Predictive Analysis: Evaluation of Regression and Classifier Metrics

Angela Diaz

National University

DDS-8551: Predictive Analysis

Abstract

This project uses predictive analytics on the Iris dataset to evaluate the effectiveness of regression and classification models in forecasting petal length. Feature engineering was conducted to enhance predictive capability, including deriving a new variable representing the ratio between sepal and petal structures. There were two regression models for this project: Petal Length vs. Sepal Length and Petal Length vs. (Sepal Length − Petal Width). Each was assessed using metrics such as ME, MAE, MSE, MPE, and MAPE. Each of the models exhibited signs of fitting issues, with low training error but higher testing error. In contrast, classification models demonstrated strong performance, achieving high precision, recall, and accuracy when predicting petal length categories. These results suggest that classification modeling is more effective than regression modeling for the iris dataset. These findings highlight the importance of feature engineering, model selection, and evaluation metrics used while developing reliable predictive models. These findings also present the need for additional or more informative predictors to improve regression accuracy.

Keywords: predictive analysis, regression modeling, classification modeling, model evaluation, supervised learning, machine learning

Predictive Analysis: Evaluation of Regression and Classifier Metrics

Data Analytics has four key types of analysis: Descriptive Analytics, Diagnostic Analytics, Predictive Analytics, and Prescriptive Analytics. These analytic types each have their own role in aiding business professionals to make data driven decisions. Descriptive Analytics provide trends from raw data to reveal what has occurred and what is occurring now in the dataset. Diagnostic Analytics compares trends and finds correlations among variables in a dataset. Predictive Analytics uses trends from past data to predict future outcomes. Prescriptive Analytics uses all the findings from Descriptive Analytics, Prescriptive Analytics, and Predictive Analytics to suggest actions to improve outcomes. It is visible that data analytics flows in a series of insightful research: first, “Reveal what happened to the data in the past”, then “Find out what trends caused that to happen”, followed by “Logically guess what will happen in the future based on the past and current trends”, lastly, “Implement a strategy to make informed decisions based on the previous analytic findings”. (Cote, 2021)

The focus will be on Predictive Analytics from a data scientist perspective. A data scientist will use predictive analysis to review past data trends in conjunction with summary statistics and incorporate the findings. Based on these findings, the data scientist can build a predictive model that will forecast an outcome. (Sheposh, 2025)

This paper will introduce a predictive model and use regression and classification metrics to determine if petal length or sepal length has better prediction of sepal width from the Iris dataset. For the regression analysis of this paper, multiple metrics will be used to determine the predictive model’s accuracy and reliability. The regression analysis metrics include Mean Error (ME), Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Mean Percentage Error (MPE). The classification analysis metrics include Accuracy, Precision, Recall, and F-score.

Iris Dataset

Iris is a well-known and widely used dataset for data science projects. The dataset comes from US Irvine's Machine Learning Repository. The Iris dataset represents parts of a plant, specifically an iris flower. The size of this dataset is small consisting only of 5 variables (sepal length (cm)', 'sepal width (cm)', 'petal length (cm)', 'petal width (cm)' and 'type') and 150 observations. (*Iris* 2025)

Since the dataset is already tidy and there are no missing values that need to be addressed, the dataset is ready to go into the next step towards predictive modeling, which is feature engineering. Feature engineering can be described as a task of creating or transforming variables so that the predictive model can understand the dataset better and ultimately contribute to better predictive outcomes.

Feature Engineering Iris Dataset

As mentioned, the dataset is very small, therefore feature engineering is applied. A new variable called 'new' was created. The values for this new variable represent the proportion of the sepal area and petal area. The reason is that there are some plants that have larger petals when compared to sepals or have smaller petals compared to sepals. This change will also allow for data analysis used during regression analysis through the ability to use Decision Trees and/or KNN in the predictive model. This change will also work well for classification analysis as it provides another way to separate (classify) the types of iris flowers. This data started with 5 variables and now has 5 variables total after feature engineering.

Testing and Training Data Split

With the number of variables that will be used for the predictive model complete, the dataset can be split into testing and training sets. Splitting the data into a training set is necessary so that the model can learn the past and current dataset trends. The testing set is then used after the model has been trained to evaluate how well it generalizes to new, unseen observations. Using the results of the testing set, the model's prediction performance is determined.

