FALL 2024

ENG 345 TERM PROJECT REPORT

Submitted by

Alpaslan KURT 200106006015

Yusuf YAZICIOĞLU 200106006013

Under The Guidance of

Asst. Prof. Mahmud Rasih ÇELENLIOĞLU

DEPARTMENT OF ENGINEERING

GEBZE TECHNICAL UNIVERSITY

01/2025

Table of Contents

1. PROBLEM DEFINITON

Scaled Sound Pressure (or Sound Pressure Level, SPL) is used in aeroacoustics to analyze noise generated by the airfoil, particularly in applications like wind turbines, UAVs, or aircraft wings. Engineers and researchers use Computational Fluid Dynamics (CFD) simulations to obtain this value and analyze airflow over airfoils. These CFD simulations are computationally expensive and time consuming. Establishing a good ML model can decrease the use of the computationally expensive CFD simulations for basic analysis of airflow over airfoils and save resources. The goal of this problem is predicting the target variable SPL (Sound Pressure Level) as accurately as possible, given a set of features (independent variables).

2. DATA COLLECTION

The Airfoil Self-Noise dataset provided by NASA is used in this project.

NASA data set, obtained from a series of aerodynamic and acoustic tests of two and three-dimensional airfoil blade sections conducted in an anechoic wind tunnel. The NASA data set comprises different size NACA 0012 airfoils at various wind tunnel speeds and angles of attack. The span of the airfoil and the observer position were the same in all of the experiments ("UCI machine learning repository," n.d.). The variables table of the dataset is provided in Figure 1.

Variables Table

Variable Name	Role	Type	Description	Units
frequency	Feature	Integer		Hz
attack-angle	Feature	Continuous		deg
chord-length	Feature	Continuous		m
free-stream-velocity	Feature	Continuous		m/s
suction-side-displacement-thickness	Feature	Continuous		m
scaled-sound-pressure	Target	Continuous		dB

Figure 1. Variables table of the dataset ("UCI machine learning repository," n.d.)

2.1. Feature Explanations:

Frequency: frequency refers to the rate at which sound waves oscillate due to airflow interactions with the airfoil (Ahmed Mohamed Nagib Elmekawy, 2017).

Angle of Attack (AoA): The angle between the chord line of the airfoil and the oncoming airflow. It significantly influences lift and drag forces, as well as the stall characteristics of the airfoil (Ahmed Mohamed Nagib Elmekawy, 2017).

Chord Length: The distance from the leading edge to the trailing edge of the airfoil. It's a key scaling factor in aerodynamic analysis, affecting Reynolds number calculations (Ahmed Mohamed Nagib Elmekawy, 2017).

Free-Stream Velocity: The velocity of the undisturbed airflow far ahead of the airfoil. It is a critical parameter in determining the aerodynamic forces acting on the airfoil and influences the Reynolds number, which affects flow characteristics (Ahmed Mohamed Nagib Elmekawy, 2017).

Suction-Side Displacement Thickness: A measure of the boundary layer's effect on the external flow, representing the distance by which the external flow is displaced due to the presence of the boundary layer on the suction (upper) side of the airfoil (Ahmed Mohamed Nagib Elmekawy, 2017).

Scaled Sound Pressure Level (Target): A logarithmic measure of the effective pressure of a sound relative to a reference value. In airfoil studies, SPL is used to quantify the noise generated by the airfoil which is crucial in applications like wind turbines and aircraft wings ("Sound pressure level SPL," 2023).

An informative figure of an airfoil is provided in Figure 2.

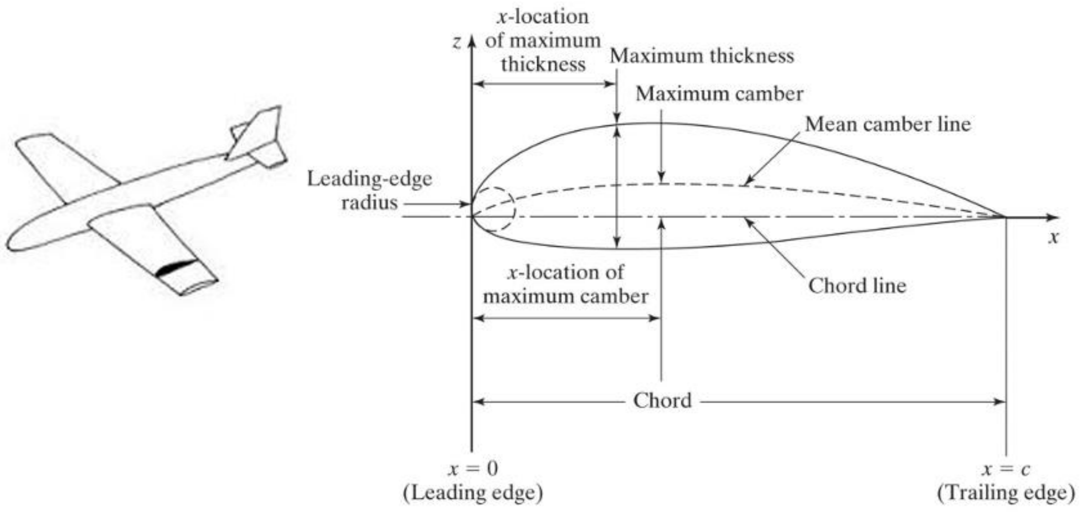


Figure 2. Airfoil structure (Ahmed Mohamed Nagib Elmekawy, 2017 p.6)

3. EDA AND FEATURE ENGINEERING

3.1. Exploratory Data Analysis (EDA):

Exploratory Data Analysis (EDA) is performed on the dataset to understand the structure of the dataset, detect patterns, and identify potential issues such as missing values, outliers, or inconsistencies. The data is summarized using descriptive statistics and visualizations are created to understand distributions and relationships between variables.

