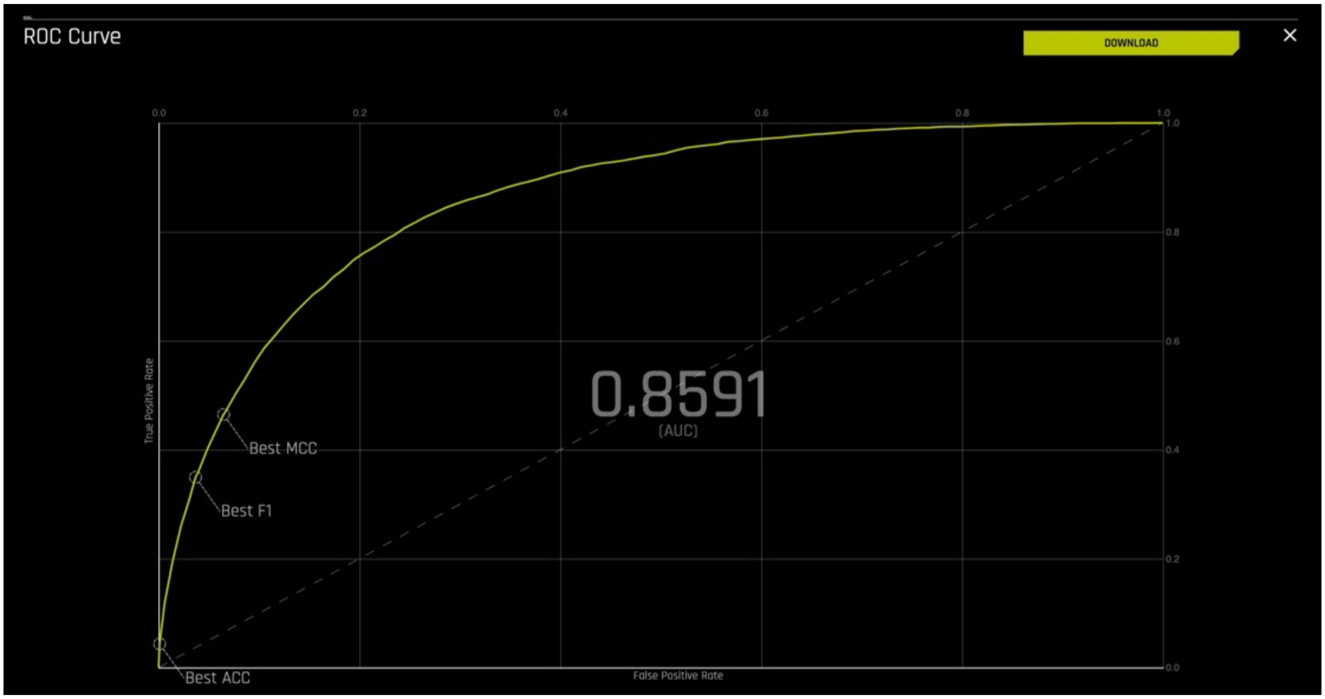
机器学习实验评分和分析教程-财务重点

h2oai.github.io/tutorials/machine-learning-experiment-scoring-and-analysis-tutorial-financial-focus

8.任务6：ER：ROC

在“诊断”页面上，单击**ROC曲线**。将会出现类似于以下图像的图像：



回顾一下，ROC曲线显示以下内容：

- 它显示了灵敏度（真阳性率或TPR）和特异性（1-FPR或假阳性率）之间的权衡。敏感性的任何提高都将伴随特异性的降低。
- 曲线越接近ROC空间的左边界和上边界，模型越精确。
- 曲线越接近ROC空间的45度对角线，模型的精度就越差。
- 切点处切线的斜率给出了该测试值的似然比（LR）。您可以在上方的图表中查看。
- 曲线下的面积是模型准确性的度量。