Regression Predictive Modeling: Petal Length vs Sepal Length

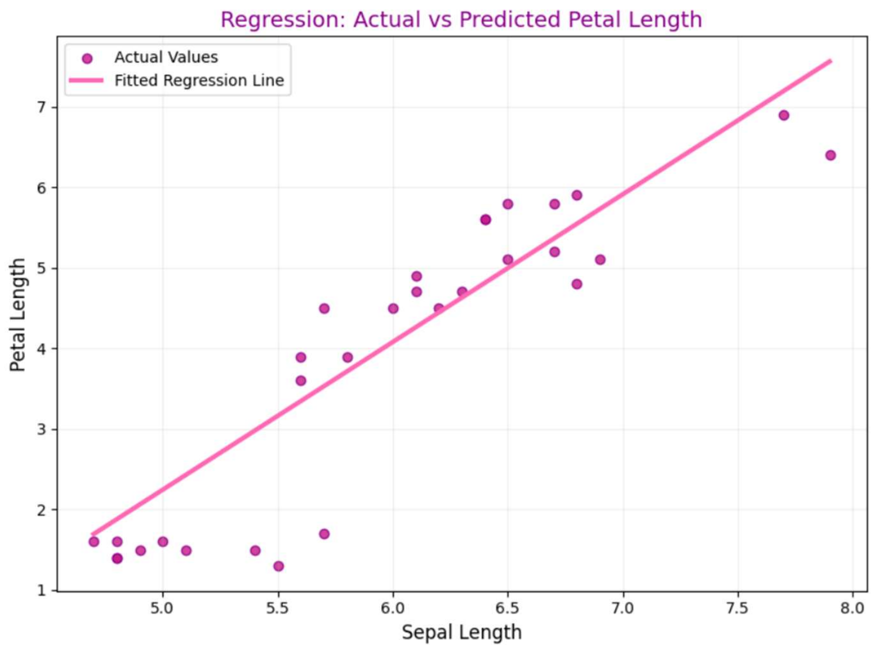
Regression Predictive Modeling is used to examine the dataset and predict a specific numerical value. These models use quantitative dependent variables to evaluate the model's performance via statistical metrics. The metric results may measure variability and bias.

Training: Petal Length vs Sepal Length

The mean of petal length was used for the estimators. The metrics used to evaluate the model were the mean error, mean percentage error, mean absolute error, mean squared error, and mean absolute percentage error.

Metric	Result
Mean Error	0.0
Mean Percentage Error	-0.121
Mean Absolute Error	0.129
Mean Squared Error	0.182
Mean Absolute Percentage Error	0.288

The training results of the mean error reveal the model has predictions that are not high nor low. The results of the mean percentage error reveal the model has a -1.3% unpredicted percentage. The results of the mean absolute error reveal the model is differs by 0.237 centimeters. The results of the mean squared error reveal the model has a small error. The results of the mean absolute squared error reveal the model is incorrect by an average of 8%. Overall, the model has no bias, slightly underprediction, good accuracy, and low average error.