The dataset contains 1503 rows and 6 columns. This means it has 1503 samples and 6 features.

```
--- Basic Information ---

Shape of the dataset: (1503, 6)

Column Data Types:

frequency                               int64
attack_angle                          float64
chord_length                          float64
free_stream_velocity                  float64
suction_side_displacement_thickness   float64
scaled_sound_pressure                 float64
dtype: object
```

Figure 3. Basic information about data

All the features except the frequency are continuous, frequency is integer as can be seen from Figure 3.

The first five rows of the dataset is given in Figure 4.

```
First 5 Rows of the Dataset:

   frequency  attack_angle  chord_length  free_stream_velocity  \
0        800           0.0        0.3048                  71.3
1       1000           0.0        0.3048                  71.3
2       1250           0.0        0.3048                  71.3
3       1600           0.0        0.3048                  71.3
4       2000           0.0        0.3048                  71.3

   suction_side_displacement_thickness  scaled_sound_pressure
0                             0.002663                126.201
1                             0.002663                125.201
2                             0.002663                125.951
3                             0.002663                127.591
4                             0.002663                127.461
```

Figure 4. First five rows of the dataset

There is no missing value in the dataset as shown in Figure 5. Since there is no missing value and it is not desirable to delete outliers because the data comes from expensive simulation, there was no need for a cleaning process for this dataset.

```
--- Missing Values ---

frequency                              0
attack_angle                           0
chord_length                           0
free_stream_velocity                   0
suction_side_displacement_thickness    0
scaled_sound_pressure                  0
dtype: int64
```

Figure 5. Missing values

The statistical summary of the dataset is given in Figure 6.

```
--- Statistical Summary ---

          frequency  attack_angle  chord_length  free_stream_velocity  \
count   1503.000000   1503.000000   1503.000000           1503.000000
mean    2886.380572      6.782302      0.136548             50.860745
std     3152.573137      5.918128      0.093541             15.572784
min      200.000000      0.000000      0.025400             31.700000
25%      800.000000      2.000000      0.050800             39.600000
50%     1600.000000      5.400000      0.101600             39.600000
75%     4000.000000      9.900000      0.228600             71.300000
max    20000.000000     22.200000      0.304800             71.300000

       suction_side_displacement_thickness  scaled_sound_pressure
count                          1503.000000            1503.000000
mean                              0.011140             124.835943
std                               0.013150               6.898657
min                               0.000401             103.380000
25%                               0.002535             120.191000
50%                               0.004957             125.721000
75%                               0.015576             129.995500
max                               0.058411             140.987000
```

Figure 6. Statistical summary

Distributions of the features are shown in Figure. 7

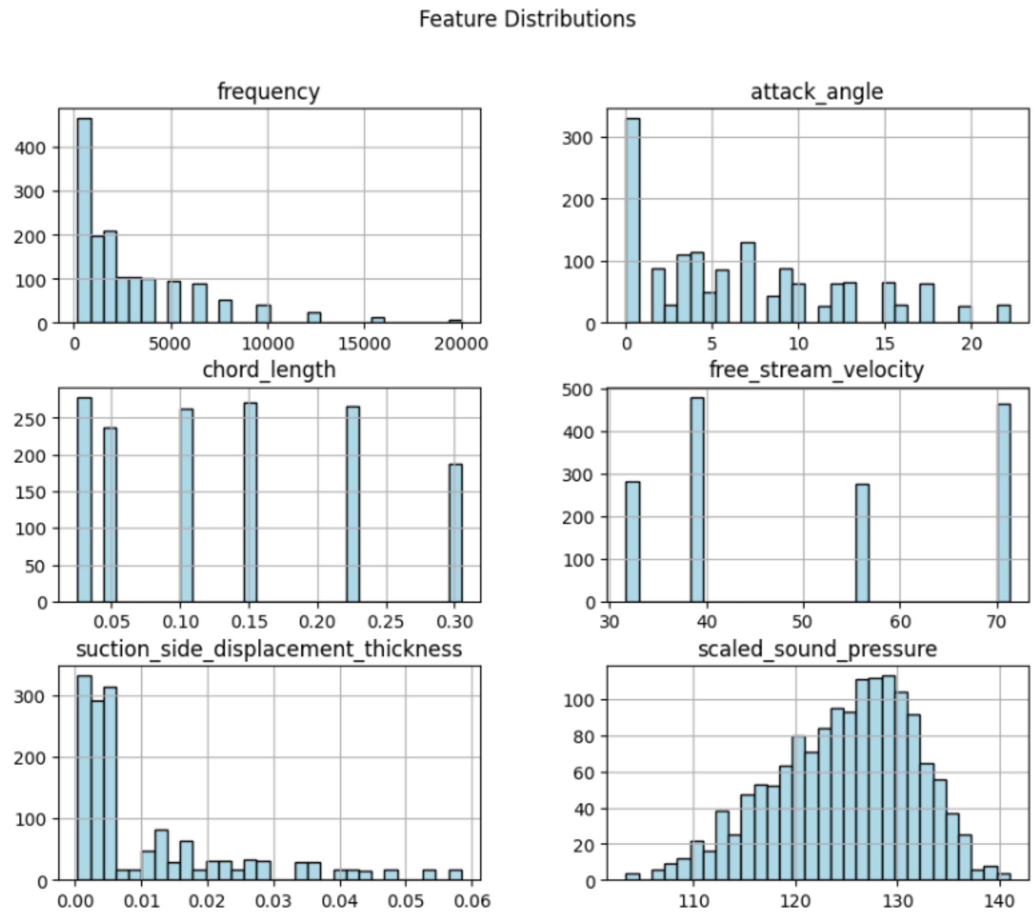


Figure 7. Distributions of the features

As shown in Figure 8, the correlation matrix shows that scaled sound pressure level (SPL) decreases with higher frequency (-0.39), thicker suction-side boundary layers (-0.31), and longer chord lengths (-0.24), suggesting these factors help reduce noise. Attack angle (-0.16) has a minor impact, while free-stream velocity (+0.13) slightly increases SPL.

