Spaceship Titanic

PRZEMYSŁAW KWIECIŃSKI

Introduction

Spaceship Titanic is an interstellar passenger liner launched a month ago. With nearly 13,000 passengers on board, the ship embarked on its maiden voyage transporting emigrants from our solar system to three newly settled exoplanets orbiting nearby stars. The Titanic spacecraft collided with a space-time anomaly hidden in a cloud of dust. Although the ship remained intact, almost half of the passengers were transferred to an alternate dimension! Your task is to predict which passengers were transported by the anomaly, using records recovered from the ship's damaged computer system.

Goals

Get highest classification accuracy score calculated by percentage of predicted labels for transported passengers

Project Phases

DATA ANALYSIS

Analysis report from Pandas profiling, data statistics and vizualizations

DATA PREPROCESSING

Handle nulls and nan values. Replacing categorical values via encoding.

MODEL TRAINING

Train 3 models XGBoos, LightGBM, CatBoost

HYPERPARAMETERS TUNING

Using Optuna framework tune relevant models parameters

SUMMARY

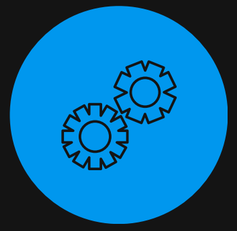
Choosing the best model and predict submission

Metrics

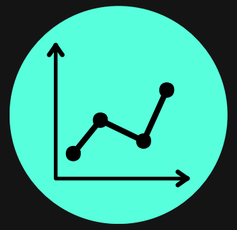
Accuracy

Recall

Precision



F1 score



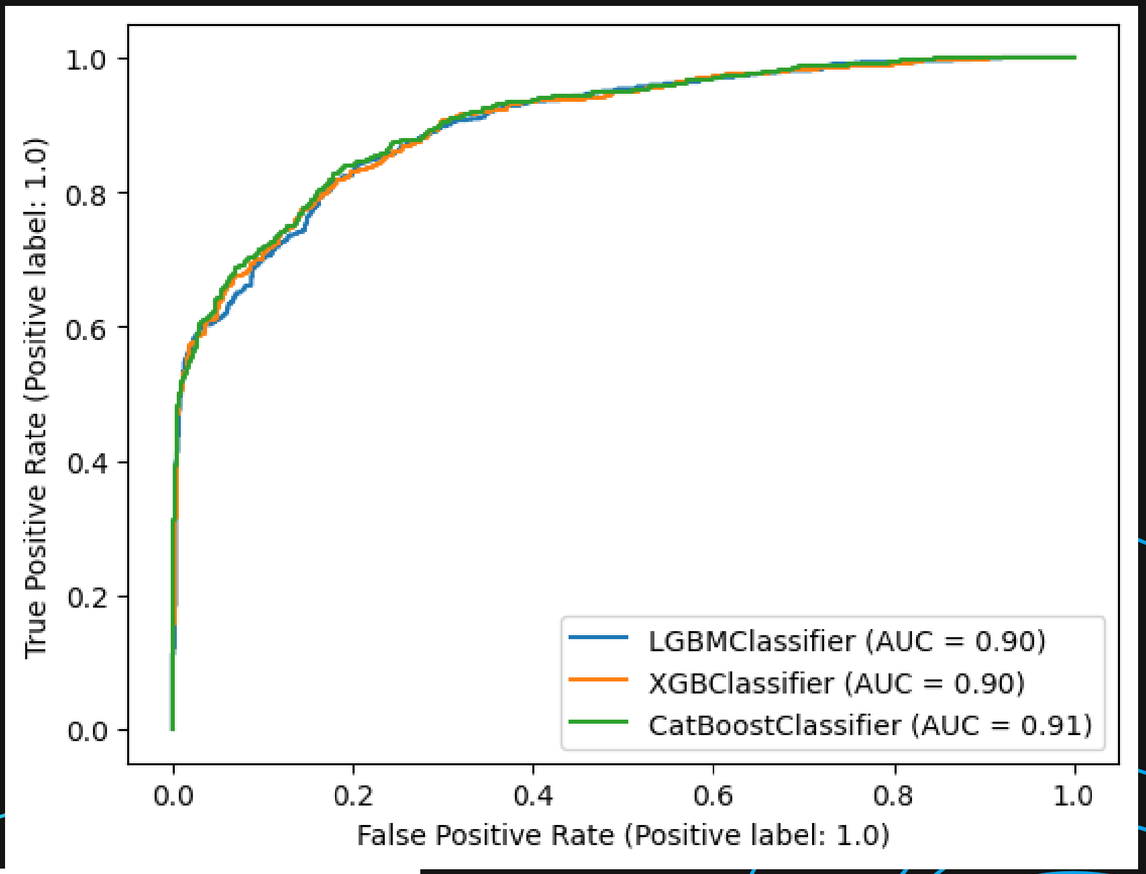
AUC

Models

	XGBOOST	LIGHTGBM	CATBOOST
ACCURACY	0.845066	0.826200	0.855365
PRECISION	0.823799	0.801927	0.821310
RECALL	0.814480	0.847285	0.837104
F1	0.819113	0.823982	0.829132
AUC	0.90	0.90	0.91

Results

CATBOOST HAD THE HIGHEST CV ACCURACY, WHICH SUGGESTS THAT IT IS MORE ROBUST TO VARIATIONS IN THE DATA. ADDITIONALLY, CATBOOST HAD A HIGH F1 SCORE, WHICH INDICATES A GOOD BALANCE BETWEEN PRECISION AND RECALL THAT IS WHY WILL BE USED TO SUBMISSION.

