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.

Table

Description automatically generated