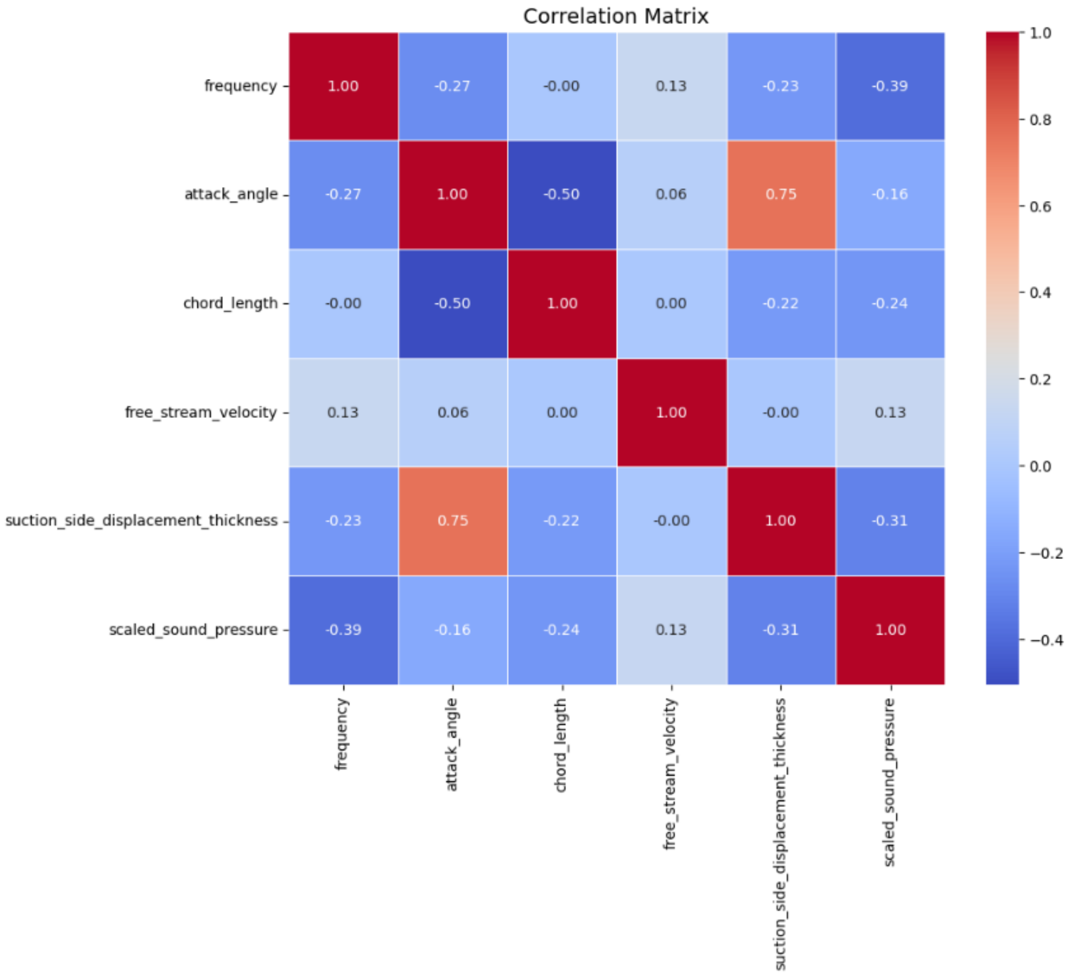


Figure 8. Correlation matrix

The boxplots of the features are given in Figure 9. As can be seen, there are outliers in some features but it is undesirable to delete those outliers.