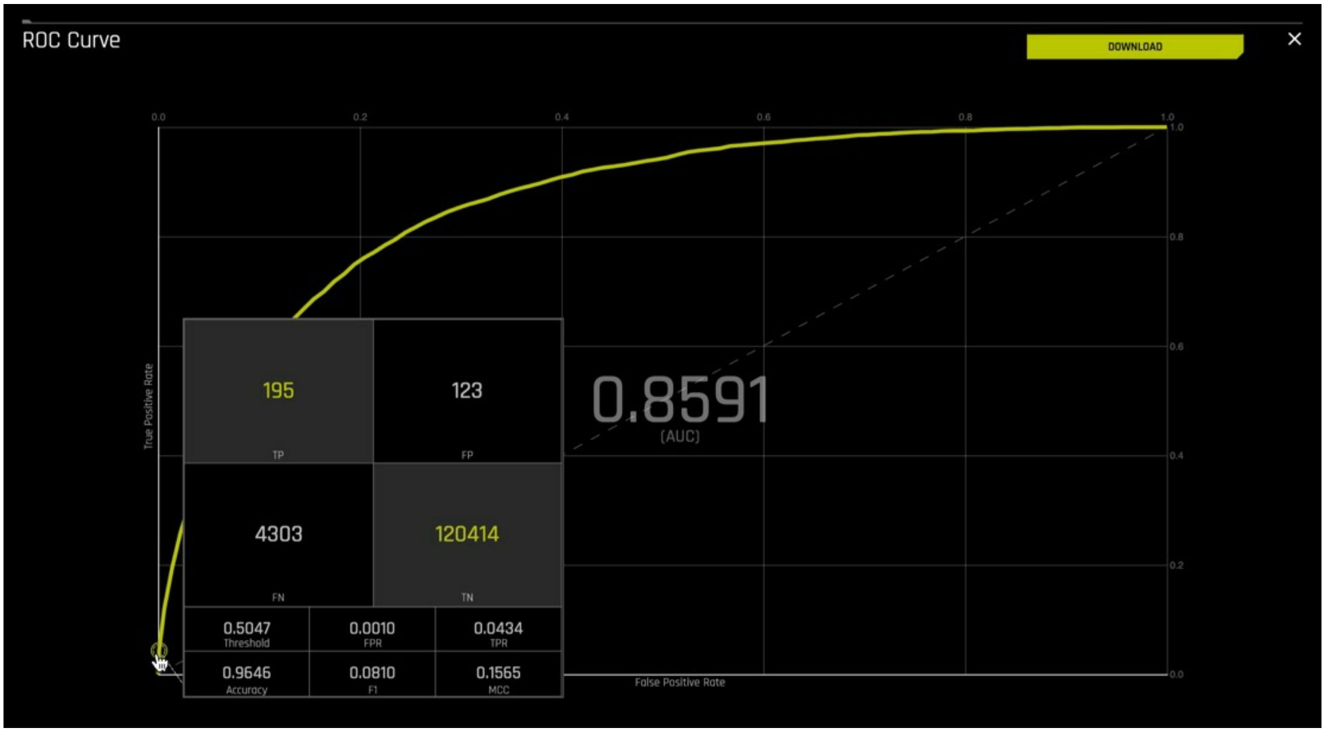
回到Freddie Mac数据集，即使该模型使用对数据损失进行评分以补偿误差，我们仍然可以查看ROC曲线结果，并查看其是否支持我们从混淆矩阵和分数部分的分析得出的结论。诊断页面。

1.根据无人驾驶AI模型为实验生成的ROC曲线，确定AUC。回想一下，理想的分类模型的AUC为1。

2.对于曲线上的以下每个点，将鼠标悬停在下面的每个点上，即可确定正阳性率，假阳性率和阈值，如下图所示：

- 最佳精度

- 最佳F1
- 最佳我的客户中心



195 TP	123 FP
4303 FN	120414 TN

0.5047 Threshold	0.0010 FPR	0.0434 TPR
0.9646 Accuracy	0.0810 F1	0.1565 MCC

回想一下，对于二进制分类问题，准确性是做出的正确预测的数量与所做的所有预测的比率。概率将转换为预测的类别，以定义阈值。对于此模型，已确定在阈值.5098处发现最佳精度。

在此阈值下，模型预测：

- TP = 1 = 195个案例预计为违约和违约
- TN = 0 = 120,414个案例被预测为未违约且未违约
- FP = 1 = 123个案例被预测为违约且未违约
- FN = 0 = 4,303个案例，预计没有违约和被违约

3. From the AUC, Best MCC, F1, and Accuracy values from the ROC curve, how would you qualify your model, is it a good or bad model? Use the key points below to help you asses the ROC Curve.

