Cloud Gaming

Mohammad Sakka
m.sakka@innopolis.university

1 Motivation

The use of cloud services has spread significantly in recent times in order to reduce the computational burden on personal devices, and therefore many applications resorted to implementing on the cloud to save the computational cost and memory cost, and among these applications are game, Cloud gaming is a new way of online gaming, which renders the game data on the high-speed cloud server instead of the end user's system and is forwarded via a high-speed network. However, users do not always have ideal or stable internet for high streaming quality which leads to loss of data packets during data transfer. In this project we will use machine learning to detect unstable network which can be used to take steps towards adapting the data streaming and minimizing data packet loss. Our project consists of two parts, the first part is to predict the bit rate of internet connection, so we can predict the quality of connection, the next part is to classify if the stream quality data is good or bad, so we can decide doing some thing to improve network quality.

2 Data

We have dataset for each part, for the first part we have dataset and its features are some statistical information about the connection measurements like FPS and RRT. The second dataset also have sttistical measurements about the connection and the target of it is boolean value tell us if the quality good or bad.

3 Exploratory data analysis

we can take general view about the bitrate data using the following figure.

Overview | Alerts 20 | Reproduction

Dataset statistics

Number of variables	9
Number of observations	379021
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	530
Duplicate rows (%)	0.1%
Total size in memory	26.0 MiB
Average record size in memory	72.0 B

Variable types

Numeric	9

as we see, all feature are numerical and no missing values

and by the same way, we can take general view about stream quality data using the following figure.

Overview | Alerts 25 | Reproduction

Dataset statistics

Number of variables	11
Number of observations	406572
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	8959
Duplicate rows (%)	2.2%
Total size in memory	34.1 MiB
Average record size in memory	88.0 B

Variable types

Numeric	9
Categorical	2

as we see, There are 2 categorical features and no missing values.

4 Task

Our Task for bitrate data is to fit a function that takes the data features as inputs and outputs the pridcted bitrate. For stream quality data, Our Task is to create function that takes the data features as input and outputs 1 if the quality is good and 0 if the quality is bad.

4.1 Input Format

We used numerical encoding for the categorical features, for example, for Feature "auto fec state" we encoded "off" as 0 and "partial" as 0.5.

Bitrate dataset was scaled using robust scaler, Stream quality data was scales using standard scaler

Feature Selection for both two datasets using sickit-learn library, we selected the best 3 features. For bitrate dataset, chi2 score was used, and mutual information score was used for stream quality data.

Before fitting the models, we tried to remove the outlier using Z-score, by computing the Z-score for the rows of the data, then deleting the rows that have absolute Z-score value smaller than 3.

4.2 Regression

We used 3 algorithms for regression. The algorithms are the following: Linear Regression, Polynomial Linear Regression with Ridge Regularization, Neural Network.

For Polynomial Regression we used validation set to tune the value of regularization factor alpha, and made the degree of polonomial function 3.

For NN, we used one hidden layer with 20 neurons and number of iterations equals to 100

4.3 Classification

For classification problem, we used Logistic Regression and Logistic Regression with L2 regularization.

5 Results

5.1 Regression Results

The following tables illustrate the results of regression task, The underlined is the best.

Table 1. Regression Results On Training Data

Model	MAE	MSE	RMSE	R2
Linear-reg	1097	3903829	1976	0.894
Poly-linear	<u>1065</u>	<u>3847947</u>	<u>1962</u>	<u>0.896</u>
NN	1066	3864726	1966	0.895

Table 2. Regression Results On Testing Data

Model	MAE	MSE	RMSE	R2
Linear-reg	1077	3783467	1945	0.894
Poly-linear	<u>1049</u>	3761927	1940	<u>0.895</u>
NN	<u>1049</u>	<u>3757020</u>	<u>1938</u>	<u>0.895</u>

We notice that Poly-liner-reg is the best in training data, but NN is the best in testing, so we can consider NN is better for generalization.

We notice that the there is no significant difference between the training and testing results, so we can say that our models are not over fitted.

5.2 classification Results

The following tables illustrate the results of classification task, The underlined is the best.

Table 3. Classification Results On Training Data

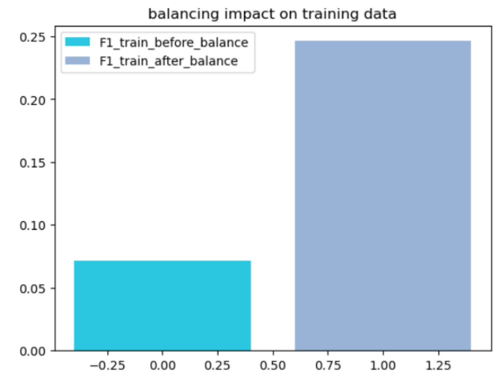
Model	Acc.	Precision	Recall	F1	F1-weighted
Log-reg	0.95	0.6	0.038	0.071	0.929
Log-reg-L2	0.95	0.6	0.038	0.071	0.929

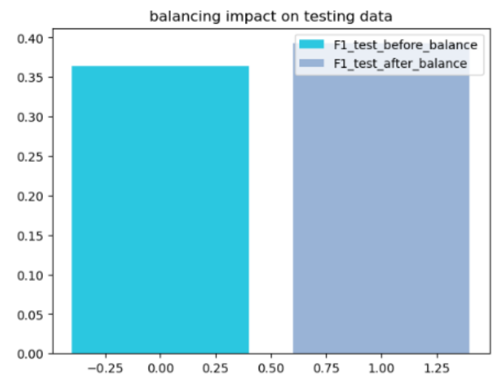
Table 4. Classification Results On Testing Data

Model	Acc.	Precision	Recall	F1	F1-weighted
Log-reg	0.929	0.437	0.311	<u>0.364</u>	0.924
Log-reg-L2	0.929	0.437	0.311	0.363	0.924

We notice that there is no significant difference between the Logistic regression model before and after regularization.

We also notice that F1 score in testing is better than training, but F1 weighted is better in training and that may be because the difference in number of records between training and testing data, so we can not decide which are the better results, but we can say that our model doesn't suffer from over or under fitting.





6 Data Imbalance

Over Sampling method has been implemented on the data to balance it, we implemented it from scratch by doubelecating the minority class rows by selecting random minor class rows each time.

Figures 1 and 2 shows the impact of data balancing

7 Conclusion

In this project we tried to solve the problem of poor connection quality in cloud gaming applications by predicting the bitrate and classify the quality of stream data.

We could achive R2 equals about 0.9 in bitrate prediction task, Neural Network was the superior in regression task.

For the classification task we achived F1 about 0.364 in testing data, we could improve it to reach about 0.4 after doing data balancing.