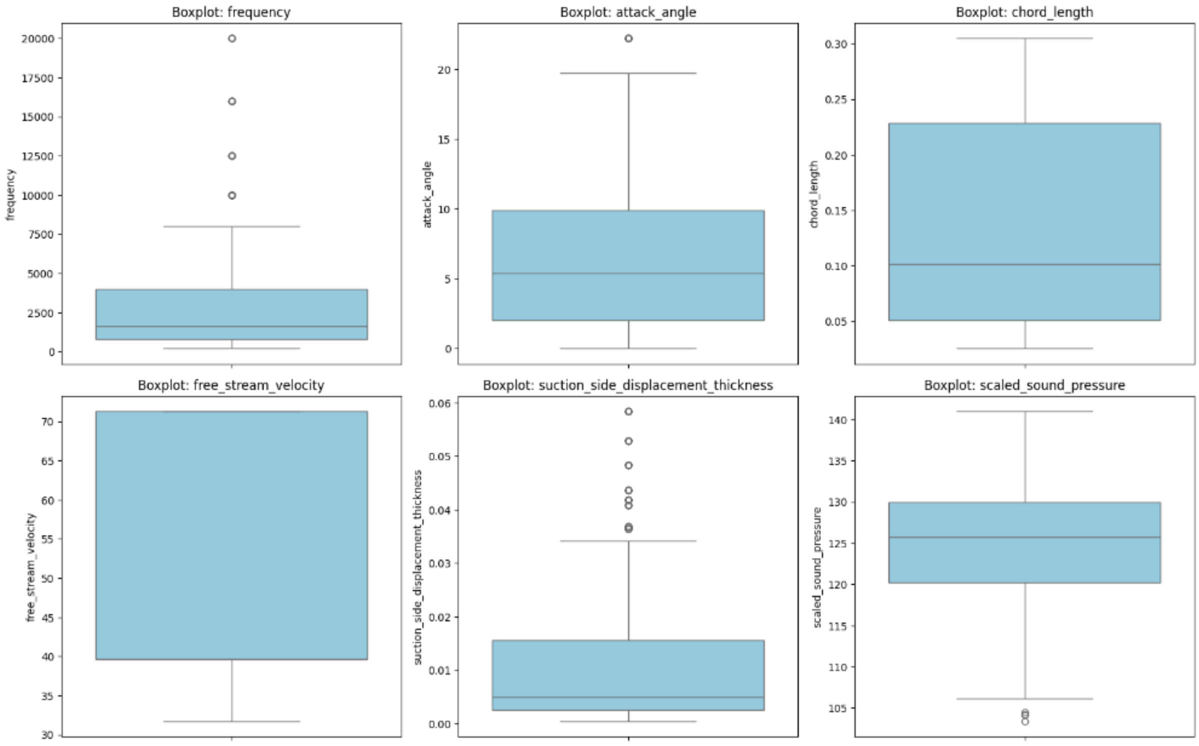
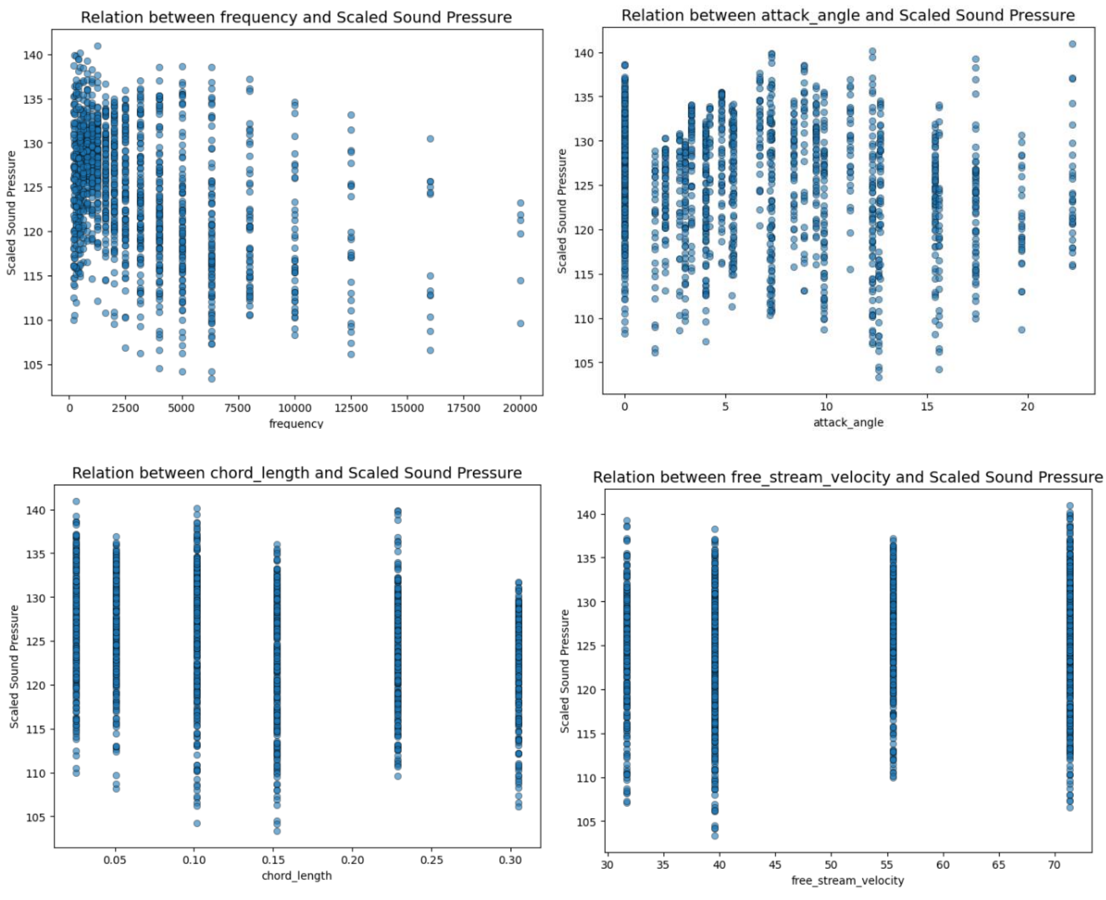


Figure 9. Boxplots of the features

Scatterplots of each feature vs. target are given in Figure 10.