Remember that for the **ROC** curve:

- The perfect classification model has an AUC of 1
- MCC is measured in the range between -1 and +1 where +1 is the perfect prediction, 0 no better than a random prediction and -1 all incorrect predictions.
- F1 is measured in the range of 0 to 1, where 0 means that there are no true positives, and 1 when there is neither false negatives nor false positives or perfect precision and recall.

- Accuracy is measured in the range of 0 to 1, where 1 is perfect accuracy or perfect classification, and 0 is poor accuracy or poor classification.

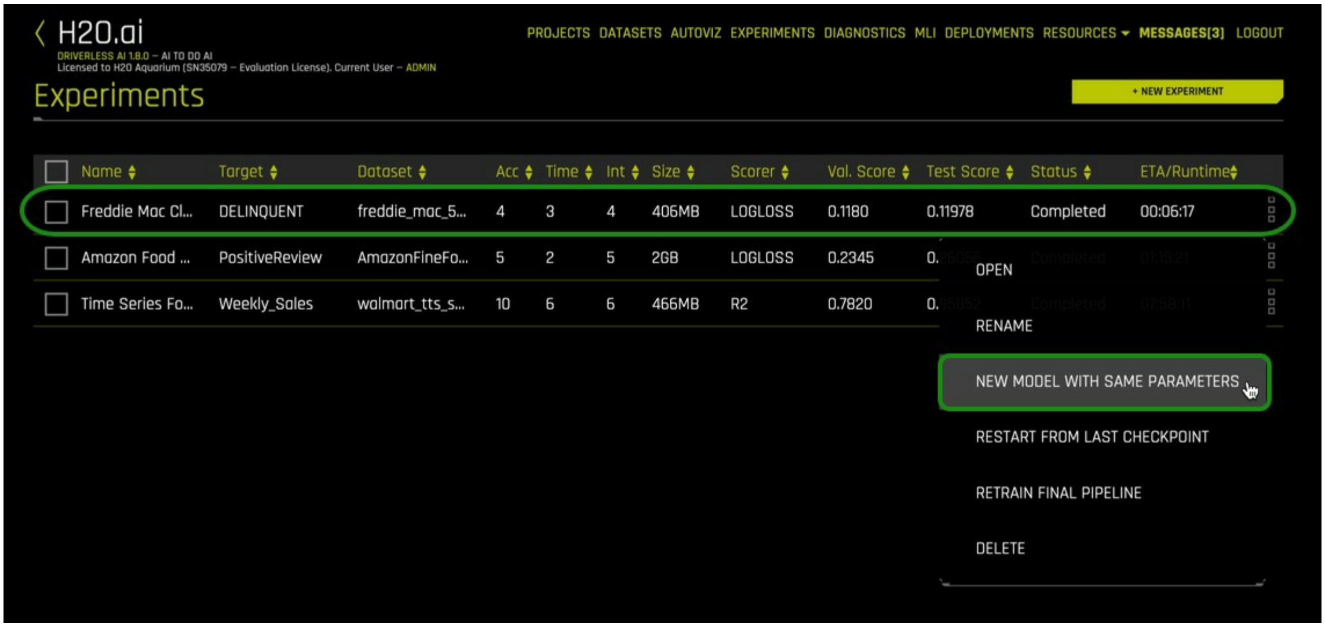
Note: If you are not sure what AUC, MCC, F1, and Accuracy are or how they are calculated review the concepts section of this tutorial.

New Model with Same Parameters

In case you were curious and wanted to know if you could improve the accuracy of the model, this can be done by changing the scorer from Logloss to Accuracy.

