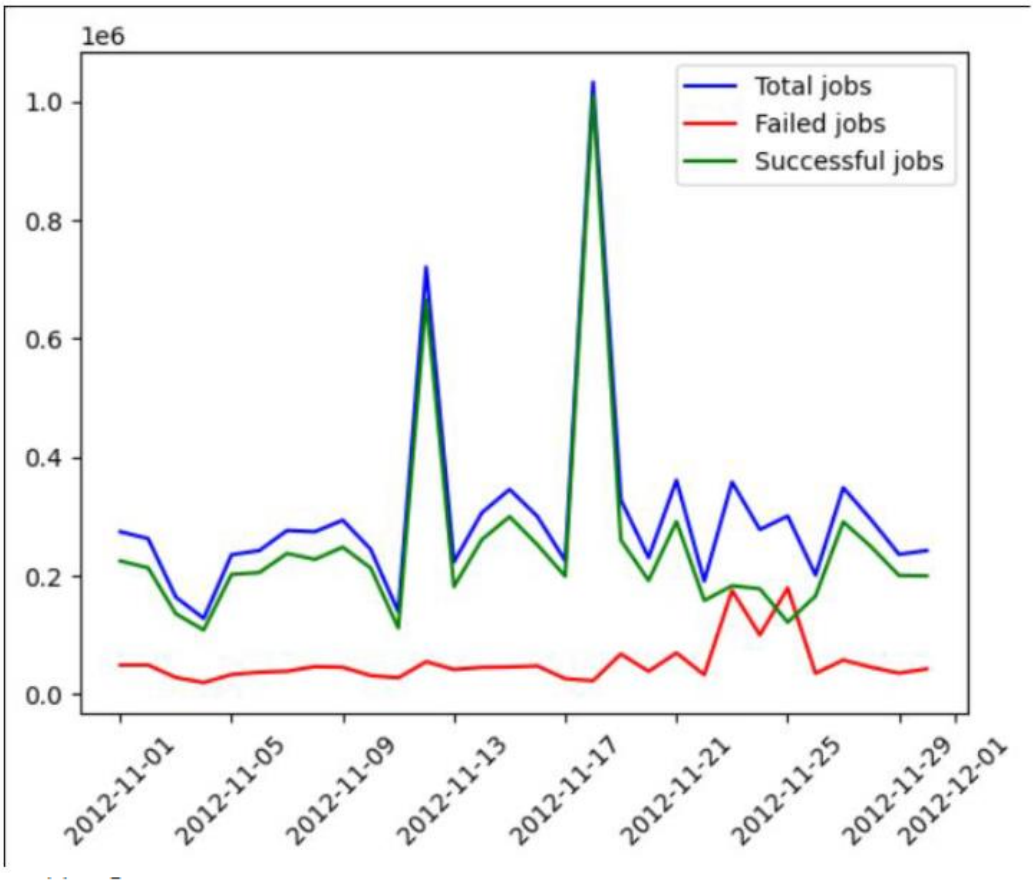
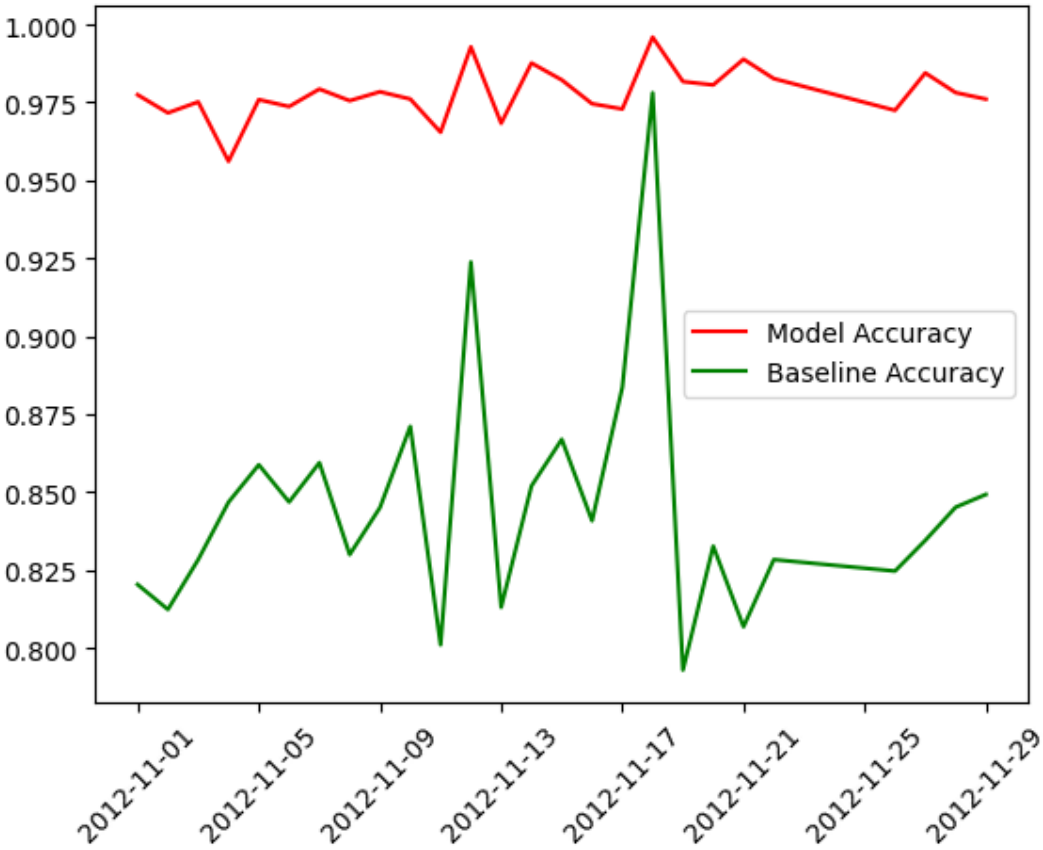
Dataset –:Netbatch Intel Log D