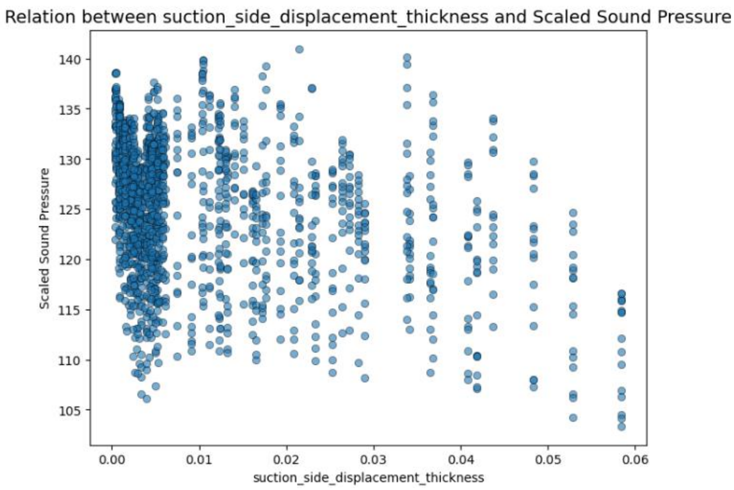


Figure 10. Scatterplots

3.2. Feature Engineering

Since there are not many features in the dataset and because of the nature of the problem, not many feature engineering techniques like PCA can be applied to the dataset. Also, EDA doesn't give a clear clue on which feature engineering techniques can be applied. Only a new feature called Reynolds Number is added. The Reynolds Number (Re) significantly affects the noise generated by an airfoil by influencing boundary layer behavior, flow separation, and turbulence.

Reynolds Number is partly represented by free stream velocity and chord length, since the density and the viscosity features of the formula is not present in the dataset, it is assumed that they don't change. So, this newly added Reynold Number feature is created by multiplying free stream velocity with chord length.

$$\text{Reynolds Number} = \text{free stream velocity} \times \text{chord length}$$

	frequency	attack_angle	chord_length	free_stream_velocity	suction_side_displacement_thickness	scaled_sound_pressure	reynolds_number
0	800	0.0	0.3048	71.3	0.002663	126.201	21.73224
1	1000	0.0	0.3048	71.3	0.002663	125.201	21.73224
2	1250	0.0	0.3048	71.3	0.002663	125.951	21.73224
3	1600	0.0	0.3048	71.3	0.002663	127.591	21.73224
4	2000	0.0	0.3048	71.3	0.002663	127.461	21.73224

Figure 11. First five rows of the dataset with the newly added Reynolds Number feature

4. APPLIED MACHINE LEARNING MODELS

The target of the dataset is numerical. So, regression models are used to predict Sound Pressure Level (target). For the training and testing of algorithms, Sci-kit Learn's train_test_split method is used. Since dataset has 1503 samples, it is appropriate to split the data 80% for training and 20% for testing. Also, to make the evaluation consistent, random state is set to 42 for each model.

In the evaluation part, mean squared error (MSE) and R^2 metrics are used to evaluate regression models.

4.1 Linear Regression

The output of the linear regression model is given in Figure 12.

```
Mean Squared Error (MSE): 22.12825415939868
R² Score: 0.5583057433518974

Model Coefficients:
frequency: -0.0012716170570063812
attack_angle: -0.40569910314622176
chord_length: -34.18381430358943
free_stream_velocity: 0.09879277969049927
suction_side_displacement_thickness: -139.47161501113263
reynolds_number: -0.005641095151919781
Intercept: 132.4927103710732
```

Figure 12. Output of the linear regression

Since the target values range between 100 and 140, an MSE of 22.13 means the squared error is relatively small compared to overall range of the target values.

An R^2 of 0.5583 means that the model explains approximately 55.83% of the variance in the target data

By looking at these indicators, it can be said that model is moderate for predicting the target.

4.2. Polynomial Regression

Since the relationship can be nonlinear, polynomial regression is also used to see if a better model can be fitted. It is tried with different degrees. The output of the polynomial regression model is given in Figure 13.

```
At Degree = 1 --> Mean Squared Error (MSE): 22.128254159398743
At Degree = 1 --> R² Score: 0.5583057433518961

At Degree = 2 --> Mean Squared Error (MSE): 15.930427281773989
At Degree = 2 --> R² Score: 0.6820183740830178

At Degree = 3 --> Mean Squared Error (MSE): 11.91043965276014
At Degree = 3 --> R² Score: 0.7622599256641522

At Degree = 4 --> Mean Squared Error (MSE): 11.429741416017388
At Degree = 4 --> R² Score: 0.7718549731912054

At Degree = 5 --> Mean Squared Error (MSE): 18.264583700472535
At Degree = 5 --> R² Score: 0.6354271031752063
```

Figure 13. Output of the polynomial regression model

As can be seen from the output, at degree = 1, the result is the same as the simple linear regression model and the best fit is the degree = 4.

For the best case (degree = 4), since the target values range between 100 and 140, an MSE of 11.43 means the squared error is relatively small compared to the overall range of the target values.

An R^2 score of 0.7719 means that the model explains approximately 77.19% of the variance in the target data.

4.3. Random Forest Regressor

Random Forest can model complex, non-linear relationships between features and the target variable without explicitly specifying a functional form. Different hyperparameter values (estimators, depth) were tried iteratively to find better fitted model. The output of the random forest regression model is given in Figure 14.

```
At Estimators = 10 and Depth = 5:
Mean Squared Error: 13.626633805293796
R-squared: 0.7280035810376508

At Estimators = 10 and Depth = 10:
Mean Squared Error: 4.187151325389644
R-squared: 0.9164217529851729

At Estimators = 10 and Depth = 15:
Mean Squared Error: 3.411445261589002
R-squared: 0.9319053474323348

At Estimators = 20 and Depth = 5:
Mean Squared Error: 13.446405417676809
R-squared: 0.7316010561534887

At Estimators = 20 and Depth = 10:
Mean Squared Error: 4.137925138435785
R-squared: 0.917404339496408

At Estimators = 20 and Depth = 15:
Mean Squared Error: 3.2813135157992934
R-squared: 0.9345028611950051

At Estimators = 30 and Depth = 5:
Mean Squared Error: 13.256829261762684
R-squared: 0.7353851187669023

At Estimators = 30 and Depth = 10:
Mean Squared Error: 4.3876265688298925
R-squared: 0.9124201375396057

At Estimators = 30 and Depth = 15:
Mean Squared Error: 3.4248428214012465
R-squared: 0.9316379234783463

At Estimators = 40 and Depth = 5:
Mean Squared Error: 12.920715301561781
R-squared: 0.7420941706754067

At Estimators = 40 and Depth = 10:
Mean Squared Error: 4.449710362813922
R-squared: 0.9111809048809361

At Estimators = 40 and Depth = 15:
Mean Squared Error: 3.4354232216981986
R-squared: 0.9314267318492853

At Estimators = 50 and Depth = 5:
Mean Squared Error: 12.925807406298373
R-squared: 0.7419925289733456

At Estimators = 50 and Depth = 10:
Mean Squared Error: 4.42884060116722
R-squared: 0.9115974788135455

At Estimators = 50 and Depth = 15:
Mean Squared Error: 3.498130178596811
R-squared: 0.9301750604560302

 With maximum r2 score the Hyper parameters: Estimators = 20 and Depth = 15
 r2 score: 0.9345028611950051
```

