**Detailed Project Report**

Thyroid Disease Detection

|  |  |
| --- | --- |
| Written By | AFRIN R PESHIMAM |
| Version | 1.0 |
| Date | 08-02-2022 |

**Document Change Control Record**

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Author** | **Comments** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Reviews

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Reviewer** | **Comments** |
|  |  |  |  |

### Approval Status:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Version** | **Review date** | **Reviewed By** | **Approved By** | **Comments** |
|  |  |  |  |  |

# **Introduction**

Inventory Management a very common problem. Sales forecasting helps you attain this revenue efficiency by offering insight into the likely behavior of your most valuable customers. You can predict future sales, as well as improve pricing, advertising, and product development. Machine learning algorithms can help to predict future sales. This study demonstrates the how different classification algorithms can forecasts sales. Different classification algorithms such as Logistic regression, Random Forest, Decision Tree, Naïve Bayes, Support Vector Machine, XGBOOST have been tested and compared to predict the better outcome of the model.

1. **Objective**

The main goal of this project is to predict the sales based on various products and the store’s factors. It will play a decisive role in order to predict, accurate sales of each product and each stores.

**3. Architecture**

Following workflow was followed during the entire project.

**Dataset**

|  |  |
| --- | --- |
| **Data** | **Description** |
| Item Identifier | Unique product ID |
| Item Weight | Weight of product |
| Item Fat | Low Fat or Regular |
| Item Visibility | The % of total display area of all products |
| Item Type | The category to which the product belongs |
| Item MRP | Maximum Retail Price |
| Outlet Identifier | Unique Store ID |
| Outlet Establishment Year | Year store was created |
| Outlet Size | The size of store in term of ground area. |
| Outlet Location | Type of City store is located |
| Outlet Type | Whether outlet is grocery or supermarket |
| Item Outlet Sales | Sales of Product |

**Data Analysis Steps**

DATA

COLLECTION

DATA

PREPROCESSING

EXPLORATORY DATA

ANALYSIS

FEATURE

SELECTION

MODEL

CREATION AND

EVALUATION

In step 1, we collect

data which is generally

present in a database

or on internet.

In step 2, we

preprocess the data

which involves data

cleaning by handling

outliers, null values

etc.

In step 3, we explore

the data by performing

univariate and

bivariate analysis on

the features.

In step 4, we use

feature selection

techniques to filter out

the most important

features to perform

model creation

In step 5, we finally

build models on our

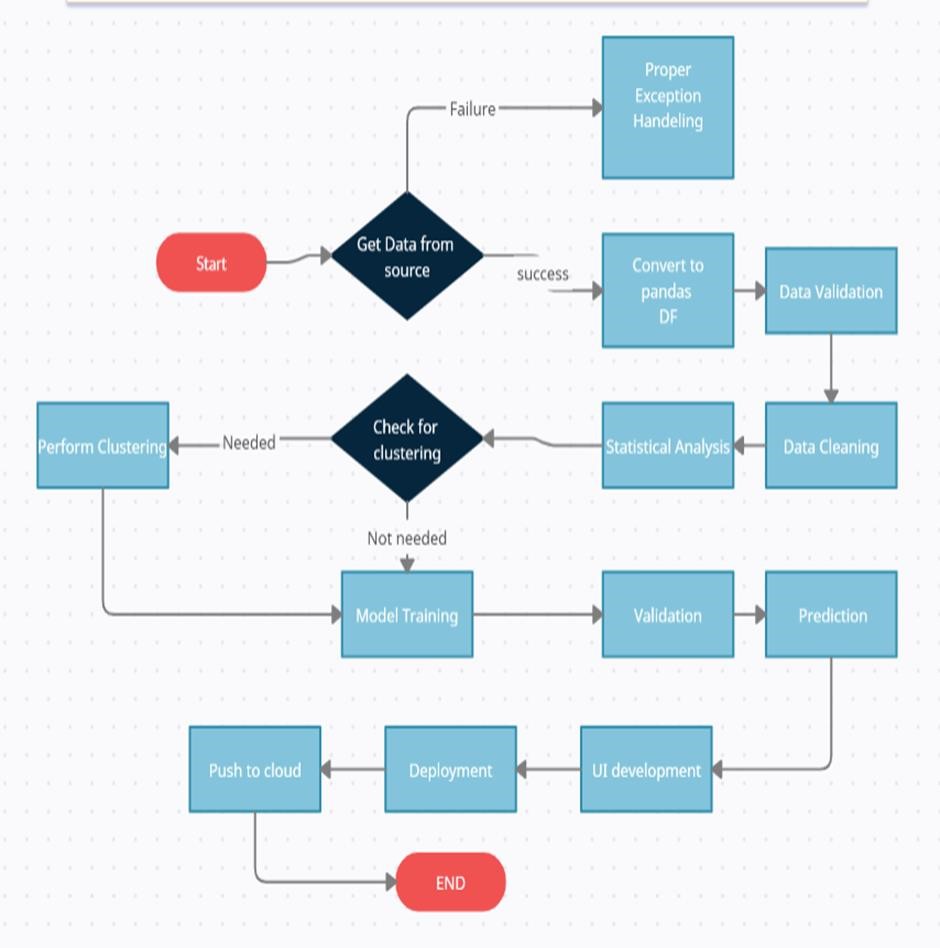
dataset and choose

the model which gives

the best accuracy.

