**Using BigQuery in ML for NCAA Basketball**

Here we are creating a BigQuery ML model for NCAA basketball using the public dataset. We are providing our query details and output screenshots.

First the input features of the model is queried using mbb\_teams\_games\_sr table in the NCAA Basketball public dataset. The input features include the mean and standard deviation of previous game statistics for both home teams and away teams using different time windows. The time windows used are 10, 5 and 1 game before the current game. The team\_id and season columns are used for one-hot features. After generating the input features, we will generate training data.

After generating, we will create a linear regression model. The model is used to predict the combined three point field goal attempts based on the previous game statistics. Below is a bigquery to create a linear regression model using three point goal attempts related features.

CREATE OR REPLACE MODEL bqml\_tutorial.ncaa\_model OPTIONS(

model\_type='linear\_reg', max\_iteration=50) AS

SELECT

\* EXCEPT(

game\_id, season, scheduled\_date,

total\_three\_points\_made,

total\_three\_points\_att),

total\_three\_points\_att as label

FROM

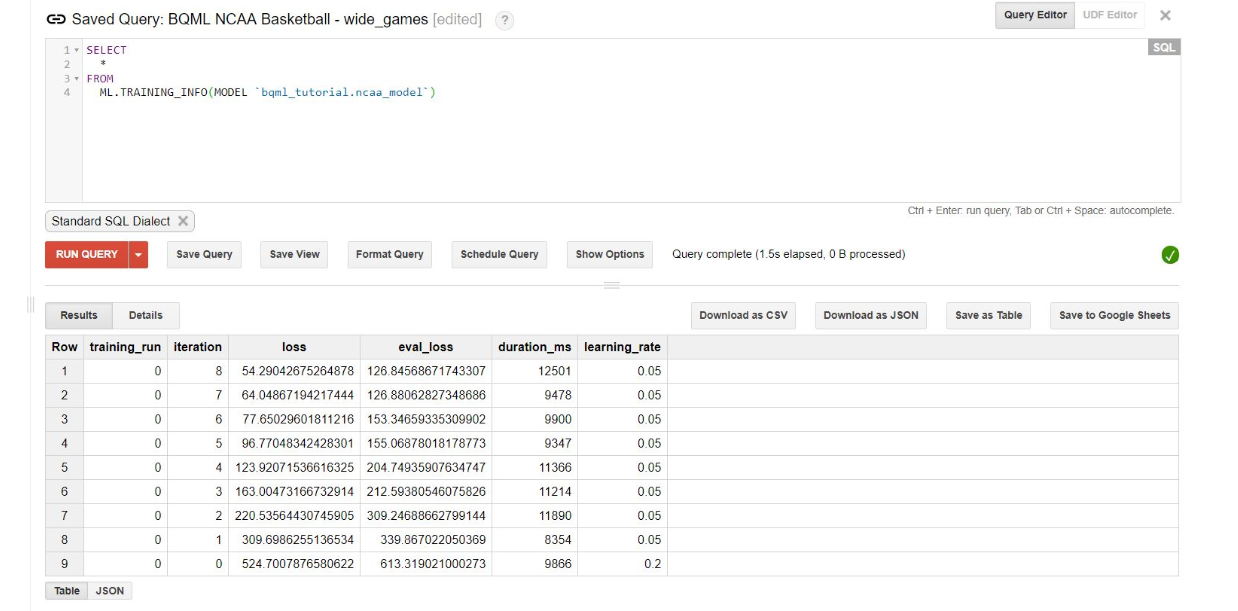
bqml\_tutorial.wide\_games

WHERE

# remove the game to predict

game\_id != 'f1063e80-23c7-486b-9a5e-faa52beb2d83'

On querying it, we get the model:



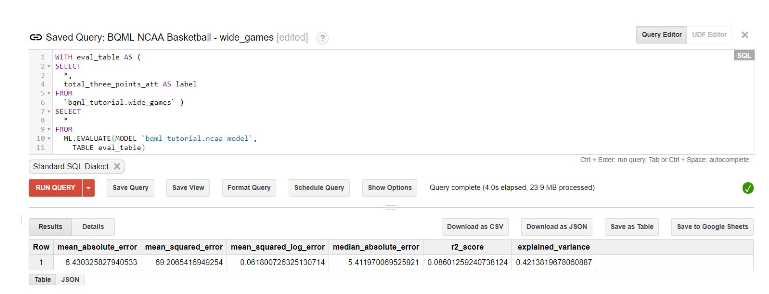
After creating the model, we will evaluate the performance of the model using the ML.EVALUATE function. Below is the query for the same.