Start time : 12:58:29
Size of the Data Set: 7875826
Total Data Set Class Distribution:
Class 0: 6801710 samples
Class 1: 1074116 samples
Mean Cross-Validation Accuracy**: 0.95949**
Individual Cross-Validation Accuracies:
Fold 1: 0.96012
Fold 2: 0.95929
Fold 3: 0.95956
Fold 4: 0.95957
Fold 5: 0.95890
Baseline Accuracy: 0.86362
F1 Score: 0.9290815844642195
Precision: 0.973757957623296
Recall: 0.8883249109034779
Accuracy: 0.9815046701133316
Counts:
Predicted Positive: 979878
Actual Positive: 1074116
Difference Positive: 94238
Predicted Negative: 6895948
Actual Negative: 6801710
Difference Negative: -94238
Confusion Matrix (XGBoost Model):
[[6775996 25714]
[119952 954164]]
Mismatch Counts:
False Positives: 25714
False Negatives: 119952
Total TP+FN, which is defference in Test data count : 145666
Records where Predicted does not match Binary Exit Code: 145666, of Train data: 7875826
Percentage of records where Predicted does not match Binary Exit Code: 1.85%


1e6
1.0
0.8
0.6
0.4
0.2
0.0
Total jobs
Failed jobs
Successful jobs
2012-11-01
2012-11-05
2012-11-09
2012-11-13
2012-11-17
2012-11-21
2012-11-25
2012-11-29
2012-12-01


1.000
0.975
0.950
0.925
0.900
0.875
0.850
0.825
0.800
Model Accuracy
Baseline Accuracy
2012-11-01
2012-11-05
2012-11-09
2012-11-13
2012-11-17
2012-11-21
2012-11-25
2012-11-29

****

Dataset –:HotLum HPE Local custer year 22

Size of the Data Set: 344993

Total Data Set Class Distribution:
Class 0: 316081 samples
Class 1: 28912 samples

