

Evaluation Report

Task:- To predict developer's role

Model Prediction

| <i>Model</i> | <i>Accuracy</i> | <i>Macro F1</i> |
|---------------|-----------------|-----------------|
| Logistic Reg. | 0.76 | 0.75 |
| Random Forest | 0.98 | 0.98 |
| LightGBM | 0.98 | 0.98 |
| XGBoost | 0.98 | 0.98 |

Final chosen model: LightGBM, accuracy = 98.3%.

Feature Importance

Column-wise features are shown in detail in ReflectionSummary pdf in this rep

But still the most important was

CommitMessage :- TF-IDF was strongest (words like *test*, *UI*, *API*).

Fileextension := Grouping in (frontend,backend etc.) was predictive

Raw numeric data :- Made columns like churn,ave_file_added,avg_file_deleted,net_files and many more was useful