1. To do this, click on the **Experiments** page.

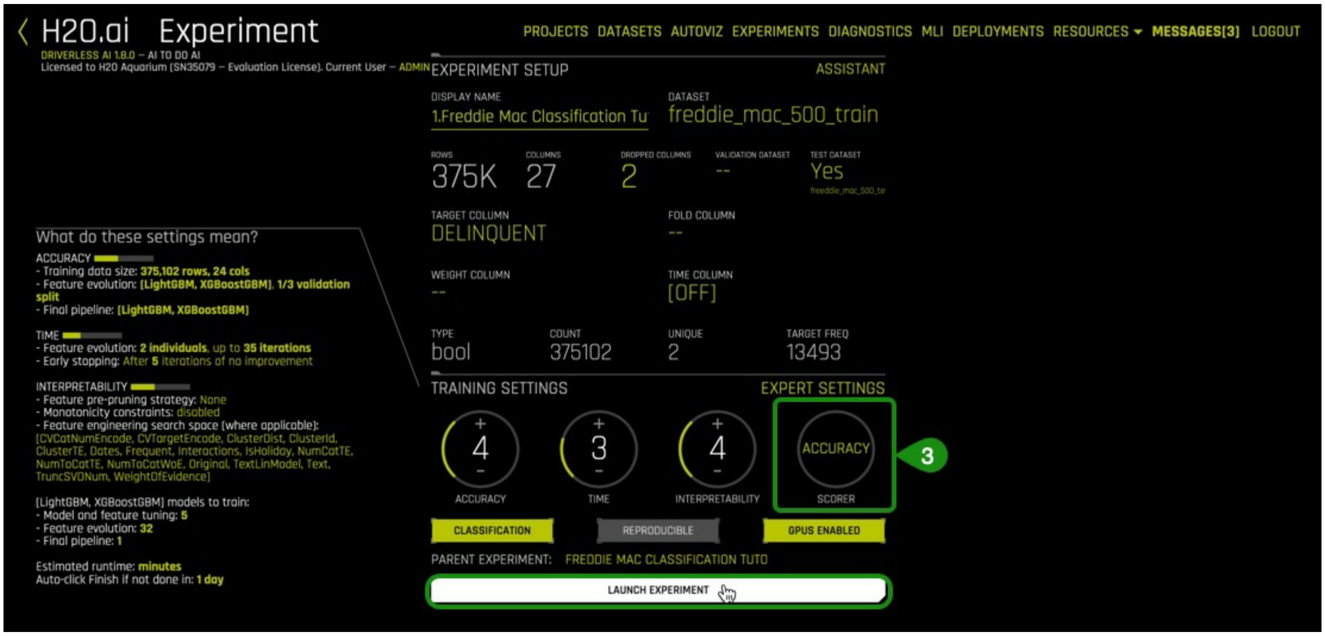
2. Click on the experiment you did for task 1 and select **New Model With Same Params**


H2O.ai
DRIVERLESS AI 1.8.0 – AI TO DO AI
Licensed to H2O Aquarium (SN35079 – Evaluation License). Current User – ADMIN
PROJECTS DATASETS AUTOVIZ EXPERIMENTS DIAGNOSTICS MLI DEPLOYMENTS RESOURCES MESSAGES[3] LOGOUT
Experiments
+ NEW EXPERIMENT
OPEN
RENAME
NEW MODEL WITH SAME PARAMETERS
RESTART FROM LAST CHECKPOINT
RETRAIN FINAL PIPELINE
DELETE

□	Name	Target	Dataset	Acc	Time	Int	Size	Scorer	Val. Score	Test Score	Status	ETA/Runtime
□	Freddie Mac Cl...	DELINQUENT	freddie_mac_5...	4	3	4	406MB	LOGLOSS	0.1180	0.11978	Completed	00:06:17
□	Amazon Food ...	PositiveReview	AmazonFineFo...	5	2	5	2GB	LOGLOSS	0.2345	0.26056	Completed	01:15:21
□	Time Series Fo...	Weekly_Sales	walmart_tts_s...	10	6	6	466MB	R2	0.7820	0.95852	Completed	07:58:11

An image similar to the one below will appear. Note that this page has the same settings as the setting in Task 1. The only difference is that on the **Scorer** section **Logloss** was updated to **Accuracy**. Everything else should remain the same.

3. If you haven't done so, select **Accuracy** on the scorer section then select **Launch Experiment**



DRIVERLESS AI 1.8.0 – AI TO DO AI
Licensed to H2O Aquarium (SN35079 – Evaluation License). Current User – ADMIN

EXPERIMENT SETUP

DISPLAY NAME: 1.Freddie Mac Classification Tu
DATASET: freddie_mac_500_train

ROWS: 375K
COLUMNS: 27
DROPPED COLUMNS: 2
VALIDATION DATASET: --
TEST DATASET: Yes (freeddie_mac_500_te)

TARGET COLUMN: DELINQUENT
FOLD COLUMN: --
WEIGHT COLUMN: --
TIME COLUMN: [OFF]

TYPE: bool
COUNT: 375102
UNIQUE: 2
TARGET FREQ: 13493

TRAINING SETTINGS

ACCURACY: 4
TIME: 3
INTERPRETABILITY: 4
SCORER: ACCURACY

PARENT EXPERIMENT: FREDDIE MAC CLASSIFICATION TUTO

What do these settings mean?