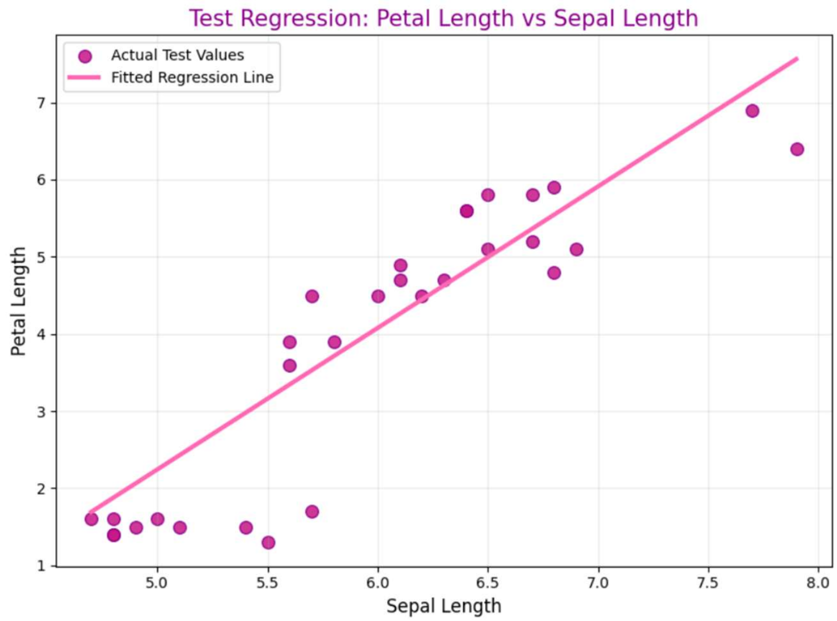


Testing: Petal Length vs Sepal Length

The same was done on the test data. The results are slightly different.

Metric	Result
Mean Error	0.0
Mean Percentage Error	-0.121
Mean Absolute Error	0.729
Mean Squared Error	0.782
Mean Absolute Percentage Error	0.288

The test results of the mean error reveal the model has a mean error close to 0 meaning that predictions that are not high nor low. The results of the mean percentage error reveal the model has an unpredicted percentage of -12.1%. The results of the mean absolute error reveal the model differs by 0.729 centimeters. The results of the mean squared error reveal the model has a small error, 0.782. The results of the mean absolute percentage error reveal the model is incorrect by an average of 28.8%. Overall, the test metrics reveal no bias, but they do indicate stronger underprediction, higher error, and weaker generalization performance. This means that using only sepal length leads to overfitting and does not hold a strong enough predictive power.



Training and Testing Comparison: Petal Length vs Sepal Length

The model appears to perform better on the training data and weaker on the testing data; errors are lower on training data than testing data. The training error is a lot lower than the testing error, which means there is enough evidence to suggest that the model is overfitted. Overall, the model is limited when using only petal length to predict sepal length.

Regression Predictive Modeling: Petal Length vs (Sepal Length - Petal Width)

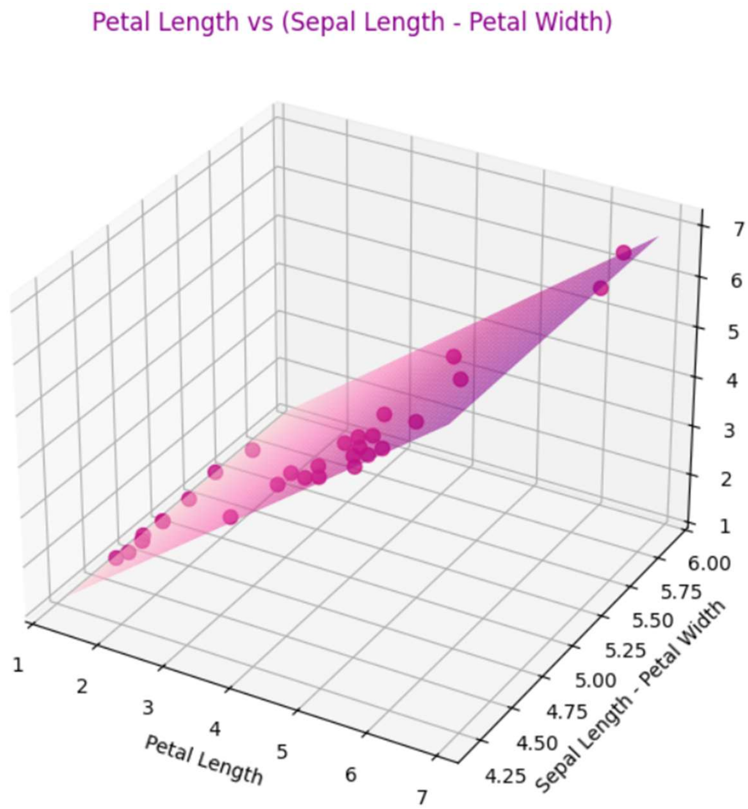
Regression Predictive Modeling can also be used with features that are combined.

Training: Petal Length vs (Sepal Length - Petal Width)

The mean of petal length vs (Sepal Length - Petal Width) was used for the second estimators The metrics used to evaluate the model were the mean error, mean percentage error, mean absolute error, mean squared error, and mean absolute percentage error.

