

JANUARY 18, 2024.

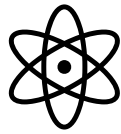
DATA SCIENCE - HEALTHCARE

PERSISTENCY OF A DRUG.

abc FARMA



Jorge Jaramillo Bermudez
HEALTH DATA SCIENTIST.



SUMMARY OF THE MACHINE LEARNING MODELS



SUMMARY AND CONCLUSIONS:

In the exploration of medication persistence, four classification models underwent training: Logistic Regression, Decision Tree Classifier, MLP Classifier, and RandomForest Classifier. The summarized results for each model are as follows:

Logistic Regression:

- Overall Accuracy: 80%
- Area under the ROC curve (ROC-AUC): 0.87
- Precision and Recall: Reasonable for both classes.

Decision Tree Classifier:

- Overall Accuracy: 73.72%
- ROC-AUC: 0.72
- Precision and Recall: Acceptable but with room for improvement.

MLP Classifier:

- Overall Accuracy: 74.74%
- ROC-AUC: 0.82
- Precision and Recall: Reasonable but with a convergence warning.

RandomForest Classifier:

- Overall Accuracy: 79.27%
- ROC-AUC: 0.86
- Precision and Recall: Balanced for both classes.

Furthermore, a hyperparameter search was executed for RandomForest Classifier, yielding the optimal parameters: max_depth: 20 and n_estimators: 200.

Conclusion:

In this analysis, the RandomForest Classifier emerged as the most effective model, boasting an overall accuracy of 79.27% and a ROC-AUC of 0.86. This model surpasses others in achieving a balance between precision and recall for both classes. Its ability to distinguish between positive and negative classes, highlighted by the area under the ROC curve, underscores its robust performance. The findings endorse the RandomForest Classifier as the preferred choice for drug persistency identification, offering a promising avenue for further refinement and application.