ACCURACY
- Training data size: 375,102 rows, 24 cols
- Feature evolution: [LightGBM, XGBoostGBM], 1/3 validation split
- Final pipeline: [LightGBM, XGBoostGBM]

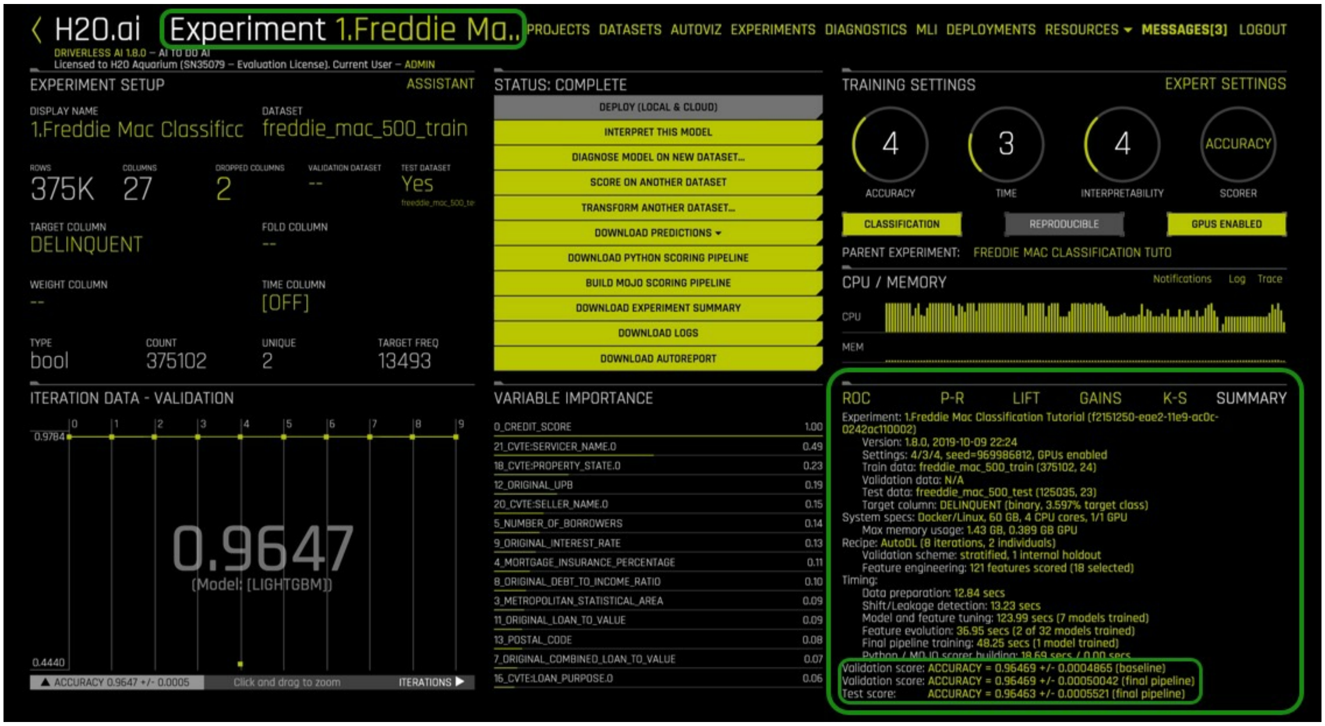
TIME
- Feature evolution: 2 individuals, up to 35 iterations
- Early stopping: After 5 iterations of no improvement

INTERPRETABILITY
- Feature pre-pruning strategy: None
- Monotonicity constraints: disabled
- Feature engineering search space (where applicable): [CVCatNumEncode, CVTargetEncode, ClusterDist, ClusterId, ClusterTE, Dates, Frequent, Interactions, IsHoliday, NumCatTE, NumToCatTE, NumToCatWoE, Original, TextLinModel, Text, TruncSVDNum, WeightOfEvidence]

[LightGBM, XGBoostGBM] models to train:
- Model and feature tuning: 5
- Feature evolution: 32
- Final pipeline: 1

Estimated runtime: minutes
Auto-click Finish if not done in: 1 day

Similarly to the experiment in Task 1, wait for the experiment to run. After the experiment is done running, a similar page will appear. Note that on the summary located on the bottom right-side both the validation and test scores are no longer being scored by **Logloss** instead by **Accuracy**.



