Untitled document

Camarines Sur Polytechnic Colleges

Document Details

Submission ID
trn:oid:::23367:82975736

Submission Date
Feb 21, 2025, 10:31 PM GMT+8

Download Date
Feb 21, 2025, 10:31 PM GMT+8

File Name
Untitled document

File Size
3.1 KB

3 Pages

433 Words

2,693 Characters

2% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

- **1** Not Cited or Quoted 2%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

- 0% Internet sources
- 0% Publications
- 2% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- 1 Not Cited or Quoted 2%
  Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%
  Matches that are still very similar to source material
- 0 Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

0% Internet sources
0% Publications
2% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1 Submitted works

Georgia Institute of Technology Main Campus on 2024-11-04 **2%**

SEAN XANDER B. AQUINO BSCS - 2A PROBLEM SET 3

The experimental data comes from the Breast Cancer Wisconsin (Diagnostic) Dataset. The database contains nucleus-related features including texture analysis along with radius measurement and symmetry assessment in addition to smoothness observations that help determine tumor malignancy. The target variable has received binary data format preprocessing for the dataset.

Benign (B) → 0

Malignant (M) → 1

Standardization processed the features while the dataset split into training and testing portions at 80-20.

Regression Results

The selected features helped the Logistic Regression Model classify cancer tumors through its predictive operation. A total of three different versions of this model received training.

Original Model (No Regularization)

L1 Regularized Model (Lasso Regression)

L2 Regularized Model (Ridge Regression)

Model Performance:

Original Model Accuracy: 71.05%

1 L1 Regularized Model Accuracy: 71.05%

L2 Regularized Model Accuracy: 71.05%

All three models demonstrated identical accuracy results which implied that the implementation of regularization did not affect the performance of predictions within this analysis.

Effect of Regularization on Accuracy

Regularization implements as a technique to enhance model performance as well as minimize overfitting scenarios. The accuracy results from these models remained identical to each other. Possible reasons include:

Feature Selection and Redundancy

Through L1 Regularization (Lasso) some features obtain zero coefficient values resulting in sparse models. The effect of feature selection on the data could minimize when the current dataset possesses well-selected features.

While the L2 Regularization (Ridge) minimizes coefficient values instead of eliminating any features from the model. The use of L2 regularization would lead to no practical benefit when overfitting is absent within the system.

Dataset Characteristics

When the dataset contains powerful predictors then the benefits of regularization will likely diminish.

Regularization shows maximum impact on datasets that have large numbers of correlated features while lacking in this particular analysis.

Hyperparameter Tuning

Default value of C=1.0 may not provide the most suitable regularization strength. Management of C value from 0.1 to 10 will produce distinct algorithm outcomes.

The analysis of feature importance needs to verify if the L1 regularization causes any coefficients to decrease, identifying any useless features.

Different values should be tested to study if regularization strength leads to enhanced outcomes. It should supplement the performance metrics by examining the values of precision, recall and F1-score for deeper evaluation results.

The effect of regularization on a dataset depends on its unique characteristics even though it provides value through overfitting and irrelevant feature control. The models demonstrated equivalent performance after regularization because it did not enhance predictive power. Additional parameter adjustments and feature examination would lead to an enhanced model accuracy rate.