Metric	Result
Mean Error	-0.008
Mean Percentage Error	0.018
Mean Absolute Error	0.025
Mean Squared Error	0.025
Mean Absolute Percentage Error	0.018

The training results of the mean error reveal the model has predictions that are not high or low. The results of the mean percentage error reveal the model has a 1.8% which suggest the model aligns with actual values. The results of the mean absolute error reveal the model differs by 0.025 centimeters. The results of the mean squared error reveal the model has a small error. The results of the mean absolute percentage error reveal the model is incorrect by an average of 1.8%. Overall, the model has na little bias, small errors, okay accuracy, and low average error.



Training and Testing Comparison: Petal Length vs Sepal Length

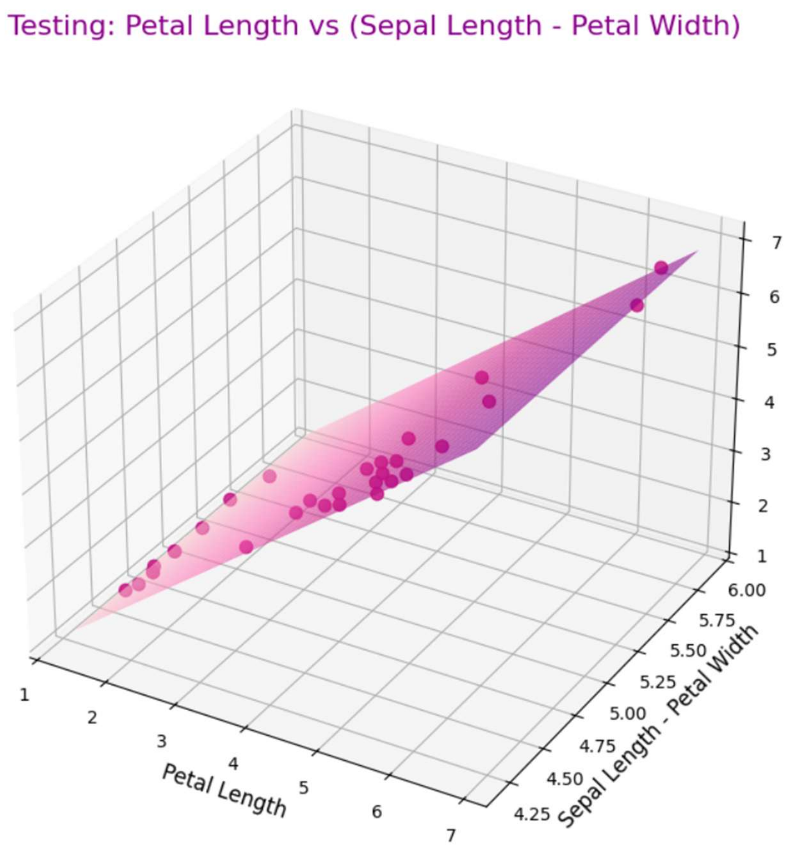
The model appears to perform better on the training data and weaker on the testing data; training data has lower errors than testing errors across all the metrics. The test error is higher, which may be indicative of overfitting. These two predictors hold limited predictive power. Overall, the model on testing data suggests that more predictors are likely needed to predict petal length.

Testing: Petal Length vs (Sepal Length - Petal Width)

The mean of petal length vs (Sepal Length - Petal Width) was used for the second estimators on the test data. The metrics used to evaluate the model were the mean error, mean percentage error, mean absolute error, mean squared error, and mean absolute percentage error.

Metric	Result
Mean Error	-0.033
Mean Percentage Error	0.071
Mean Absolute Error	0.1
Mean Squared Error	0.1
Mean Absolute Percentage Error	0.071

The testing results of the mean error reveal the model has a small amount of bias. The results of the mean percentage error reveal the model has a 7.1% which suggests the model has a little proportional deviation from actual values. The results of the mean absolute error reveal the model differs by 0.01 centimeters. The results of the mean squared error reveal the model has a small error. The results of the mean absolute percentage error reveal the model is incorrect by an average of 7.1%. Overall, the model has no bias, small error, high accuracy, and low average error.



