Auto ML

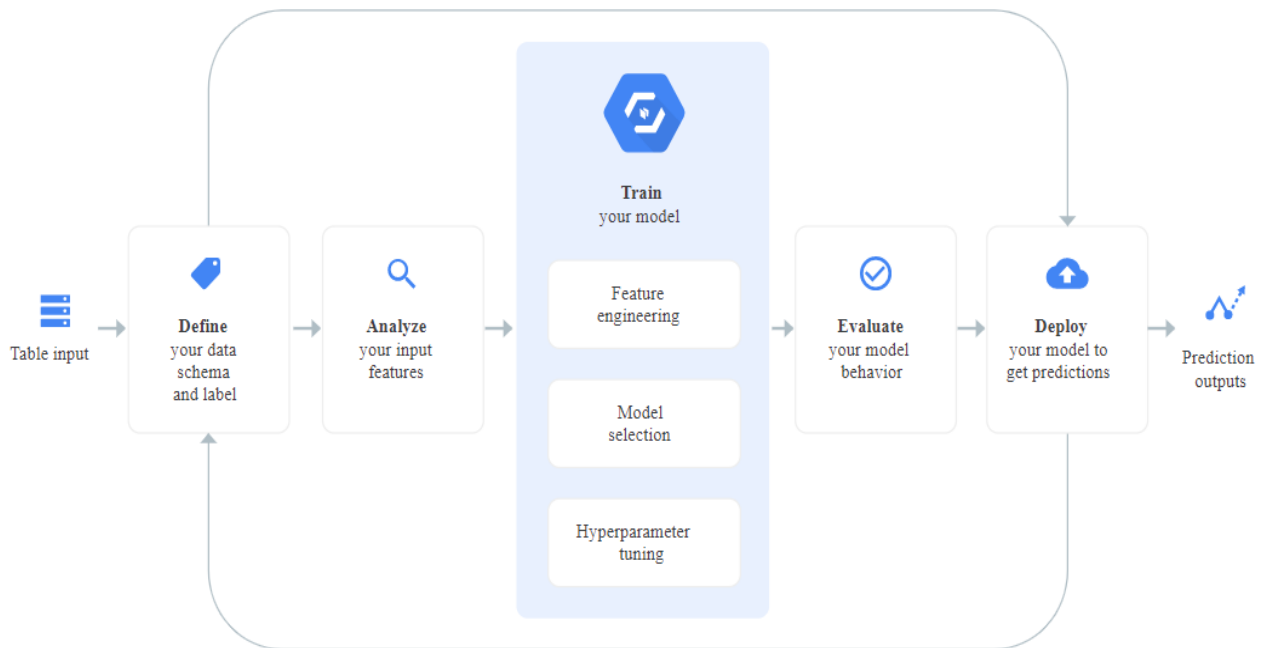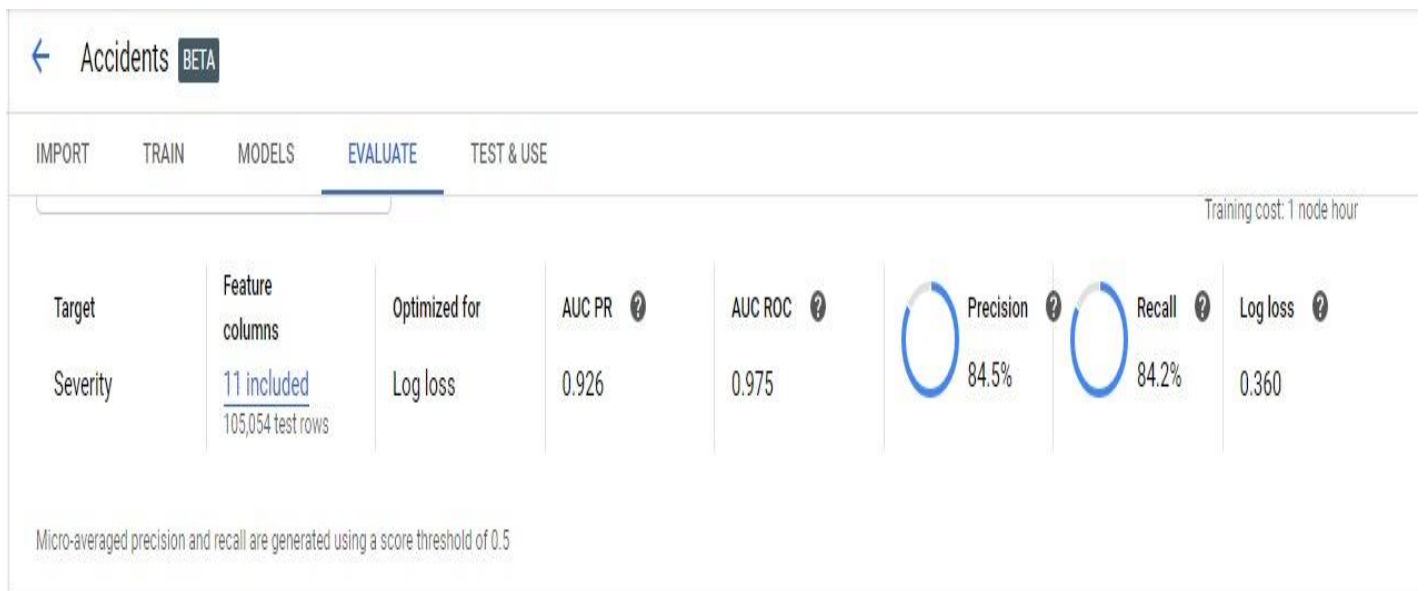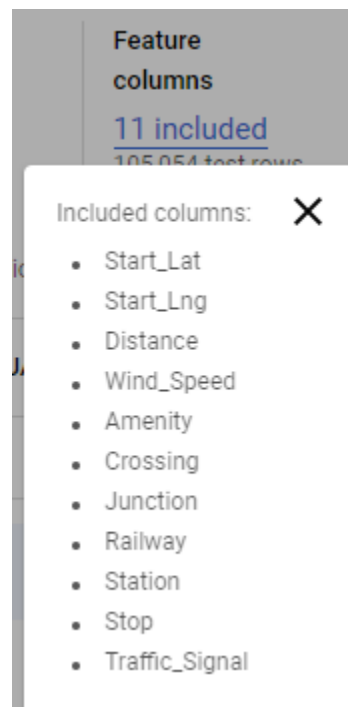## HOW AUTOML WORKS:

