

Through this lab, I learned to make a machine learning model is select the eigenvalue that it uses for training first, then use historical data for training, then determined if the eigenvalue is reasonable base on the result of prediction. If not reasonable, the selected eigenvalues should be modified. If the prediction success rate is 75% or more, then this model is relatively successful. It can then be used to make formal predictions.

Write a query to determine available seasons and games

This step is to create a data sample from the database for our machine learning

Create a labeled machine learning dataset

This step is to purify the created data sample, after confirming which data should be used to determine whether the team will win or not, this step is to extract the data used to determine the result

Create a machine learning model

There are many different models in machine learning, and logistic regression is the most common one. In BigQuery ML, it only needs to select the machine learning model and then select the eigenvalues of the data.

Evaluate model performance and create table

First, historical records are used to make predictions to judge the accuracy of the machine learning model. When the accuracy is not ideal, we should consider how to obtain higher accuracy.

Using skillful ML model features

After the previous prediction results, it realized that some eigenvalues are not very reasonable, such as seed ranking and team name. Therefore, it should be found to make the eigenvalues more complex to improve the accuracy of the prediction.

Train the new model and make evaluation

After retraining the model, the accuracy of prediction is improved, indicating that the selection of eigenvalues is more scientific than the original

Run a query to create a table ncaa_2018_predictions

Using the trained model to predict with new data, it was found that the newly trained model successfully predicted the upset. This shows that the newly trained model is a relatively successful prediction model. Problems that lead to prediction failures were fixed by using new eigenvalue.

Run queries to create tables ncaa_2019_tournament and ncaa_2019_tournament_predictions

After a series of tests above, the prediction model has been a successful model, and now it can be used to predict real matches from the future.

End Lab00:07:10

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked.
[Learn more.](#)

Open Google Console

Username
student-04-827f5a933cc8

Password
LBf5VoOP1sT#

GCP Project ID
qwiklabs-gcp-01-3ccad20e

Bracketology with Google Machine Learning

ORDER BY seed, opponent_seed

Query complete (0.7 sec elapsed, 289.7 KB processed)

Job informationResultsJSONExecution details

Row	prediction	confidence	school_ncaa	seed	opponent_school_ncaa	opponent_seed
1	win	0.605	Duke	1	North Carolina	1
2	win	0.515	Duke	1	Gonzaga	1
3	loss	0.694	Duke	1	Virginia	1
4	loss	0.585	Duke	1	Michigan	2
5	loss	0.542	Duke	1	Michigan St.	2
6	loss	0.53	Duke	1	Tennessee	2
7	loss	0.508	Duke	1	Kentucky	2
8	loss	0.554	Duke	1	Purdue	3
9	loss	0.517	Duke	1	Houston	3
10	loss	0.522	Duke	1	Texas Tech	3
11	win	0.619	Duke	1	LSU	3
12	loss	0.588	Duke	1	Virginia Tech	4
13	win	0.626	Duke	1	Florida St.	4
14	win	0.537	Duke	1	Kansas St.	4

Here we filtered the model results to see all of Duke's possible games. Scroll to find the Duke vs North Dakota St. game.

Insight: Duke (1) is 88.5% favored to beat North Dakota St. (16) on 3/22/19.

Experiment by changing the school_ncaa filter above to predict for the matchups in your bracket. Write down what the model confidence is and enjoy the games!

GSP461

Overview

Setup and requirements

Open the BigQuery Console

NCAA March Madness

Find the NCAA public dataset in BigQuery

Write a query to determine available seasons and games

Understand machine learning features and labels

Create a labeled machine learning dataset

Part 1: Create a machine learning model to predict the winner based on seed and team name

Evaluate model performance

Making predictions

How many did our model get right for the 2018 NCAA tournament?

Models can only take you so far...

Part 2: Using skillful ML model features

Preview the new features

Interpreting selected metrics

Train the new model

Evaluate the new model's performance