Rakamin Academy

Homework

Stage-3

Presented by NautixTech

Final Model Testing

Rakamin Academy

Model	Accuracy		Precision	Recall	F1 Score	ROC AUC	
	Train	Test				Train	Test
Logistic Regression	0,37	0,37	0,33	0,37	0,32	0,68	0,68
Bagging	0,96	0,65	0,64	0,65	0,64	1,00	0,86
Random Forest Classifier	0,99	0,66	0,66	0,66	0,66	1,00	0,89
AdaBoost Classifier	0,50	0,50	0,50	0,50	0,47	0,78	0,78
Gradient Boosting Classifier	0,61	0,60	0,61	0,60	0,58	0,87	0,86

Model Evaluation

Rakamin Academy

In this project, the evaluation metrics used are accuracy and ROC AUC.

Logistic Regression: Consistent performance between train and test data with an accuracy of 0.37. However, it potentially underfits due to the low ROC AUC value.

Bagging: Based on the difference in accuracy and ROC AUC values, there is a potential for overfitting.

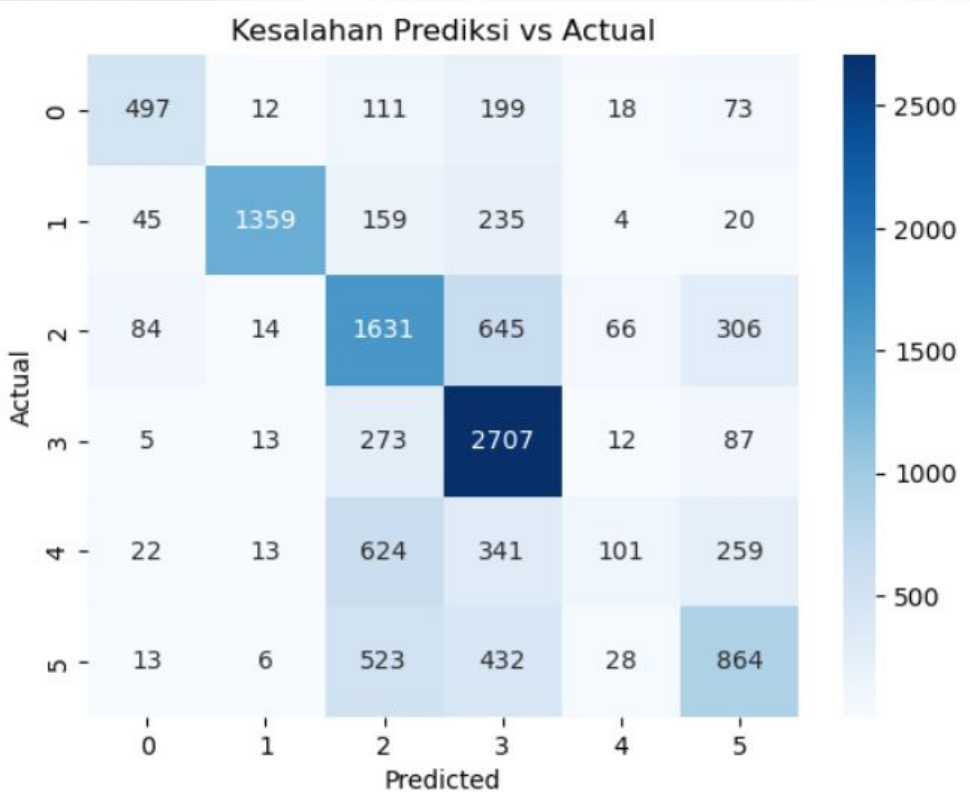
Random Forest Classifier: Based on the difference in accuracy and ROC AUC values, there is a potential for overfitting.

AdaBoostClassifier: Consistent performance between train and test data when viewed from accuracy and ROC AUC.

Gradient Boosting Classifier: Consistent performance between train and test data when viewed from accuracy and ROC AUC.

Based on the above description, **the best model is the Gradient Boosting Classifier**. This model does not have the potential for underfitting or overfitting. The similar results between accuracy and ROC AUC indicate that this model has the best generalization among the other models and its performance tends to be consistent between train and test data.

Error Analysis



Causes of Error:

- In this project, the Actual = 4 and Actual = 2 classes have a higher number of errors, indicating that the model may have difficulty predicting the dominant class.
- There may be noise in the dominant data.
- The model may not be well-suited to the data type.

Review: Since the target feature used is fixed, the data pre-processing section will be reviewed and doing hyperparameter turning.

Adjustments

Rakamin Academy

Based on the error analysis description, the following actions will be taken:

- Adjusting data pre-processing
- Trying to use other libraries for NLP cases, such as Hugging Face

Final Report Drafting

Rakamin Academy

Report GDocs:

https://docs.google.com/document/d/1PbPC7mcbhKlO41pEy-IEll615LzhWsNq/edit?usp=sharing&ouid=106398992337656366053&rtpof=true&sd=true

Git:

Final_Project_NautixTech/Stage 3 at main • tatashandharu15/Final_Project_NautixTech

Thank You

Rakamin
Academy