Figure 14. Output of the random forest regression model

As can be seen from the output, the best fit model is with 20 estimators and 15 depth.

For the best case, since the target values range between 100 and 140, an MSE of 3.28 means the squared error is very small compared to the overall range of the target values.

An R^2 score of 0.9345 means that the model explains approximately 93.45% of the variance in the target data. Such a high R^2 score suggests the Random Forest model captures most of the patterns in the data with these specific hyperparameters.

4.4. Support Vector Regressor

SVR uses kernels to transform the data into higher dimensions. This allows it to model complex, non-linear patterns between features and the target variable. Different kernels were tried to fit a better model. The output of the SVR model is given in Figure 15.

```
Kernel: linear
Mean Squared Error (MSE): 22.20129269906737
R-squared (R2) Score: 0.5568478468882525

Kernel: rbf
Mean Squared Error (MSE): 14.053315525925862
R-squared (R2) Score: 0.719486738088255

Kernel: poly
Mean Squared Error (MSE): 18.503420180696597
R-squared (R2) Score: 0.6306597726468641
```

Figure 15. Output of the SVR model

As can be seen from the output, the best fit model is the one with the radial basis function (RBF) kernel.

For the best case, since the target values range between 100 and 140, an MSE of 14.05 means the squared error is relatively small compared to the overall range of the target values.

An R^2 score of 0.7195 means that the model explains approximately 71.95% of the variance in the target data.

4.5. Artificial Neural Network

Artificial Neural Networks (ANNs) are well-suited for this dataset because they can effectively model complex, nonlinear relationships, which are common in aerodynamic problems like predicting sound pressure from airfoil characteristics.

4.5.1. ANN with Stochastic Gradient Descent

The structure includes an input layer with 64 neurons and ReLU is used as the activation function to introduce nonlinearity. Two hidden layers, each with 16 neurons and ReLU activations, are included to capture complex patterns and interactions between the features. A single neuron in the output layer is used for predicting the continuous target variable, making it suitable for regression tasks. Training is performed with a batch size of 1 which allows Stochastic Gradient Descent (SGD) to update the weights. The output of the ANN with stochastic gradient descent is given in Figure 16.

```
Epoch 50/50
961/961 ━━━━━━━━━━━━━━━━━━━━ 1s 2ms/step - loss: 12.9529 - mae: 2.8242 - val_loss: 11.4298 - val_mae: 2.5143
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step
Mean Squared Error (MSE): 9.287161325869697
R-squared (R2) Score: 0.8146222567468696
```

Figure 16. Output of the ANN with SGD

Since the target values range between 100 and 140, an MSE of 9.287 means the squared error is relatively small compared to the overall range of the target values.

The R^2 score of 0.8146 means that approximately 81.46% of the variance in the target variable is explained by the model.