Mean Cross-Validation Accuracy: 0.94011
Individual Cross-Validation Accuracies:
Fold 1: 0.94025
Fold 2: 0.93962
Fold 3: 0.94059
Fold 4: 0.93977
Fold 5: 0.94032
Majority Class: 0
Count of Majority Class: 316077
Total Number of Samples: 344989
Baseline Accuracy: 91.62%

F1 Score: 0.6395544564743043
Precision: 0.8010828257587589
Recall: 0.5322357498616491
Accuracy: 0.9497236175806466

Counts:
Predicted Positive: 19209
Actual Positive: 28912

Difference Positive: 9703

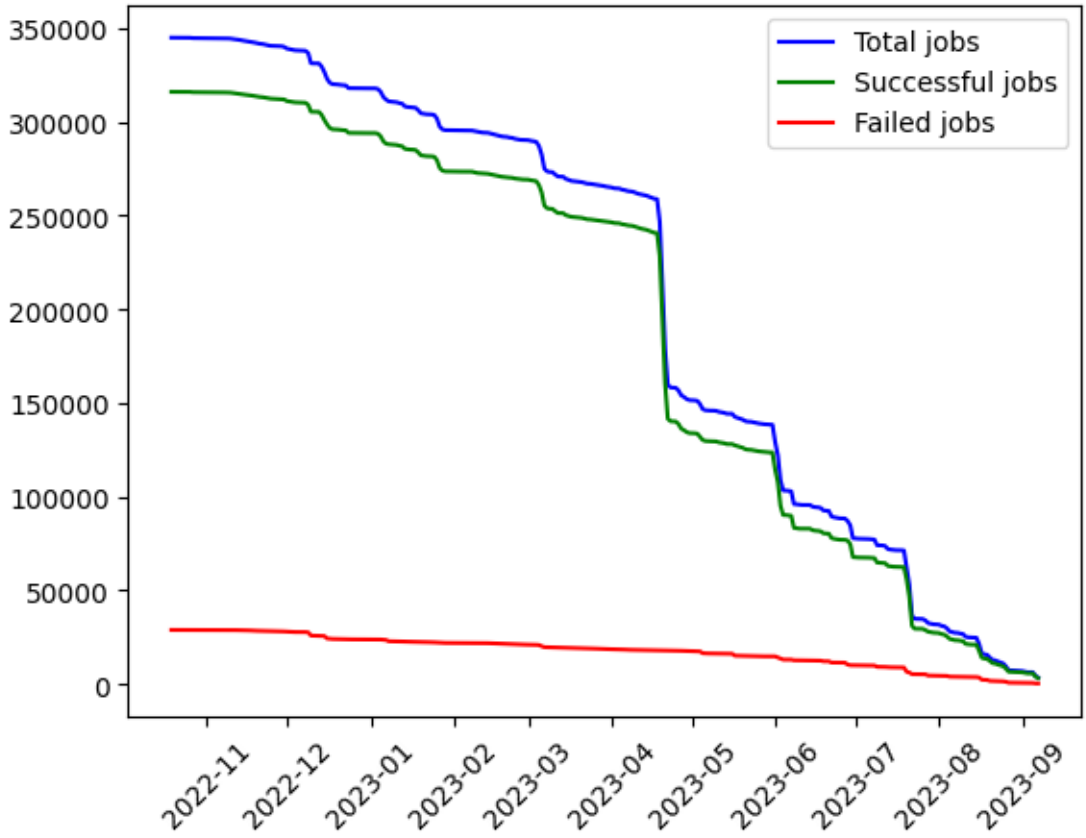
Predicted Negative: 325784
Actual Negative: 316081

