Cloud Gaming

Ahmad Taha
a.taha@innopolis.university

1 Motivation

Nowadays cloud gaming is gaining more popularity, and it's becoming more common, but still face a lot of troubles, one of the major problems that cloud gaming face is unstable networks, We intend to optimize cloud gaming environments by detecting unstable networks, specially for people who do not have stable internet connection.

2 Data

We will deal with statistical data which come from one of the company that provide cloud gaming services. In Our dataset we have 5 main features, by applying statistical measures on those features we get nine features as predictor (independent features) and we have a target variable (dependent) which we want to predict (approximate) as for regression task The dataset was splitted in two files, train and test data, the structre of the train data is as follow:

Number of instances: 379021 rows
Number of features: 10 columns
The test set on the other hand as follow:
Number of instances: 228145 rows
Number of features: 10 columns

3 Exploratory data analysis

My main EDA started with the Regression file, as the classification task contains some of the same features we used in the regression, we shall not repeat what we have done as analysis in the classification file, I went deep in the EDA, starting with general overview then descriptive statistics, I analysed each feature on it's own in order to detect the distribution of this feature, how useful is it for predicting the target in regression task and in classification task as well, that gave me the best ideas of how should I clean and process the data in order to best prepare it for the modeling stage.

4 Task

AS for regression task: we will use machine learning models (Linear Regression and its variables) to predict the bit rate. AS for classification task: we will use machine learning models (Logistic Regression and its variables) to predict the stream quality based on our features.

4.1 Regression

As for Regression Models and Algorithms we used three models Simple Linear Regression based on Sickit-Learn Class LinearRegresion, Stochastic Gradient Linear Regression Based on the class SGDRegressor and finally Polynomial Linear Regression After generating some high order variables given our base features.

4.2 Classification

As for Classification Models and Algorithms we used two models Simple Logistic Regression based on Sickit-Learn Class LogisticRegresion with L2 Regularization, LogisticRegresion with L2 Regularization with balanced data.

5 Outlier Detection

The outlier detection process affected the result of our regression task more than classification task that because of the nature of regression metrics which yield in high error specially for outlier, we removed the outliers as one of the main steps of data cleaning.

6 Results

Let's see through tables the result of Regression model and Classification models on our test data.

Table 1. regression

Model	RMSE	Adjusted R2
Linear Regression	1959.248	0.892
SGDRegressor	1970.440	0.891
Polynomial LR(OverFitted)	106621.034	-317.860

Table 2. classification

Model	Acc.	Recall	Precision	F1-score
Ridge LR	0.950	0.941	0.128	0.226
Ridge LR (Data Balancing)	0.929	0.359	0.345	0.390

7 Data Imbalance

The target of the classification task was severely unbalanced, so we will notice much difference in term of results when we train our models on balanced dataset, rather than unbalanced dataset, that what model tend to keep predicting the major class which yield in high accuracy but poor performance on other metrics.

8 Conclusion

With out simple models we trained we have just scratched the surface of the problem of unstable networks, but still

detecting unstable networks in Cloud Gaming environments in advance would help a little in solving the problems, but as always we can go deeper to the heart of the problem by gathering more data and features and use more advanced algorithms in the future.