DRIVERLESS AI 1.8.0 – AI TO DO AI
Licensed to H2O Aquarium (SN35079 – Evaluation License). Current User – ADMIN

EXPERIMENT SETUP

DISPLAY NAME: 1.Freddie Mac Classificc
DATASET: freddie_mac_500_train

ROWS: 375K
COLUMNS: 27
DROPPED COLUMNS: 2
VALIDATION DATASET: --
TEST DATASET: Yes (freeddie_mac_500_te)

TARGET COLUMN: DELINQUENT
FOLD COLUMN: --
WEIGHT COLUMN: --
TIME COLUMN: [OFF]

TYPE: bool
COUNT: 375102
UNIQUE: 2
TARGET FREQ: 13493

TRAINING SETTINGS

ACCURACY: 4
TIME: 3
INTERPRETABILITY: 4
SCORER: ACCURACY

PARENT EXPERIMENT: FREDDIE MAC CLASSIFICATION TUTO

VARIABLE IMPORTANCE

0_CREDIT_SCORE	1.00
21_CVTE:SERVICER_NAME.0	0.49
18_CVTE:PROPERTY_STATE.0	0.23
12_ORIGINAL_UPB	0.19
20_CVTE:SELLER_NAME.0	0.15
5_NUMBER_OF_BORROWERS	0.14
9_ORIGINAL_INTEREST_RATE	0.13
4_MORTGAGE_INSURANCE_PERCENTAGE	0.11
8_ORIGINAL_DEBT_TO_INCOME_RATIO	0.10
3_METROPOLITAN_STATISTICAL_AREA	0.09
11_ORIGINAL_LOAN_TO_VALUE	0.09
13_POSTAL_CODE	0.08
7_ORIGINAL_COMBINED_LOAN_TO_VALUE	0.07
16_CVTE:LOAN_PURPOSE.0	0.06

```
Experiment: 1.Freddie Mac Classification Tutorial (f2151250-eae2-11e9-ac0c-
0242ac110002)
    Version: 1.8.0, 2019-10-09 22:24
    Settings: 4/3/4, seed=969986812, GPUs enabled
    Train data: freddie_mac_500_train (375102, 24)
    Validation data: N/A
    Test data: freeddie_mac_500_test (125035, 23)
    Target column: DELINQUENT (binary, 3.597% target class)
System specs: Docker/Linux, 60 GB, 4 CPU cores, 1/1 GPU
    Max memory usage: 1.43 GB, 0.389 GB GPU
Recipe: AutoDL (8 iterations, 2 individuals)
    Validation scheme: stratified, 1 internal holdout
    Feature engineering: 121 features scored (18 selected)
Timing:
    Data preparation: 12.84 secs
    Shift/Leakage detection: 13.23 secs
    Model and feature tuning: 123.99 secs (7 models trained)
    Feature evolution: 36.95 secs (2 of 32 models trained)
    Final pipeline training: 48.25 secs (1 model trained)
    Python / MOJO scorer building: 18.69 secs / 0.00 secs
Validation score: ACCURACY = 0.96469 +/- 0.0004865 (baseline)
Validation score: ACCURACY = 0.96469 +/- 0.00050042 (final pipeline)
Test score:       ACCURACY = 0.96463 +/- 0.0005521 (final pipeline)
```

