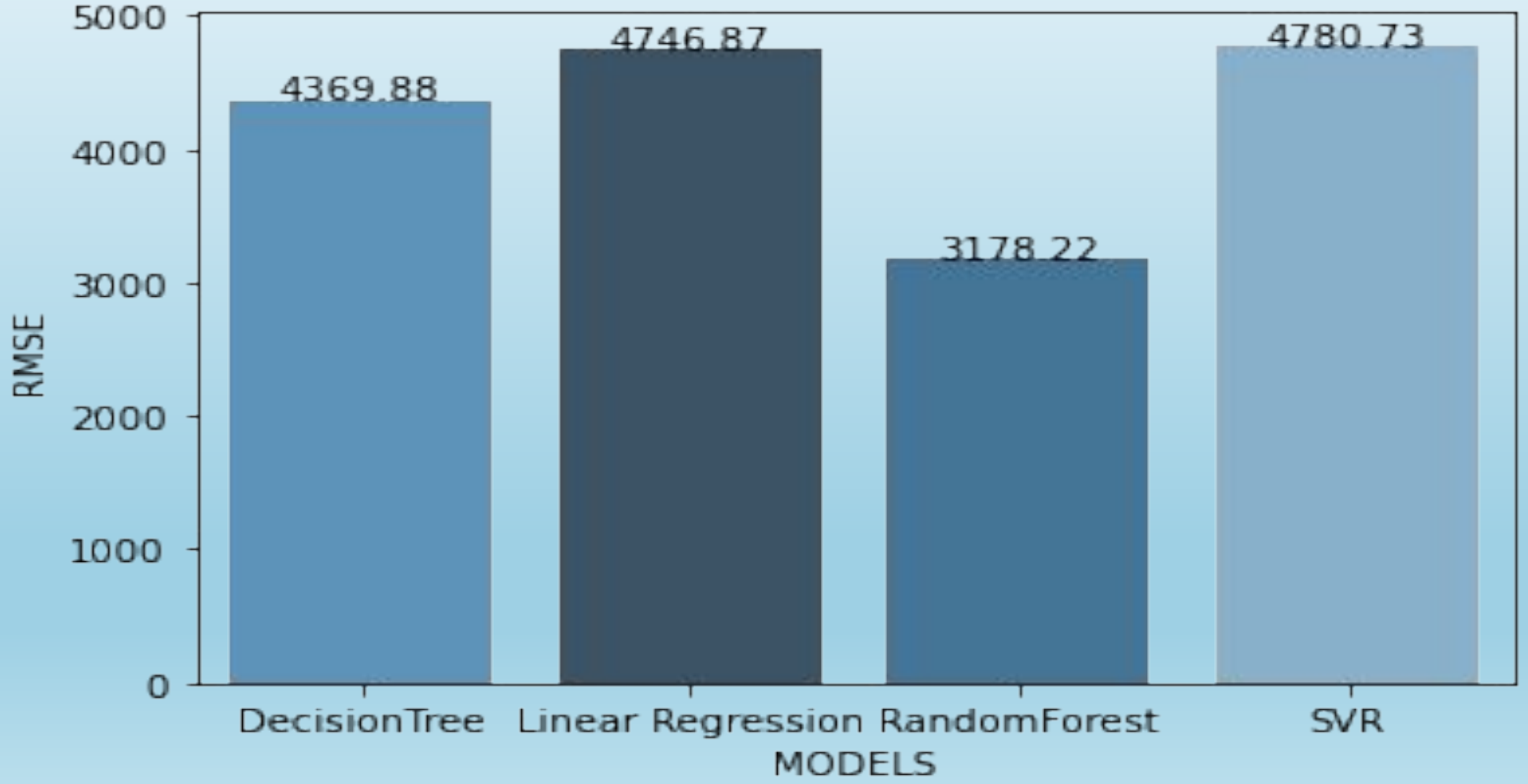








MODELS' RMSE PERFORMANCE



CONCLUSION

- Out of our 4 models we picked Random Forest because of the low RMSE.
- Better renewable energy infrastructure.
- Better investment decisions.

THANK YOU