

What is the data telling us?

We must take the data that we have to determine the demand of different parts so that companies are able to not lose out on money.

How can we use this information to come up with a supply plan?

We used the different categories as different features to run our model on.

Set up a machine learning model so that we are able to predict the supply....

Data is presented to us in categories so we use a model that can work based on that... We chose to build a Random Forest (a type of decision tree).

- Reduce overfitting, bias, params (node size)

Table Set Characteristics:

Types of Categories: 4

Types of Sub Cat for Division: 3

Types of Sub Cat for Prod Cat: 143

Types of Sub Cat for Attrib B: 132

Types of Sub Cat for Attrib C: 36

Overlap of Categories: There are tons of overlap, division data is a weak feature.

Use python programmer (good for machine learning)

```
DecisionTreeRegressor(*, criterion='squared_error', splitter='best', max_depth=None,
min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0,
max_features=None, random_state=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, ccp_alpha=0.0, monotonic_cst=None)[source]
```

History:

1. We ran program and received an error 75.922225% error (which is the average of the std)

Max_depth = 10 | random_state = 42

2.

```
model = DecisionTreeRegressor(  
    max_depth=10,  
    splitter="best",  
    min_samples_leaf=1,  
    random_state=42)
```

MAE: 75.923

3.

```
model = DecisionTreeRegressor(  
    max_depth=12,  
    splitter="best",  
    min_samples_leaf=1,  
    random_state=42)
```

MAE: 61.8236

4.

```
model = DecisionTreeRegressor(  
    max_depth=15,  
    splitter="best",  
    min_samples_leaf=1,  
    random_state=42)
```

MAE: 47.4765

5.

```
model = DecisionTreeRegressor(  
    max_depth=22,  
    splitter="best",  
    min_samples_leaf=1,  
    random_state=42)
```

MAE: 27.443

6.

```
model = DecisionTreeRegressor(  
    max_depth=25,  
    splitter="best",  
    min_samples_leaf=1,  
    random_state=42)
```

```
max_depth=30,  
splitter="best",  
min_samples_leaf=1,  
random_state=42)
```

MAE: 23.985

7.

```
model = DecisionTreeRegressor(  
    max_depth=35,  
    splitter="best",  
    min_samples_leaf=1,  
    random_state=42)
```

MAE: 23.714

We ran the model and found that we were stuck at 23%.

Debugging:

(may be causing overfitting)

1. Changing the min_samples leaf = 8 (may be causing overfitting)
 - a. It made things worse.
2. Change the min_smples leaf = 1
 - a. It lowered the error

(may be because of the testing data size = .2)

1. Change it to .15, but need to watch out because a smaller TD means larger Training set and makes it easier to match

()