Training and Testing Comparison: Petal Length vs (Sepal Length - Petal Width)

The second model appears to perform better on the training data than the test data. The test data metrics have higher error values which is indicative of the model not generalizing that well. The two metrics are still not enough to hold high predictive power. Overall, the testing data metrics suggest that more predictors that are more informative need to be added to achieve better accuracy.

Testing Comparison: Petal Length vs Sepal Length & Petal Length vs (Sepal Length and Petal Width)

The testing results of the first model use only sepal length to predict petal length. The first model captures only a basic linear relationship. The testing results also suggest that the first model is underfit and holds insufficient predictive power.

The testing results of the second model uses two predictors, making it a more flexible model. Even though the second model used these two predictors, the improvement of test results are modest.

The second model does show improvement from the first models fit on the training data. However, the second model suffers slightly from overfitting.

In the end of the regression model comparison on test data informs that each model suffers from inefficiencies for properly predicting petal length. The models should add more predictors to improve the metrics across the board.

Classification Predictive Modeling

Classification is a method in which the data is grouped into categories. (Pan et al., 2025) When it comes to predictive modeling, classification is a type of supervised learning. The goal of classification predictive modeling is to predict an outcome; typically, a categorical outcome based on an input(s). The target variable is categorical and may include binary, multiclass and ordain data. It may be used with machine learning or deep learning models.

Training: Classification of Petal Length

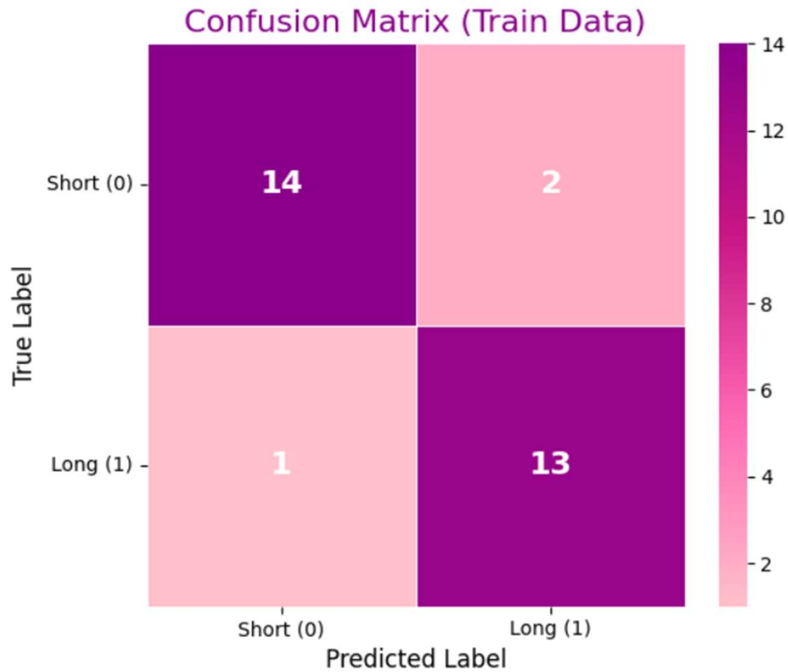
The metrics used to evaluate the classification model are precision, recall, f1-score and accuracy. The table below will display the metric results for the training data.

Class	Precision	Recall	F1-Score
0	0.98	0.97	0.98
1	0.96	0.98	0.97

The precision results for class 0 show that the model predicts class 0 98% of the time. The recall result for class 0 displays a 97% score for successfully identifying true class 0 samples. The F1 score of 0.98 means that there is a balance between the precision and recall of the model. Overall, the model is pretty accurate when detecting class 0 and properly labeling it.

The results for class 1 are similar to class 0. Meaning, that the model is pretty accurate at detecting class 0 and class 1 on the training data.

The accuracy score for the model is 97%. This a high results which means that the use of classification is highly favorable.



