Zindi Store Sales Forecasting Challenge – 8th Place Solution by Mawero Rodney G.

1.1 Overview

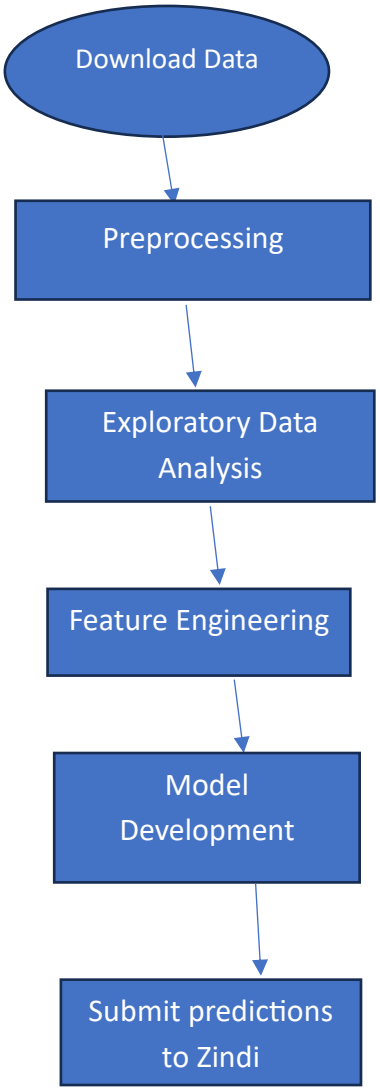
Sales forecasting is the foundation of a business's financial story. Once you have your sales forecast you can create profit and loss statements, cash flow statements and balance sheets, thus helping you set goals for your company. Proper forecasting also ensures you have the right stock at all times and leads to less wasted stock.

1.2 Objectives

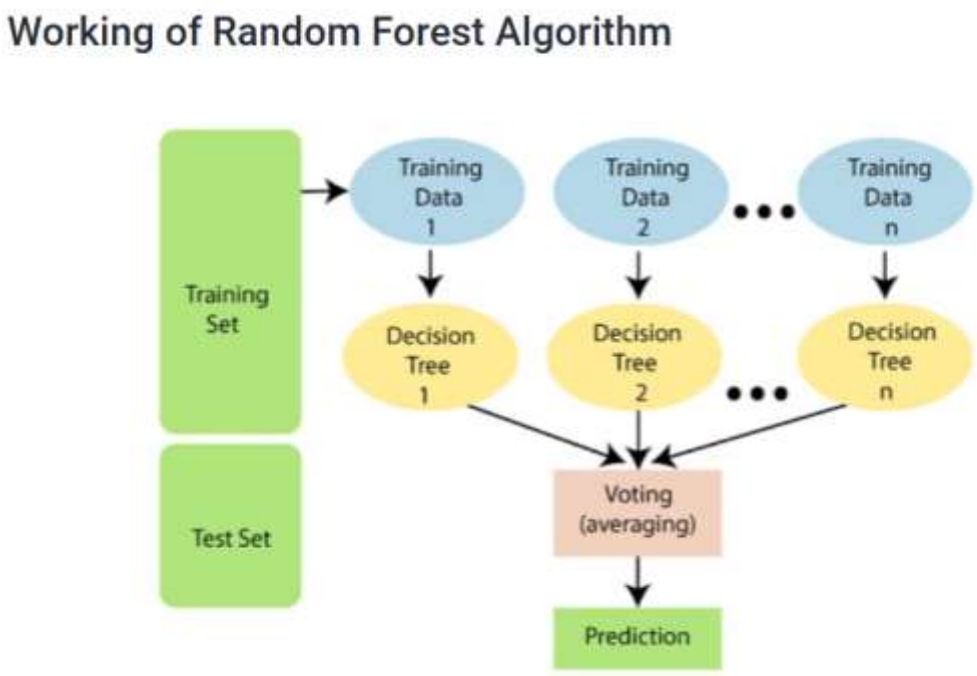
The objective of this challenge is to create a model to forecast the number of products purchased per week per store over the next eight weeks, for grocery stores located in different areas in the same country. The solution to this challenge can be used by small chain stores to know how much stock to order per week and per month.

