## CREATE TRAINING AND TESTING DATASETS

CREATE TABLE `rich-honor-327502.dsc680.test\_data` AS

SELECT \*

FROM `rich-honor-327502.dsc680.bq\_credit\_dataset`

WHERE MOD(ABS(FARM\_FINGERPRINT(CAST(uniqueID AS STRING))), 5) = 0;

CREATE OR REPLACE TABLE `rich-honor-327502.dsc680.train\_data` AS

SELECT \*

FROM `rich-honor-327502.dsc680.bq\_credit\_dataset`

WHERE NOT uniqueID IN (

  SELECT DISTINCT uniqueID FROM `rich-honor-327502.dsc680.test\_data`

);

## CHECK DEFAULT RATE FOR OVERALL, TEST AND TRAINING DATASET TO MAKE SURE DATA IS

## EVENLY DISTRIBUTED

SELECT count (distinct uniqueID)

FROM `rich-honor-327502.dsc680.bq\_credit\_dataset`

where loan\_status = 0;

##ALL

#loan status 1 = 6825

#loan status 0 = 24854

#21.5% default

#TRAIN

#loan status 1 = 5461

#loan status 0 = 19854

#5461/(19854+5461) = 21.5%

##TEST

#loan status 1 = 1364

#loan status 0 = 5000

#1364/6364 = 21.4%

## STEP 3 – CREATE TRAINING MODEL

# Create the base model

CREATE OR REPLACE MODEL `rich-honor-327502.dsc680.bq\_credit\_base\_model`

OPTIONS(input\_label\_cols=['loan\_status'], model\_type='logistic\_reg') AS

SELECT \* EXCEPT (uniqueID)

FROM `rich-honor-327502.dsc680.train\_data`;

Chart, bar chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart

Description automatically generated with medium confidence

##STEP 4

# Evaluate Model fit on train data

SELECT \*

FROM ML.EVALUATE(MODEL `rich-honor-327502.dsc680.bq\_credit\_base\_model`,

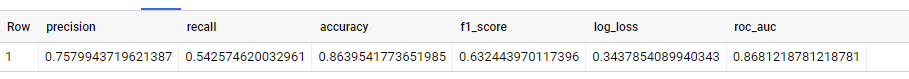
(

  SELECT \* EXCEPT (uniqueID)

  FROM `rich-honor-327502.dsc680.train\_data`

  )

)



##STEP 5

# Evaluate Model fit on test data

SELECT \*

FROM ML.EVALUATE(MODEL `rich-honor-327502.dsc680.bq\_credit\_base\_model`,

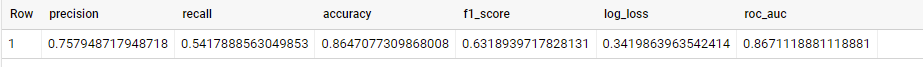
(

  SELECT \* EXCEPT (uniqueID)

  FROM `rich-honor-327502.dsc680.test\_data`

  )

)



##STEP 6

# To view your linear beta-values

SELECT \* from ML.WEIGHTS(MODEL `rich-honor-327502.dsc680.bq\_credit\_base\_model`);

#STEP 7

# To get test data confusion matrix

SELECT \*

FROM ML.CONFUSION\_MATRIX(MODEL `rich-honor-327502.dsc680.bq\_credit\_base\_model`,

(

  SELECT\* EXCEPT (uniqueID)

  FROM `rich-honor-327502.dsc680.test\_data`  )

);

Graphical user interface, text, application, table

Description automatically generated