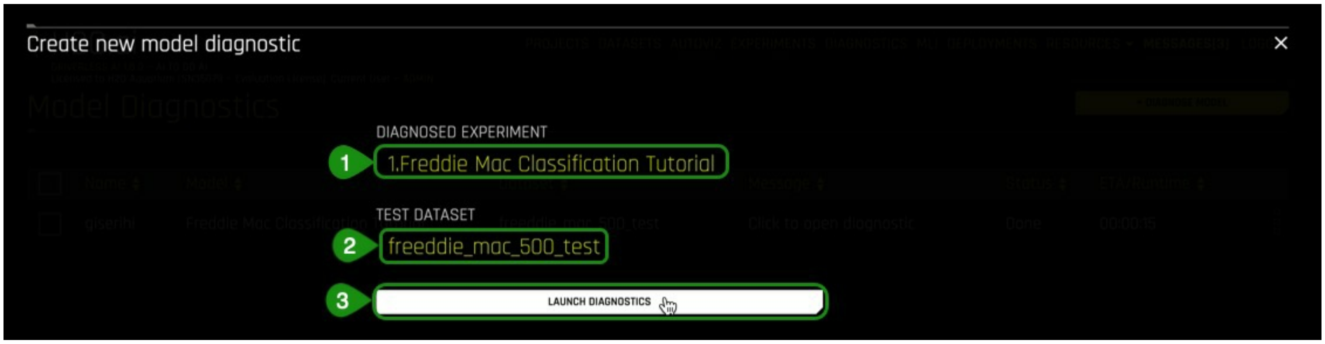
We are going to use this new experiment to run a new diagnostics test. You will need the name of the new experiment. In this case, the experiment name is **1.Freddie Mac Classification Tutorial**.

4. Go to the **Diagnostics** tab.

5. Once in the **Diagnostics** page, select **+Diagnose Model**

6. In the **Create new model diagnostics** :

1. Click on Diagnosed Experiment then select the experiment that you completed in Task in this case the experiment name is **1.Freddie Mac Classification Tutorial**
2. Click on Dataset then select the freddie_mac_500_test dataset
3. Initiate the diagnostics model by clicking on **Launch Diagnostics**