4.5.2 ANN with Mini-Batch Gradient Descent

The model design includes an input layer with a variable size (layer_size), which is selected from [32, 64, 128] to experiment with different levels of complexity. A hidden layer is included, with its size set to half the input layer's size. Both layers use the ReLU activation function to introduce nonlinearity. The hyperparameters were tuned over various configurations, including learning rates ([0.01, 0.001]), epochs ([50, 100]), and batch sizes ([16, 32]). The output of the ANN with mini-batch gradient descent is given in Figure 17.

```
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 32, Learning Rate: 0.01, Epochs: 50, Batch Size: 16, Mean Squared Error (MSE): 12.569757490831545 R2 Score: 0.7490995154354901
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 9ms/step
Input Layer Size: 32, Learning Rate: 0.01, Epochs: 50, Batch Size: 32, Mean Squared Error (MSE): 18.17950943946029 R2 Score: 0.637125240416717
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 32, Learning Rate: 0.01, Epochs: 100, Batch Size: 16, Mean Squared Error (MSE): 14.958580177230843 R2 Score: 0.7014170704882864
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 32, Learning Rate: 0.01, Epochs: 100, Batch Size: 32, Mean Squared Error (MSE): 9.943664761003342 R2 Score: 0.8015180238200655
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 32, Learning Rate: 0.001, Epochs: 50, Batch Size: 16, Mean Squared Error (MSE): 20.610426885340992 R2 Score: 0.5886025568603583
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step
Input Layer Size: 32, Learning Rate: 0.001, Epochs: 50, Batch Size: 32, Mean Squared Error (MSE): 57.83099918148211 R2 Score: -0.15434412541906117
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 32, Learning Rate: 0.001, Epochs: 100, Batch Size: 16, Mean Squared Error (MSE): 13.123331007054421 R2 Score: 0.7380498302236865
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 32, Learning Rate: 0.001, Epochs: 100, Batch Size: 32, Mean Squared Error (MSE): 19.790397144004416 R2 Score: 0.6049708805618084
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 64, Learning Rate: 0.01, Epochs: 50, Batch Size: 16, Mean Squared Error (MSE): 23.565419603991604 R2 Score: 0.5296189920990791
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 64, Learning Rate: 0.01, Epochs: 50, Batch Size: 32, Mean Squared Error (MSE): 28.955297937792345 R2 Score: 0.4220335365578126
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 64, Learning Rate: 0.01, Epochs: 100, Batch Size: 16, Mean Squared Error (MSE): 14.746301167872211 R2 Score: 0.7056542967314978
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 64, Learning Rate: 0.01, Epochs: 100, Batch Size: 32, Mean Squared Error (MSE): 11.386566608581846 R2 Score: 0.7727167702556641
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step
Input Layer Size: 64, Learning Rate: 0.001, Epochs: 50, Batch Size: 16, Mean Squared Error (MSE): 17.559011626236167 R2 Score: 0.6495107778562983
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 64, Learning Rate: 0.001, Epochs: 50, Batch Size: 32, Mean Squared Error (MSE): 30.11757736583575 R2 Score: 0.3988336878806875
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 64, Learning Rate: 0.001, Epochs: 100, Batch Size: 16, Mean Squared Error (MSE): 9.613910183874278 R2 Score: 0.8081001383317732
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 64, Learning Rate: 0.001, Epochs: 100, Batch Size: 32, Mean Squared Error (MSE): 19.257861082384597 R2 Score: 0.6156006445811975
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 128, Learning Rate: 0.01, Epochs: 50, Batch Size: 16, Mean Squared Error (MSE): 13.997364759376728 R2 Score: 0.7206035515549566
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 128, Learning Rate: 0.01, Epochs: 50, Batch Size: 32, Mean Squared Error (MSE): 13.408341501800159 R2 Score: 0.7323608365187694
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 8ms/step
Input Layer Size: 128, Learning Rate: 0.01, Epochs: 100, Batch Size: 16, Mean Squared Error (MSE): 15.01317885811839 R2 Score: 0.7003272455253735
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 128, Learning Rate: 0.01, Epochs: 100, Batch Size: 32, Mean Squared Error (MSE): 9.48743632461074 R2 Score: 0.810624638314936
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 128, Learning Rate: 0.001, Epochs: 50, Batch Size: 16, Mean Squared Error (MSE): 16.48973478896422 R2 Score: 0.670854234705074
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 11ms/step
Input Layer Size: 128, Learning Rate: 0.001, Epochs: 50, Batch Size: 32, Mean Squared Error (MSE): 21.160420062464635 R2 Score: 0.5776243375312942
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step
Input Layer Size: 128, Learning Rate: 0.001, Epochs: 100, Batch Size: 16, Mean Squared Error (MSE): 10.777975144140104 R2 Score: 0.7848646492773267
10/10 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
Input Layer Size: 128, Learning Rate: 0.001, Epochs: 100, Batch Size: 32, Mean Squared Error (MSE): 15.317511908853232 R2 Score: 0.6942525611128809
Best MSE: 9.48743632461074
Best R2 Score: 0.810624638314936
Best Hyperparameters: {'layer_size': 128, 'learning_rate': 0.01, 'epochs': 100, 'batch_size': 32}
```

Figure 17. Output of the ANN with mini-batch gradient descent

As can be seen from the output, the best hyperparameters are layer size: 128, learning rate: 0.01, epochs: 100, batch size: 32

Since the target values range between 100 and 140, an MSE of 9.487 means the squared error is relatively small compared to the overall range of the target values.

The R^2 score of 0.8106 means that approximately 81.06% of the variance in the target variable is explained by the model.

5. COMPARISON OF THE MODELS

The selection of an appropriate predictive model plays a critical role in accurately analyzing data and making reliable forecasts. This study compares the performance of various regression models and neural network methods, including Linear Regression, Polynomial Regression, Random Forest, Support Vector Regression (SVR), and Artificial Neural Networks (ANN) using different optimization techniques such as Stochastic Gradient Descent (SGD) and Minibatch. Each model's effectiveness is evaluated based on two key metrics: R^2 Score, which measures the proportion of variance explained by the model, and Mean Squared Error (MSE), which quantifies the prediction error. This analysis aims to identify the model that best balances accuracy and efficiency for the given dataset.

The MSE and R^2 scores of the models are presented in Figures 18 and Figure 19.

