

The best model considering models from step 1 and 3 is Random Forest Regressor without hyper parameter tuning. So, we consider random forest regressor algorithm for bi-directional elimination as wrapper method

```
from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

```
import pandas as pd
```

```
#reading csv file
data = pd.read_csv('/content/drive/MyDrive/new_election_dataset (1).csv')
data.head()
```

	TimeElapsed	time	territoryName	totalMandates	availableMandates	numParis
0	0	8	0	0	16	
1	0	8	0	0	16	
2	0	8	0	0	16	
3	0	8	0	0	16	
4	0	8	0	0	16	

5 rows x 29 columns

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import accuracy_score
from mlxtend.feature_selection import SequentialFeatureSelector
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import mean_squared_error
```

```
# Preparing Data set
# dropping Final Mandate variable from X
#assign the value of y for training
x = data.drop(columns=['FinalMandates'])
y = data[["FinalMandates"]]
```

```
#Standardizing value of x by using standardscaler to make the data normally distr
sc = StandardScaler()
a = sc.fit_transform(x)
df_new_x = pd.DataFrame(a,columns=x.columns)
df_new_x.head()
```

	TimeElapsed	time	territoryName	totalMandates	availableMandates	numP
0	-1.752045	-1.206238	-1.741356	-0.767282	0.979379	-(
1	-1.752045	-1.206238	-1.741356	-0.767282	0.979379	-(
2	-1.752045	-1.206238	-1.741356	-0.767282	0.979379	-(
3	-1.752045	-1.206238	-1.741356	-0.767282	0.979379	-(
4	-1.752045	-1.206238	-1.741356	-0.767282	0.979379	-(

5 rows x 28 columns

```
# Assuming X is your feature matrix and y is your target variable
X_train, X_test, y_train, y_test = train_test_split(df_new_x, y, test_size=0.2, r
```

```
# Using Random Forest Regressor
model = RandomForestRegressor()
```

```
# preform bidirectional elimination for feature selection
feature_selector = SequentialFeatureSelector(model, k_features=6, forward=True, f
feature_selector.fit(X_train, y_train)
```

```
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
selector.est_.fit(X[:, IDX], y, **fit_params)
```

```
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
```

Page 4 of 14

7. *Chlorophyll a* and *Chlorophyll b* (mg/g)

Page 6 of 14

[https://colab.research.google.com/drive/1jEnYaD\\_Ky7p2pDXL2KTAzRbJ\\_eMs3CHK#scrollTo=8IDKzIfmHYwt](https://colab.research.google.com/drive/1jEnYaD_Ky7p2pDXL2KTAzRbJ_eMs3CHK#scrollTo=8IDKzIfmHYwt)

```
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
```



[illegible]

```

    selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
    selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
    selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
    selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
    selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
    selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
    selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
    selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
    selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
    selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
    selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
    selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
    selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
    selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
    selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
    selector.est_.fit(X[:, IDX], y, **fit_params)
/usr/local/lib/python3.10/dist-packages/mlxtend/feature_selection/utilities.py
    selector.est_.fit(X[:, IDX], y, **fit_params)

```



```
# Get the selected features with indices
selected_features_indices = feature_selector.k_feature_idx_

# Convert indices to a list
selected_features_indices_list = list(selected_features_indices)

# Train the final model with the selected features
final_model = RandomForestRegressor()
final_model.fit(X_train.iloc[:, selected_features_indices_list], y_train)

# Make predictions on the test set
y_pred = final_model.predict(X_test.iloc[:, selected_features_indices_list])

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
print('Mean Squared Error:', mse)

#Get the names of the selected features
selected_features_names = X_train.columns[selected_features_indices_list]

# Display the names of the selected features
print('Selected Features:')
print(selected_features_names)
```

<ipython-input-20-de54625603cc>:9: DataConversionWarning: A column-vector y was passed when you used fit: a 1D array was expected, but got an ndarray instance. Resolving this means converting y into a column vector.

```
final_model.fit(X_train.iloc[:, selected_features_indices_list], y_train)
Mean Squared Error: 0.0014077447504772295
Selected Features:
Index(['totalMandates', 'availableMandates', 'votersPercentage', 'Party',
      'Percentage', 'Hondt'],
      dtype='object')
```

```
# Make predictions on the training set
y_train_pred = final_model.predict(X_train.iloc[:, selected_features_indices_list])

# Evaluate the model on the training set
mse_train = mean_squared_error(y_train, y_train_pred)
print('Mean Squared Error (Training):', mse_train)

# Make predictions on the test set
y_test_pred = final_model.predict(X_test.iloc[:, selected_features_indices_list])

# Evaluate the model on the test set
mse_test = mean_squared_error(y_test, y_test_pred)
print('Mean Squared Error (Testing):', mse_test)
```

```
Mean Squared Error (Training): 2.5998909041320057e-05
Mean Squared Error (Testing): 0.0014077447504772295
```

```
import matplotlib.pyplot as plt

# Plotting bar plots for training and testing MSE
labels = ['Training', 'Testing']
mse_values = [mse_train, mse_test]

plt.bar(labels, mse_values, color=['blue', 'orange'])
plt.xlabel('Dataset')
plt.ylabel('Mean Squared Error (MSE)')
plt.title('Model Performance on Training and Testing Sets')
plt.show()
```