Difference Negative: -9703

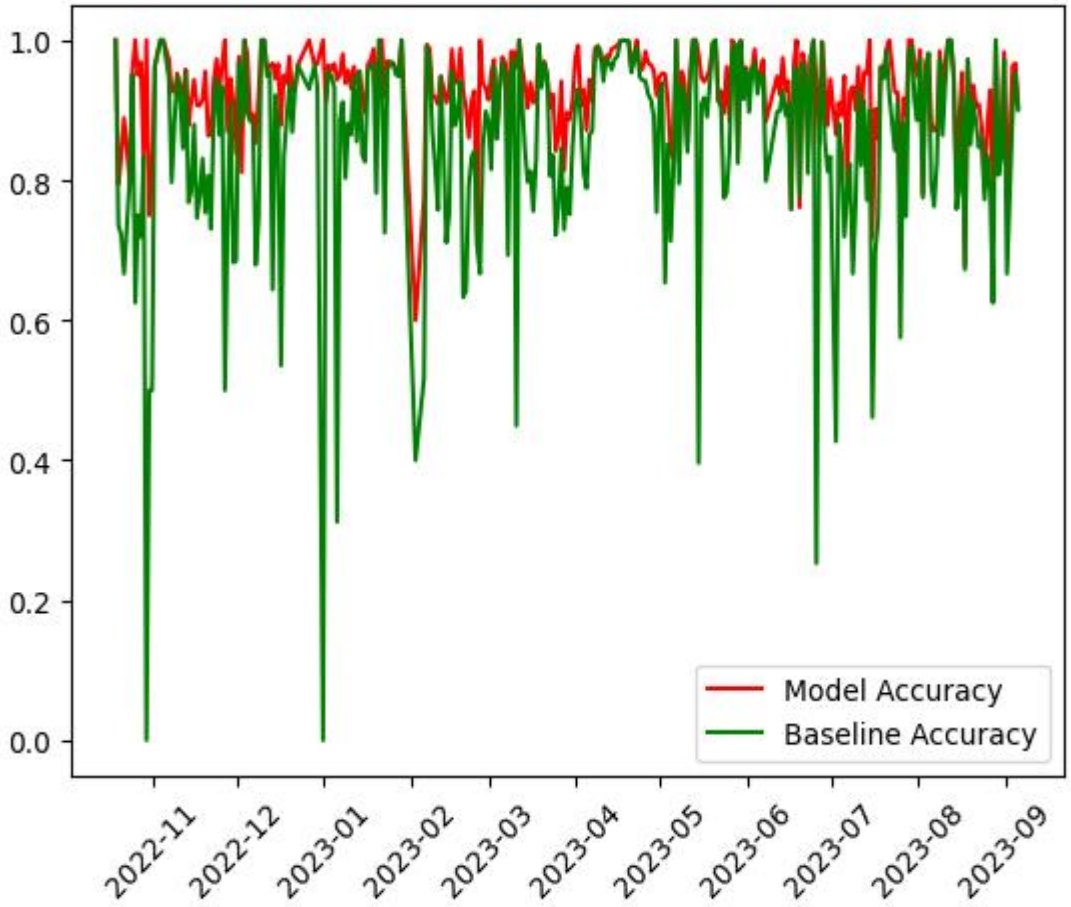
Confusion Matrix (XGBoost Model):
[[312260 3821]
 [13524 15388]]

Mismatch Counts:
False Positives: 3821
False Negatives: 13524
Total TP+FN, which is defference in Test data count : 17345

Records where Predicted does not match Binary Exit Code: 17345, of Train data: 344993
Percentage of records where Predicted does not match Binary Exit Code: 5.03%


Total jobs
Successful jobs
Failed jobs
350000
300000
250000
200000
150000
100000
50000
0
2022-11
2022-12
2023-01
2023-02
2023-03
2023-04
2023-05
2023-06
2023-07
2023-08
2023-09

:



Datset -: Los Alamos National Lab (LANL) CM-5 log

Start time : 11:07:06

Size of the Data Set: 122266

Total Data Set Class Distribution:
Class 0: 94409 samples
Class 1: 27857 samples

Training Set Class Distribution:
Class 0: 94407 samples
Class 1: 27857 samples

Testing Set Class Distribution:
Class 0: 2 samples
Class 1: 0 samples

Mean Cross-Validation Accuracy: 0.85083

Mean ROC AUC: 0.88094

Individual Cross-Validation Accuracies:
Fold 1: 0.88074
Fold 2: 0.88132
Fold 3: 0.87977
Fold 4: 0.88192
Fold 5: 0.88097

