--Building a regression model to predict survival on the Titanic using SQL

-- query from table

SELECT \*

FROM `project-binh-2023-demo.DA04.train\_titanic`

LIMIT 10;

--create logistic regression model

CREATE OR REPLACE MODEL `DA04.titanic\_logistic\_model`

OPTIONS( model\_type = 'LOGISTIC\_REG',

auto\_class\_weights= TRUE,

input\_label\_cols= ['Survived']

) AS

SELECT

Pclass, Sex, Age, SibSp, Fare, Embarked, Survived

FROM

`project-binh-2023-demo.DA04.train\_titanic`;

-----------

SELECT

\*

FROM

ML.EVALUATE( MODEL `DA04.titanic\_logistic\_model`);

SELECT

\*

FROM

ML.CONFUSION\_MATRIX(MODEL `DA04.titanic\_logistic\_model`);

SELECT

\*

FROM

ML.ROC\_CURVE(MODEL `DA04.titanic\_logistic\_model`);

--OR #standardSQL

SELECT

\*

FROM

ML.EVALUATE( `DA04.titanic\_logistic\_model`,(

SELECT Survived,Pclass, Sex, Age, SibSp, Fare, Embarked,Parch

FROM `project-binh-2023-demo.DA04.train\_titanic`

));

-- create new predict for test\_data

SELECT

\*

FROM ML.PREDICT(MODEL `DA04.titanic\_logistic\_model`,

(

SELECT Pclass, Sex, Age, SibSp, Fare, Embarked,Parch

FROM `project-binh-2023-demo.DA04.test\_titanic`

WHERE Age > 0 and Fare >0

))