We have performed Exploratory Data Analysis (EDA) & feature engineering on our dataset after merging all the given dataset. We used a ‘RandomForestRegressor()’ ensemble algorithm for prediction where we used k-fold ross validation with value of k as 10 and splitted the dataset into 75% for training set & 25% for test set. We used a loop to train K models with a 75/25% random split of the data each time between training and test samples. We scaled the data & computed accuracy, using mean absolute error (MAE) which is almost exactly the same after each iteration. It shows that the performance of the model is consistent across different random samples of the data, which is what we want. In other words, it shows a robust nature. Even though the model is predicting robustly, this value for MAE is not so good, since the average value of the target variable is around 0.51, meaning that the accuracy as a percentage was around 50%. In an ideal world, we would want the MAE to be as low as possible. This is where the iterative process of machine learning comes in. At this stage, since we only have small samples of the data, we can report back to the business with these findings and recommend that the dataset needs to be further engineered, or more datasets need to be added.

This feature importance visualisation tells us:

* The product categories were not that important
* The unit price and temperature were important in predicting stock
* The hour of day was also important for predicting stock