**XGBOOST REGRESSOR MODEL**

**MODEL TRAINING AND VALIDATION WORKFLOW**



**MODEL TRAINING AND VALIDATION WORKFLOW**

**Data Collection**

* BIG MART SALES DATASET from KAGGLE
* For Data Set: [BigMart Sales Data | Kaggle](https://www.kaggle.com/brijbhushannanda1979/bigmart-sales-data)

**Data Pre-processing**

* Missing values handling by analyzing similar data
* Categorical features handling by ordinal encoding

**MODEL TRAINING AND VALIDATION WORKFLOW**

**Model Creation and Evaluation**

❑ Various classification algorithms like Logistic Regression, Random Forest, Decision Tree, Naïve Bayes, Support Vector Machine tested.

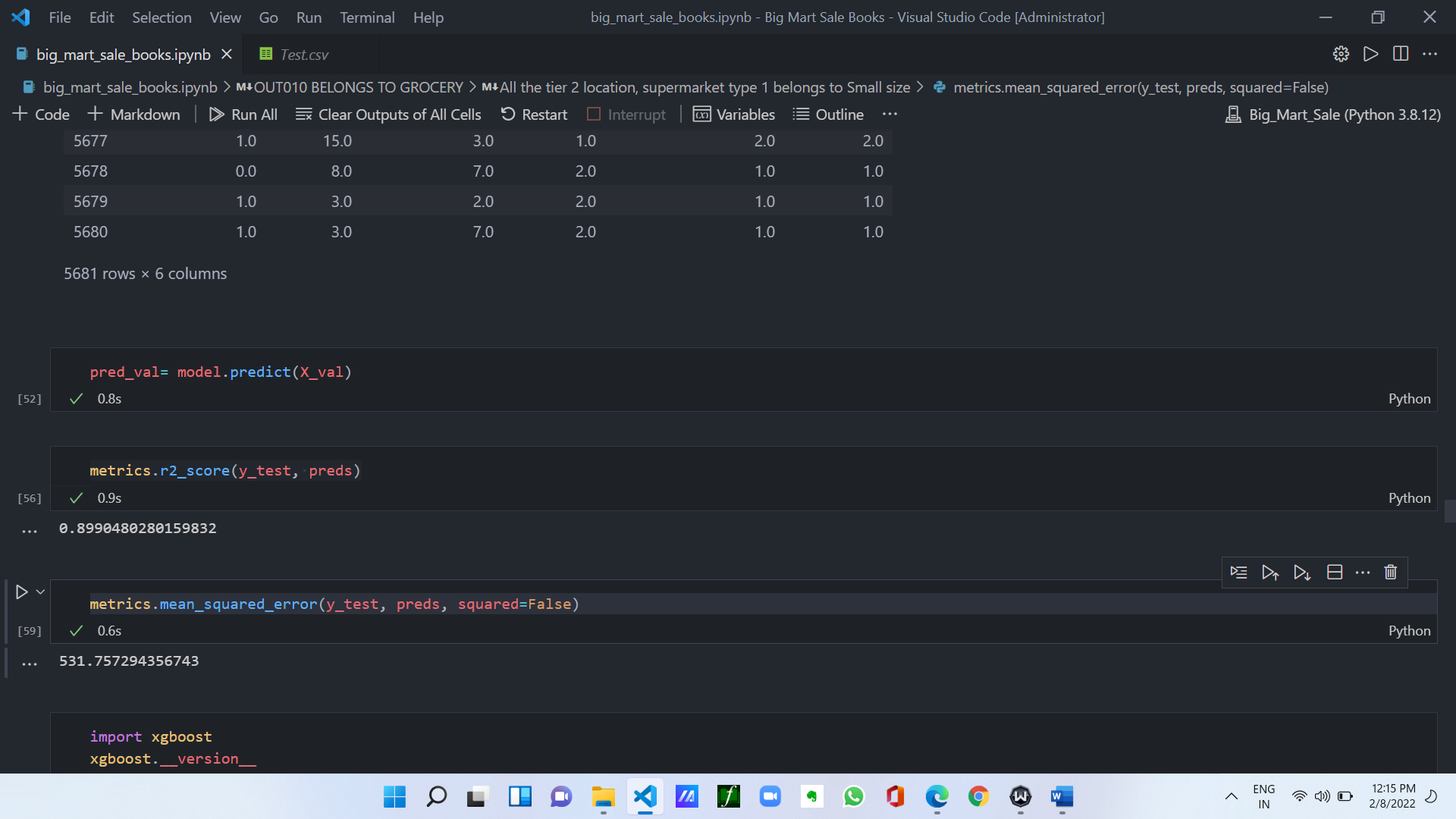
❑ Random Forest, Decision Tree, XGBOOST and Logistic regression were given better results. XGBOOST was chosen for the final model training and testing.

❑ Hyper parameter tuning was performed.

❑ Model performance evaluated based on RMSE and R2

**Model Prediction Results**

**on Test Dataset**

****

**Model Deployment**

**Model Deployment**

*  The final model is deployed on Heroku using Flask framework.

**Frequently Asked Questions**

**Thank You**