FOOD DEMAND FORECASTING FOR DISTRIBUTION SERVICE

L.Q Huy, L.Q.H Thinh, D.T Nhi, T.H.N Vu

Group 9, BigData Club, Ho Chi Minh city University of Technology, Vietnam National University - Ho Chi Minh city

INTRODUCTION

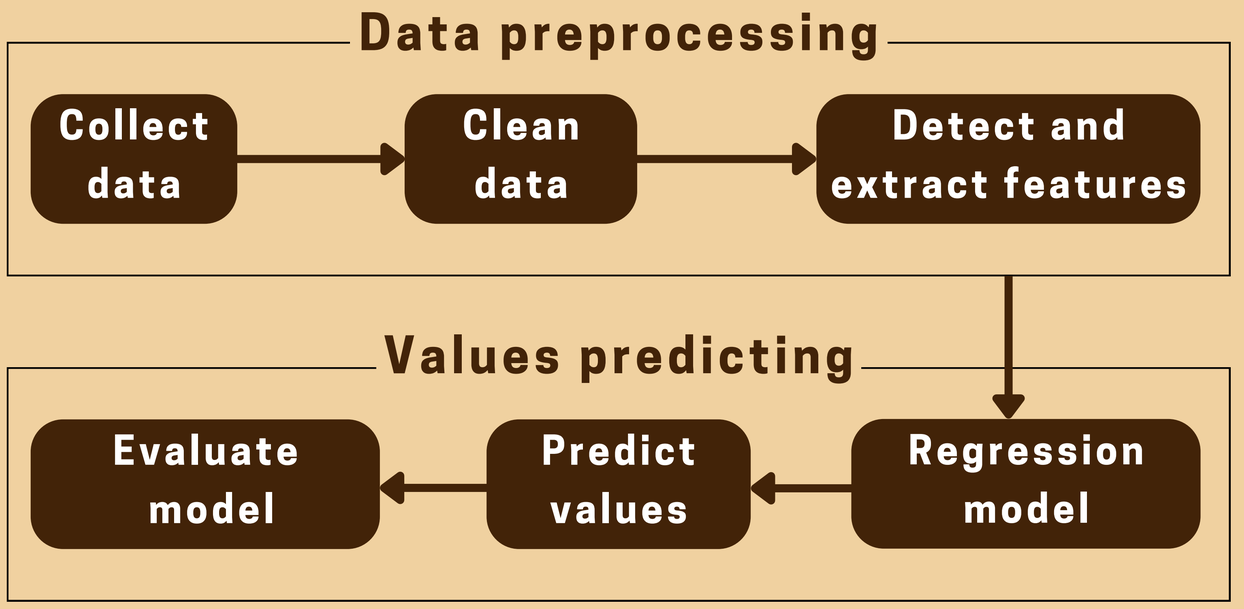
- Food is one of the most vital human's daily demand beside air and water.
- One branch of service industry involved in food is food distribution service.
- The problem of how and how much food should be prepared for distribution is extremely important in this business.

OBJECTIVES

- Build a model to predict the food demand from collected data.
- Understand the relationship between features and how they influence the predicted values.
- Revise acquired knowledge and practice applying it to practical products, also learn how to teamwork efficiently.

METHODS

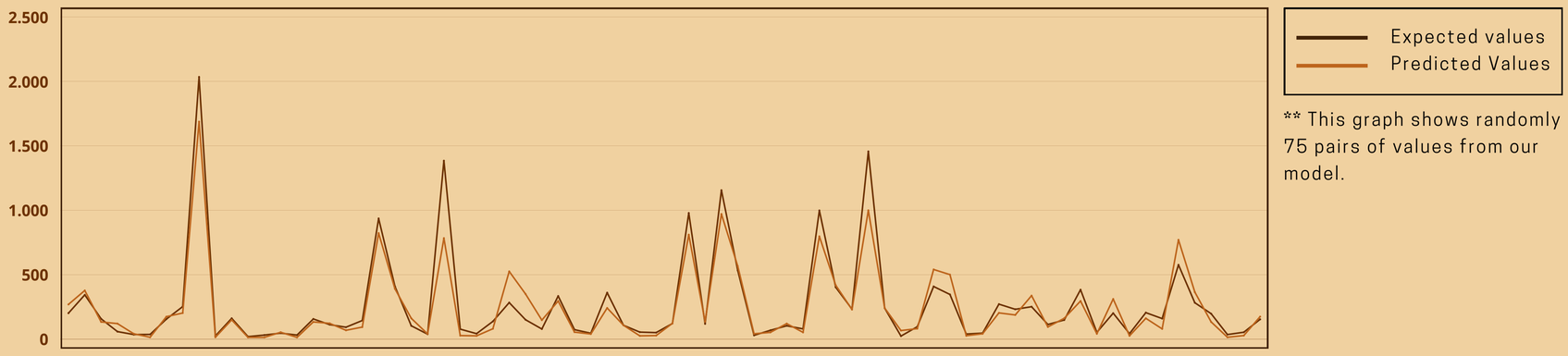
- Collect data from particular source, in this assignment we collect them from Kaggle.
- Clean data, use specified methods to detect and extract the most significant features from collected dataset.
- Use regression model on two preprocessed splitted datasets, predict the expected values and then evaluate model's performance by these values.



RESULTS

PREDICTED AND EXPECTED VALUES FROM TEST DATA

There is not so much difference between predicted and expected values.



** This graph shows randomly 75 pairs of values from our model.

Figure 1: Graph shows relation between expected and predicted values, which we obtain from our model.

RMSE AND DIFFERENCE IN VALUES IN TEST DATA

The root mean squared error of this model is:

RMSE = 344.986

** This RMSE value is obtained from model modifying and validating process, in this case we use RandomForestRegressor for solution.

The difference in values

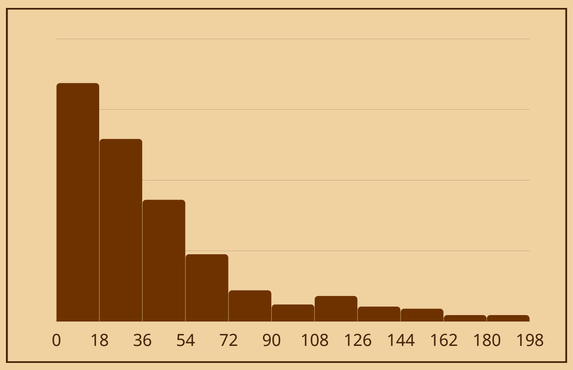


Figure 2: Histogram shows the percentage of difference between expected and predicted values as considered with expected values in test data, can see that the main part is below 100.

CONCLUSIONS

- The model can predict the trend of expected values.
- The model can also basically predict the expected values based on extracted features.
- The model has not-so-good performance, showed in high RMSE and percentage of difference between predicted and expected values.
- This poor performance may have many causes, the most probable is from collected data.
- Our model can be used as a reference for shops in food distribution service such as Bach Hoa Xanh, ... help them predict the trend of food demand.