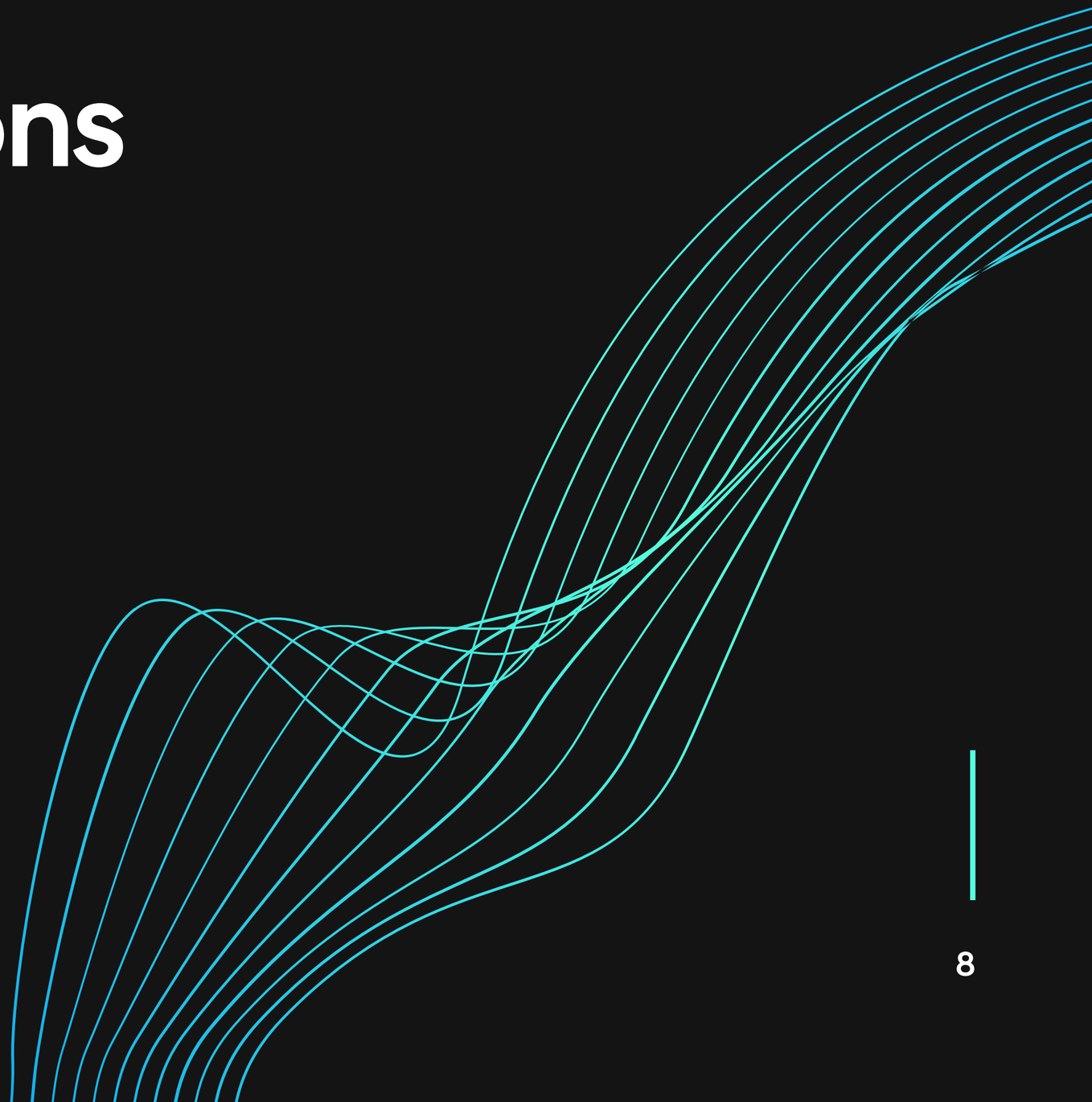


155	**G1llar**		0.80851

Your Best Entry!
Your most recent submission scored 0.80851, which is an improvement of your previous score of 0.80757. Great job!

Recommendations

There is still place for improve results by finding other correlations between features, selecting most important features we can do this using in example SelectKBest module from Sci-kit Learn. Good idea is also tune hyperparameters that I didn't used in my project yet.



Let's work together

EMAIL

przemyslaw.kwiecinski91@gmail.com

LINKEDIN

linkedin.com/in/przemysław-kwieciński-632212203

GITHUB

Przemek9110

KAGGLE

Przemyslaw Kwiecinski