Majority Class: 0
Count of Majority Class: 94407
Total Number of Samples: 122264
Baseline Accuracy: 77.22%

F1 Score: 0.78351375508988
Precision: 0.877045535396656
Recall: 0.7080087590192771
Accuracy: 0.9108582925752049

Counts:
Predicted Positive: 22488
Actual Positive: 27857

Difference Positive: 5369

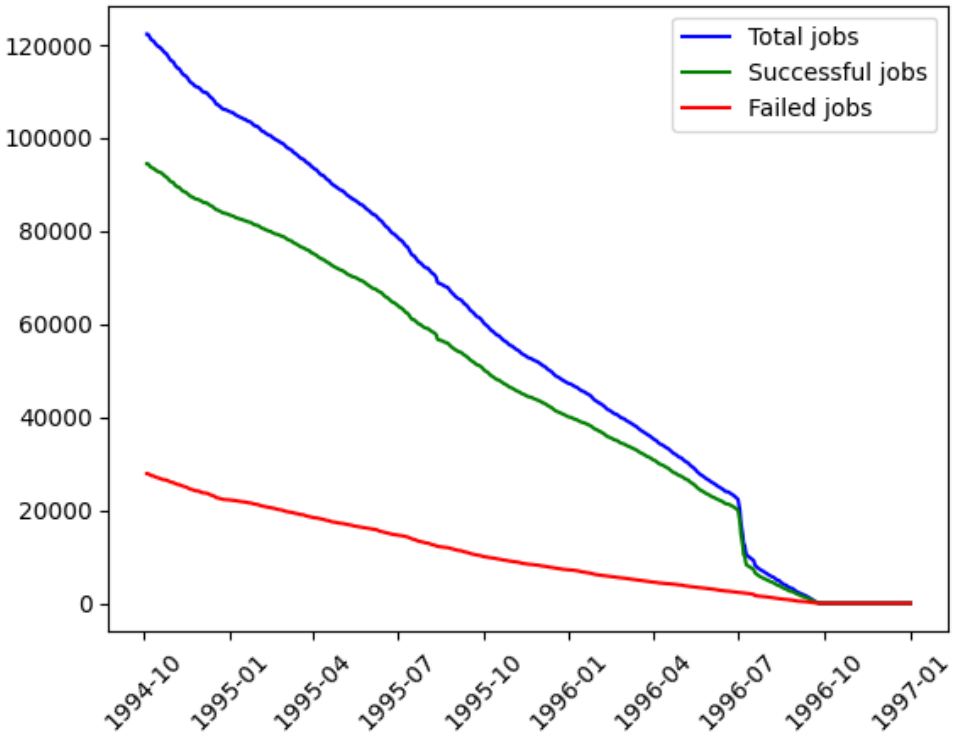
Predicted Negative: 99778
Actual Negative: 94409