MODELLING RESULTS:

Below features are included for training our model, these features are chosen as per their correlation with the target variable **" Severity"**

Feature columns

11 included

105.054 test rows

Included columns:

- Start_Lat
- Start_Lng
- Distance
- Wind_Speed
- Amenity
- Crossing
- Junction
- Railway
- Station
- Stop
- Traffic_Signal

MODEL USED IN AUTO ML: GRADIENT BOOSTING DECISION TREE ALGORITHM

Gradient boosting algorithm sequentially blends weak learners so that each new learner matches the residuals of the previous step in order to improve the model. The final model aggregates the outcomes from each move and reaches a strong learner. Gradient-boosted algorithm for decision trees uses decision trees as week-learners. The residuals are identified via a loss function. For example, *for a regression task, mean squared error (MSE) can be used, and logarithmic loss (log loss)* can be used for classification tasks. It should be noted that the current trees in the model do not change with the introduction of a new tree

Hyperparemets are key components of learning algorithms that affect the performance and accuracy of the model. Learning rate and n estimators are two important hyperparameters for gradient boosting decision trees in gradients. Learning rate, called α, simply means how quickly the model learns. Every added tree modifies the model overall. Learning rate regulates the extent of the modification.

LOGS VIEWER:

```
{
  "insertId": "w72sc0g4m7lvs5",
  "jsonPayload": {
    "@type": "type.googleapis.com/google.cloud.automl.master.TablesModelStructure",
    "modelParameters": [
      {
        "hyperparameters": {
          "Max tree depth": 9,
          "Tree L2 regularization": 3,
          "Tree L1 regularization": 3,
          "Center Bias": "False",
          "Tree complexity": 1,
          "Model type": "GBDT",
          "Number of trees": 150
        }
      },
      {
        "hyperparameters": {
          "Number of trees": 150,
          "Max tree depth": 9,
          "Center Bias": "False",
          "Tree L2 regularization": 3,
          "Tree L1 regularization": 3,
          "Tree complexity": 1,
          "Model type": "GBDT"
        }
      },
      {
        "hyperparameters": {
          "Number of trees": 300,
          "Max tree depth": 12,
          "Tree L2 regularization": 3,
          "Tree L1 regularization": 1,
          "Center Bias": "False",
          "Tree complexity": 3,
          "Model type": "GBDT"
        }
      }
```

RESULTS IN BIGQUERY

ONLINE PREDICTION:

We Used **Charleston Boulevard and Lamb Boulevard(**36.159629° -115.080135 **)** in las vegas for online prediction, given that there is a stop, junction and presence of traffic signals, distance affected 10 mph, we obtained a ***severity level prediction of 2*** which is justified. As there is a chance traffic can be affected for several distance we can also see that Severity of 3 has 41% probability.
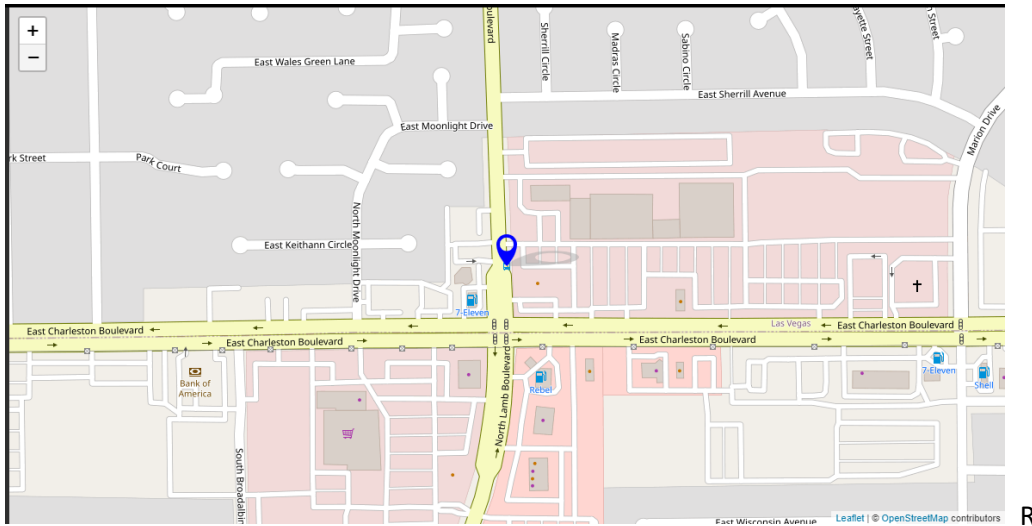
Predict label

Severity

Prediction result

Baseline prediction value: 0.716

2
Confidence score: 0.58

3
Confidence score: 0.411

4
Confidence score: 0.009

1
Confidence score: 0.001