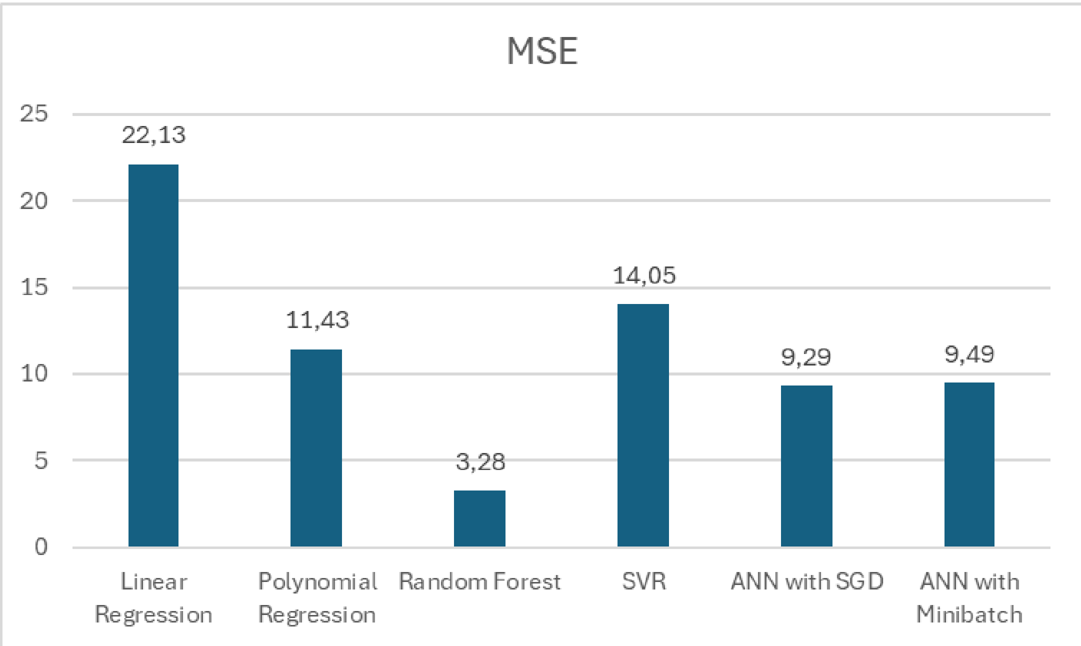


Figure 18. MSE scores of the models

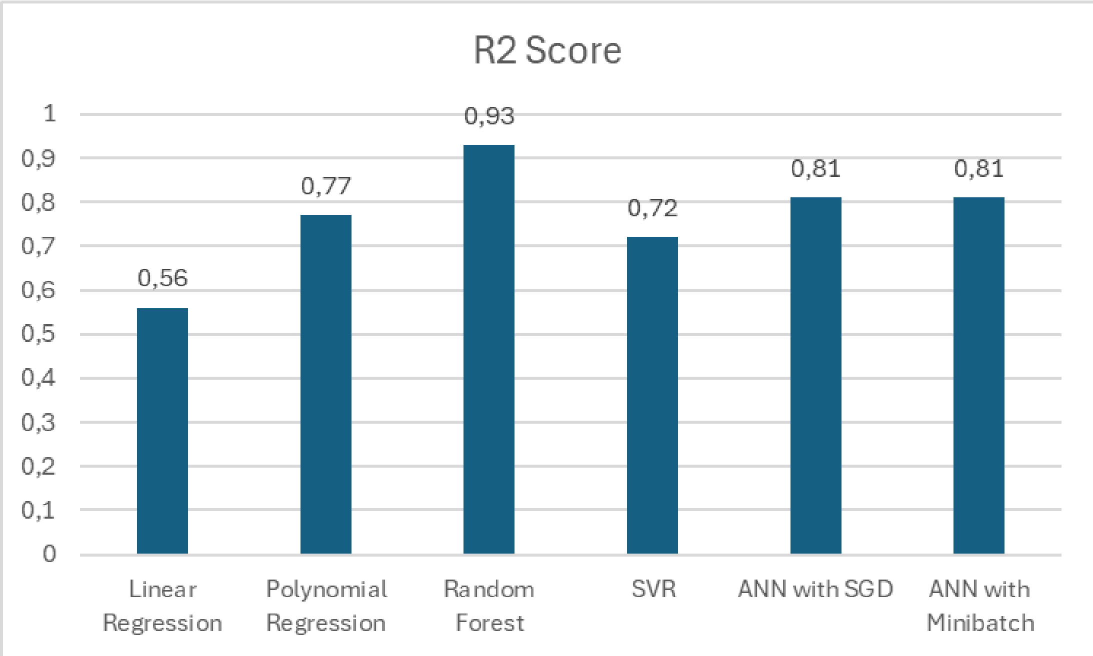


Figure 19. R^2 scores of the models

1. Random Forest is the best performer with the highest R^2 and lowest MSE. This makes it the most suitable model for this dataset.
2. Neural networks perform similarly with both optimization strategies (SGD and Minibatch), offering a good balance between complexity and accuracy.
3. Linear Regression performs the worst because of its inability to capture non-linear relationships in the data.
4. Polynomial Regression provides a substantial improvement over Linear Regression because of its ability to model non-linear relationships.
5. SVR performs reasonably well but not as good as Random Forest and neural networks.

6. CONCLUSION

In conclusion, this study explores the application of various machine learning models to predict the Scaled Sound Pressure Level (SPL) based on aerodynamic data. Among the models tested, the Random Forest Regressor demonstrated the best performance with the lowest Mean Squared Error (MSE) and the highest R^2 score. Neural networks with Stochastic Gradient Descent and Mini-Batch Gradient Descent, provided second best results and showed that they can handle complex, non-linear relationships effectively. Polynomial Regression offered substantial improvements over Linear Regression by capturing non-linearity. Support Vector Regression (SVR) achieved moderate performance. This comparative analysis highlights the importance of selecting advanced models like Random Forest and ANN for complex aerodynamic problems which emphasizes the potential of machine learning in reducing reliance on computationally expensive simulations.

7. REFERENCES

UCI machine learning repository. (n.d.). UCI Machine Learning Repository. https://archive.ics.uci.edu/dataset/291/airfoil+self+noise

Ahmed Mohamed Nagib Elmekawy. (2017, April 30). *Airfoil Parameters*. https://drahmednagib.com/. https://drahmednagib.com/onewebmedia/SPC307-2017/SPC307-Lec.8.pdf

Sound pressure level SPL. (2023, September 20). SVANTEK - Sound and Vibration. https://svantek.com/academy/sound-pressure-level-spl