The final model submitted is a random forest regressor.

2.1 Machine Learning Pipeline



2.1 Architecture Diagram



Source: Simplilearn https://www.simplilearn.com/tutorials/machine-learning-tutorial/random-forest-algorithm

The following steps explain the working Random Forest Algorithm:

Step 1: random samples from a given data or training set are selected

Step 2: This algorithm will construct a decision tree for every training data.

Step 3: Voting will take place by averaging the decision trees.

Step 4: Finally, selection of the most voted prediction result as the final prediction result. This combination of multiple models is called Ensemble. The ensemble method used in random forest is **bagging**

Bagging: Creating **a different training subset** from sample training data **with replacement is called Bagging.** The final output is based **on majority voting**.

3.0 ETL – Extraction Transformation and Loading of Data

Data

Description	Files
Contains the target. This is the dataset used to train the model	**Train.csv**
Resembles Train.csv but without the target-related columns. This is the dataset used to apply the trained model.	**test.csv**
Information about holidays	**holidays.csv**
Information about the different stores such as their locations	**stores.csv**
	dates.csv

2.6 MB

Information about holidays

holidays.csv

1.8 KB

stores.csv

864 B

Information about the time periods with their associated date features e.g. day of the week, day of the year, etc

dates.csv

79.8 KB

Shows the submission format for this competition, with the ‘ID’ column mirroring that of Test.csv. The order of the rows does not matter, but the names of the ‘ID’ must be correct.

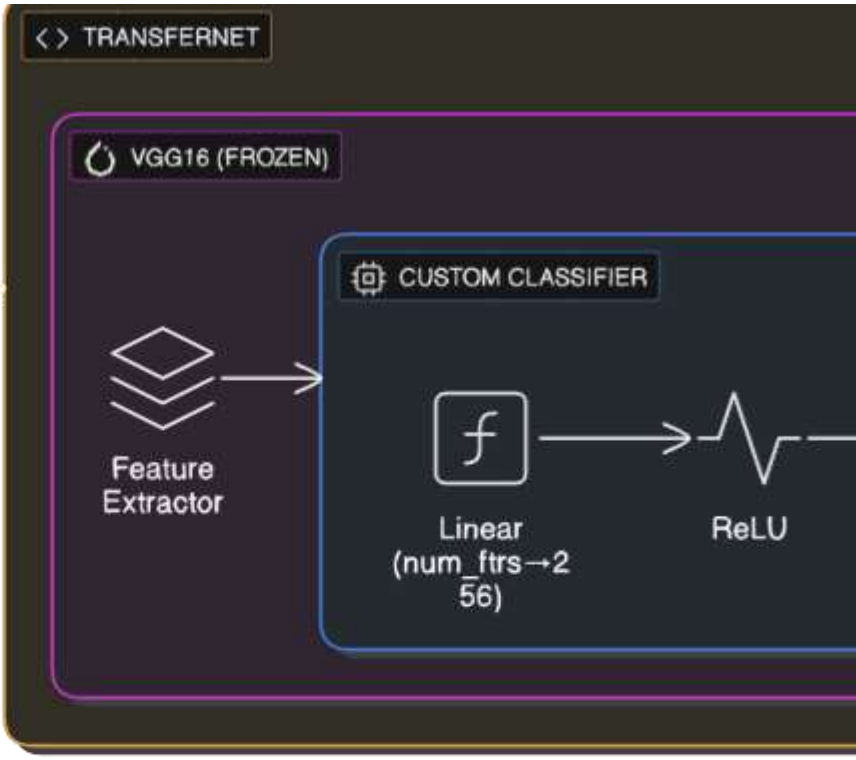
SampleSubmission.csv

508.6 KB

Contains the target. This is the dataset that you will use to train your model.

train.csv

83.6 MB


TRANSFERNET
VGG16 (FROZEN)
CUSTOM CLASSIFIER
Feature Extractor
Linear (num_ftrs→256)
ReLU