Create new model diagnostic

DIAGNOSED EXPERIMENT

1 1.Freddie Mac Classification Tutorial

TEST DATASET

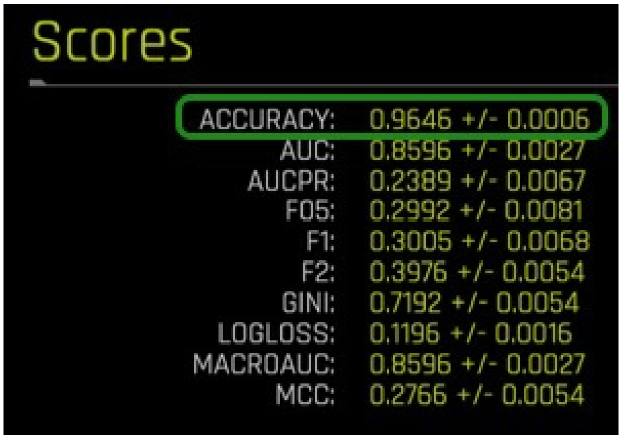
2 freeddie_mac_500_test

3 LAUNCH DIAGNOSTICS

7. After the model diagnostics is done running a new diagnostic will appear

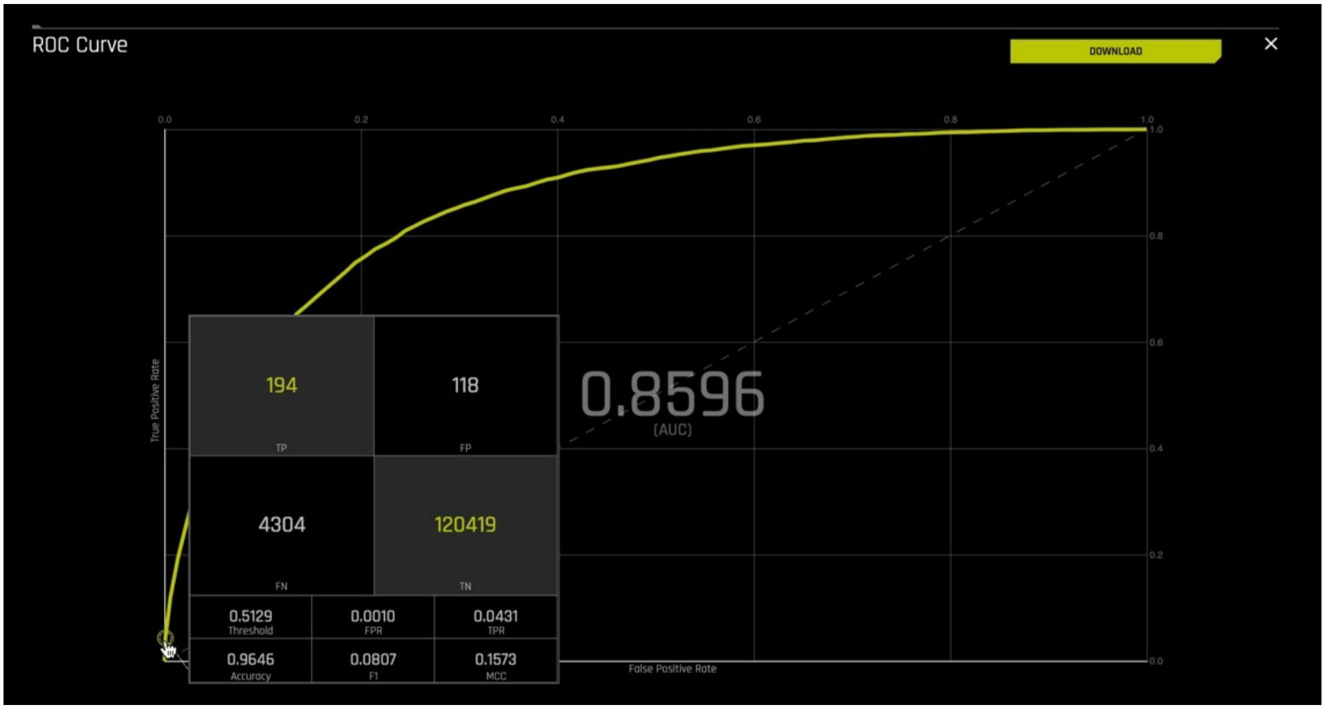
8. Click on the new diagnostics model. On the **Scores** section observe the accuracy value. Compare this Accuracy value to the Accuracy value from task 6.

9. Next, locate the new ROC curve and click on it. Hover over the **Best ACC** point on the curve. An image similar to the one below will appear:



Scores

ACCURACY:	0.9646 +/- 0.0006
AUC:	0.8596 +/- 0.0027
AUCPR:	0.2389 +/- 0.0067
F05:	0.2992 +/- 0.0081
F1:	0.3005 +/- 0.0068
F2:	0.3976 +/- 0.0054
GINI:	0.7192 +/- 0.0054
LOGLOSS:	0.1196 +/- 0.0016
MACROAUC:	0.8596 +/- 0.0027
MCC:	0.2766 +/- 0.0054



194 TP	118 FP
4304 FN	120419 TN

0.5129 Threshold	0.0010 FPR	0.0431 TPR
0.9646 Accuracy	0.0807 F1	0.1573 MCC

How much improvement did we get from optimizing the accuracy via the scorer?

The new model predicted:

- Threshold = .5129
- TP = 1 = 194 cases predicted as defaulting and defaulted
- TN = 0 = 120,419 cases predicted as not defaulting and did not default
- FP = 1 = 118 cases predicted as defaulting and did not default
- FN = 0 = 4,304 cases predicted not to default and defaulted

The first model predicted:

- Threshold = .5047
- TP = 1 = 195 cases predicted as defaulting and defaulted
- TN = 0 = 120,414 cases predicted as not defaulting and did not default
- FP = 1 = 123 cases predicted as defaulting and did not default
- FN = 0 = 4,303 cases predicted to not default and defaulted

The threshold for best accuracy changed from .5047 for the first diagnostics model to .5129 for the new model. This increase in threshold improved accuracy or the number of correct predictions made as a ratio of all predictions made. Note, however, that while the number of FP decreased the number of FN increased. We were able to reduce the number of cases that were predicted to falsy default, but in doing so, we increased the number of FN or cases that were predicted not to default and did.

The takeaway is that there is no win-win; sacrifices need to be made. In the case of accuracy, we increased the number of mortgage loans, especially for those who were denied a mortgage because they were predicted to default when, in reality, they did not. However, we also increased the number of cases that should not have been granted a loan and did. As a mortgage lender, would you prefer to reduce the number of False Positives or False Negatives?

10. Exit out of the ROC curve by clicking on the **x** located at the top-right corner of the plot, next to the **Download** option

Deeper Dive and Resources