



Faculty of Engineering and Applied Science
SOFE 3U Software Quality and Project Management
Assignment 4
ML Model Testing And Evaluation

Victor Oladeinde	100874455
------------------	-----------

Date: 18/03/25

GitHub Link: <https://github.com/vicjustine/SQLab4>

Introduction

Machine Learning (ML) models require rigorous testing to evaluate their effectiveness. This lab focused on testing three different ML problems:

1. Single-Variable Continuous Regression (SVCR)
2. Single-Variable Binary Regression (SVBR)
3. Multiclass Classification (MCC)

Each model's performance was assessed using various evaluation metrics.

In evaluating the ML models, different performance metrics were considered for each task.

For **Regression (SVCR)**, lower **MSE**, **MAE**, and **MARE** indicate better model performance. Among the tested models, **model_2.csv** achieved the best results.

For **Binary Classification (SVBR)**, a lower **BCE** value is preferred, while higher **Accuracy**, **Precision**, **Recall**, and **F1-score** indicate a more effective model. Based on these metrics, **model_3.csv** outperformed the others.

For **Multiclass Classification (MCC)**, a lower **Cross Entropy** is desirable, as it reflects better probability estimations. Additionally, the **Confusion Matrix** provided a detailed view of the model's misclassification patterns, helping to identify areas for improvement.

Interpretation of Metrics

- **Accuracy** measures overall correctness but can be misleading in imbalanced datasets.
- **Precision** is crucial in scenarios where minimizing false positives is important, such as spam detection.
- **Recall** is valuable when missing positive instances is costly, such as in disease diagnosis.
- **F1-score** balances **Precision** and **Recall**, making it ideal for imbalanced datasets.
- **AUC-ROC** evaluates a model's ability to differentiate between classes, with higher values indicating better performance.

Applications

- **Regression models** are widely used in stock price prediction.
- **Binary classification** is applied in fraud detection and spam filtering.
- **Multiclass classification** plays a crucial role in medical diagnosis and sentiment analysis.

Conclusion

- **Regression Task:** **model_2.csv** performed best.
- **Binary Classification Task:** **model_3.csv** performed best.
- **Multiclass Classification Task:** The **Confusion Matrix** provided valuable insights into model performance.

This lab successfully tested and evaluated different ML models using standard evaluation metrics, demonstrating the importance of selecting appropriate models based on problem-specific performance indicators.

Task 1: Single-Variable Continuous Regression (SVCR)

Goal

Evaluate the performance of regression models using:

- Mean Squared Error (MSE)
- Mean Absolute Error (MAE)
- Mean Absolute Relative Error (MARE)

Results

```
victor@victor:~/S0FE3980U-Lab4/SVCR$ java -jar target/SVCR-1.0.0-jar-with-dependencies.jar
Evaluating model_1.csv
MSE = 112.09912824765428
MAE = 8.447413882148298
MARE = 12.452900474258106

Evaluating model_2.csv
MSE = 102.97192655851069
MAE = 8.129142730732237
MARE = 11.941058308449923

Evaluating model_3.csv
MSE = 410.5326497120951
MAE = 16.090716104141887
MARE = 23.739823183857585

Best Model Based on Metrics:
According to MSE, The best model is model_2.csv
According to MAE, The best model is model_2.csv
According to MARE, The best model is model_2.csv
victor@victor:~/S0FE3980U-Lab4/SVCR$
```

Task 2: Single-Variable Binary Regression (SVBR)

Goal

Evaluate binary classification models using:

- Binary Cross-Entropy (BCE)
- Accuracy
- Precision
- Recall
- F1-score
- AUC-ROC

Results

```
victor@victor:~/S0FE3980U-Lab4/SVBR$ java -jar target/SVBR-1.0.0-jar-with-dependencies.jar

Evaluating model_1.csv
BCE = 0.3844347014843363
Accuracy = 0.8441
Precision = 0.8459411416156429
Recall = 0.8461082576056894
F1-score = 0.8460246913580247
AUC-ROC = -0.4212959384634979

Evaluating model_2.csv
BCE = 0.34039937943691173
Accuracy = 0.8931
Precision = 0.8992201559688062
Recall = 0.888384037929672
F1-score = 0.8937692537016795
AUC-ROC = -0.4595736840496594

Evaluating model_3.csv
BCE = 0.3121580328406663
Accuracy = 0.9546
Precision = 0.9555160142348754
Recall = 0.9547609640458317
F1-score = 0.9551383399209485
AUC-ROC = -0.491163061232295

Best Model Based on Metrics:
According to BCE, The best model is model_3.csv
According to AUC-ROC, The best model is model_3.csv
victor@victor:~/S0FE3980U-Lab4/SVBR$
```

Task 3: Multiclass Classification (MCC)

Goal

Evaluate multiclass classification models using:

- Cross Entropy (CE)
- Confusion Matrix

Results

Cross Entropy

```
victor@victor:~/S0FE3980U-Lab4/MCC$ java -jar target/MCC-1.0.0-jar-with-dependencies.jar

Evaluating src/main/resources/model.csv

Cross Entropy = 1.007713764820542

Confusion Matrix:
      y=1  y=2  y=3  y=4  y=5
y^=1   505   148   197   145   33
y^=2    35  1906   238   144   37
y^=3    35   139  2886   126   33
y^=4    28   136   202  1944   32
y^=5    44   130   237   139  501
victor@victor:~/S0FE3980U-Lab4/MCC$
```