WITH eval\_table AS (

SELECT \*, total\_three\_points\_att AS label FROM `bqml\_tutorial.wide\_games` )

SELECT \* FROM ML.EVALUATE(MODEL `bqml\_tutorial.ncaa\_model`, TABLE eval\_table)

Below is the output we will get after executing the above query.



Now that we have evaluated the model, the next step is to use the ML.PREDICT function to predict the total three point field goal attempts in the 2018 NCAA final game: Michigan versus Villanova.

Below is the query for the same.

WITH game\_to\_predict AS (

SELECT \* FROM `bqml\_tutorial.wide\_games` WHERE game\_id='f1063e80-23c7-486b-9a5e-faa52beb2d83' )

SELECT truth.game\_id AS game\_id, total\_three\_points\_att, predicted\_total\_three\_points\_att

FROM (

SELECT game\_id, predicted\_label AS predicted\_total\_three\_points\_att

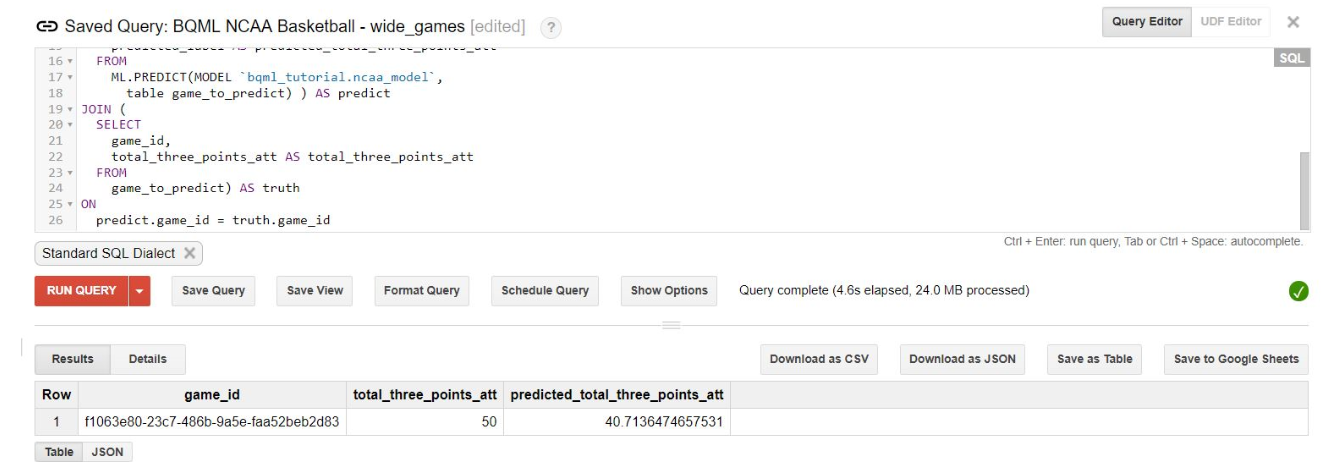
FROM ML.PREDICT(MODEL `bqml\_tutorial.ncaa\_model`, table game\_to\_predict) ) AS predict

JOIN (

SELECT game\_id, total\_three\_points\_att AS total\_three\_points\_att

FROM game\_to\_predict) AS truth ON predict.game\_id = truth.game\_id

Below is the output we will get after executing the above query.



The total\_three\_points\_att value is the actual number of field goals that occurred in the final game – 50. The model’s prediction is 43.41.