Training: Classification of Petal Length

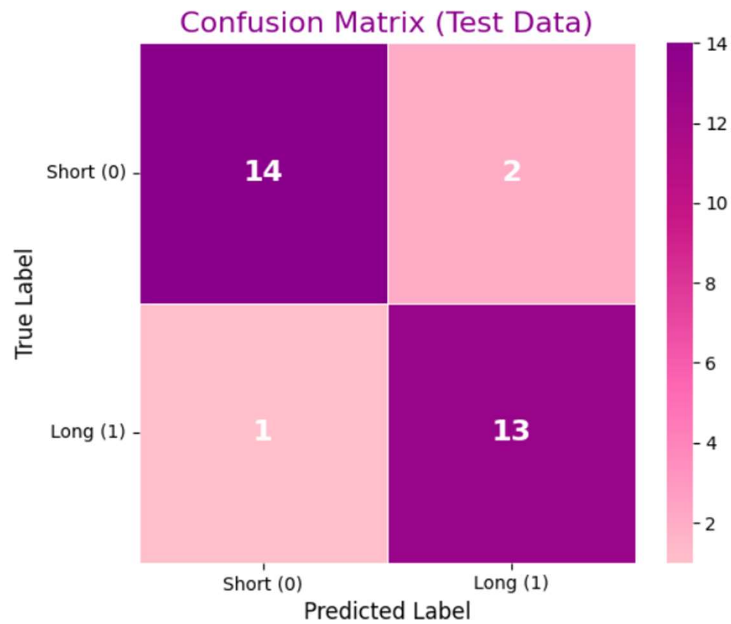
The table below will display the metric results for the training data.

Class	Precision	Recall	F1-Score
0	0.93	0.88	0.90
1	0.87	0.93	0.90

The precision results for class 0 show that the model predicts class 0 93% of the time. The recall result for class 0 displays an 88% score for successfully identifying true class 0 samples. The F1 score of 0.90 means that there is a balance between the precision and recall of the model. Overall, the model is pretty accurate when detecting class 0 and properly labeling it.

The results for class 1 are close to those of class 0. The precision score for class 1 has 87% correct predictions of class 1. The recall shows that the model captures 93% of the class 1 cases. The F1 score is 0.90 which means there is a balanced accuracy in the model.

The accuracy score for the model is 90%. This is a high result which means that the use of classification is highly favorable.



Conclusion

This project used the Iris dataset to create Regression and Classification models. These models were implemented to discover if sepal length or sepal length – petal width was better at predicting petal length. Metrics such as the mean error and precision were used to evaluate the performance of each model, regression and classification respectively.

The steps that were taken throughout this project were data preprocessing, feature engineering, data splitting, regression model analysis, and classification model analysis.

Overall, it was observed that the regression models struggled with underfitting and overfitting due to limited predictors being used. Classification models excelled compared to regression models. The classification models performed with higher accuracy. It was learned that predictive modeling may fall dependent on the number of feature and the type of model being used.

References

Bangert, P. (2021). *Machine Learning and data science in the oil and gas industry: Best practices, tools, and case studies*. Gulf Professional.

Bhavitha, C., Reddy, A. D., V, D. G., Pavan, BH. V. V. S. R. K. K., Pathak, D., & Kumaraswamy, I. (2025). Predictive accuracy of machine learning models in early detection of parkinson's disease. *2025 3rd International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, 493–501. https://doi.org/10.1109/icscds65426.2025.11167205

Li, J., Zhang, T., Luo, X., Chen, W., Peng, H., & Philip Chen, C. L. (2025). Multi-scenario human motion recognition framework combining skeleton key points extraction and classification modeling. *2025 40th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, 1592–1597. https://doi.org/10.1109/yac66630.2025.11150181

Pan, H., Zhou, Y., Xie, Y., Xie, D., Lin, J., Wang, Y., & He, Z. (2025). Channel Noise Characterization and classification modeling based on BG Models. *2025 IEEE 2nd International Conference on Energy and Electrical Engineering (EEE)*, 1–4. https://doi.org/10.1109/eee64897.2025.11162833

Pattnaik, A., Jena, J. J., Singh, J. P., Panda, A. R., Bandyopadhyay, A., & Gourisaria, M. K. (2025). A predictive modelling for breast cancer classification using Machine Learning Technique. *2025 International Conference on Visual Analytics and Data Visualization (ICVADV)*, 965–970. https://doi.org/10.1109/icvadv63329.2025.10961666

Tsanas, A., & Xifara, A. (2012). *Energy efficiency*. UCI Machine Learning Repository. https://archive.ics.uci.edu/dataset/242/energy+efficiency