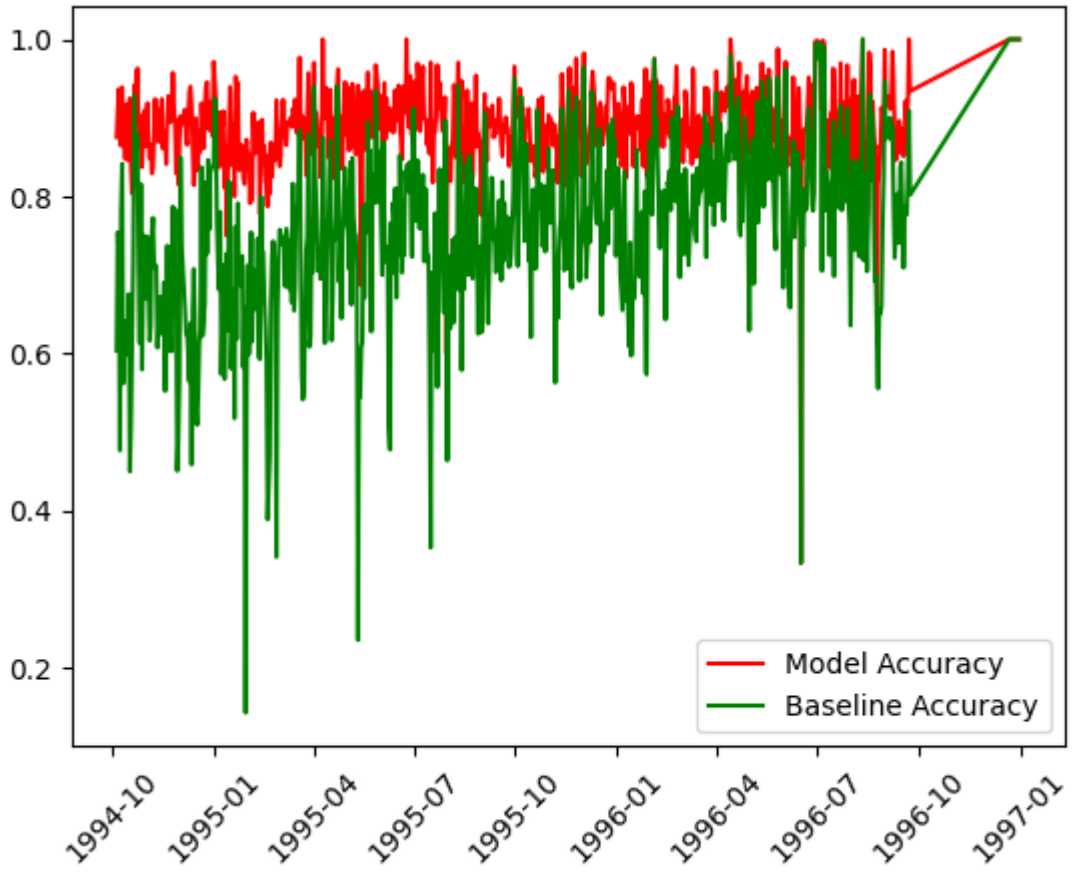
Difference Negative: -5369

Confusion Matrix (XGBoost Model):
[[91644 2765]
 [8134 19723]]

Mismatch Counts:
False Positives: 2765
False Negatives: 8134
Total TP+FN, which is defference in Test data count : 10899

Records where Predicted does not match Binary Exit Code: 10899, of Train data: 122266
Percentage of records where Predicted does not match Binary Exit Code: 8.91%


Total jobs
Successful jobs
Failed jobs
120000
100000
80000
60000
40000
20000
0
1994-10
1995-01
1995-04
1995-07
1995-10
1996-01
1996-04
1996-07
1996-10
1997-01


1.0
0.8
0.6
0.4
0.2
Model Accuracy
Baseline Accuracy
1994-10
1995-01
1995-04
1995-07
1995-10
1996-01
1996-04
1996-07
1996-10
1997-01

Data set-: The Cornell Theory Center (CTC) IBM SP2 log

Size of data set-: **79302**

Start time : 10:26:30

Size of the Data Set: 79302

Total Data Set Class Distribution:
Class 0: 62630 samples
Class 1: 16672 samples

Training Set Class Distribution:
Class 0: 62624 samples
Class 1: 16670 samples

Testing Set Class Distribution:
Class 0: 6 samples
Class 1: 2 samples

Mean Cross-Validation Accuracy: 0.82169

Mean ROC AUC: 0.77890

Individual Cross-Validation Accuracies:
Fold 1: 0.77975
Fold 2: 0.77387
Fold 3: 0.78340
Fold 4: 0.78170
Fold 5: 0.77580

Majority Class: 0
Count of Majority Class: 62624
Total Number of Samples: 79294
Baseline Accuracy: 78.98%

F1 Score: 0.8335897264397031
Precision: 0.9420256991685563
Recall: 0.7475407869481766
Accuracy: 0.9372525283095003

Counts:
Predicted Positive: 13230
Actual Positive: 16672

Difference Positive: 3442

Predicted Negative: 66072
Actual Negative: 62630

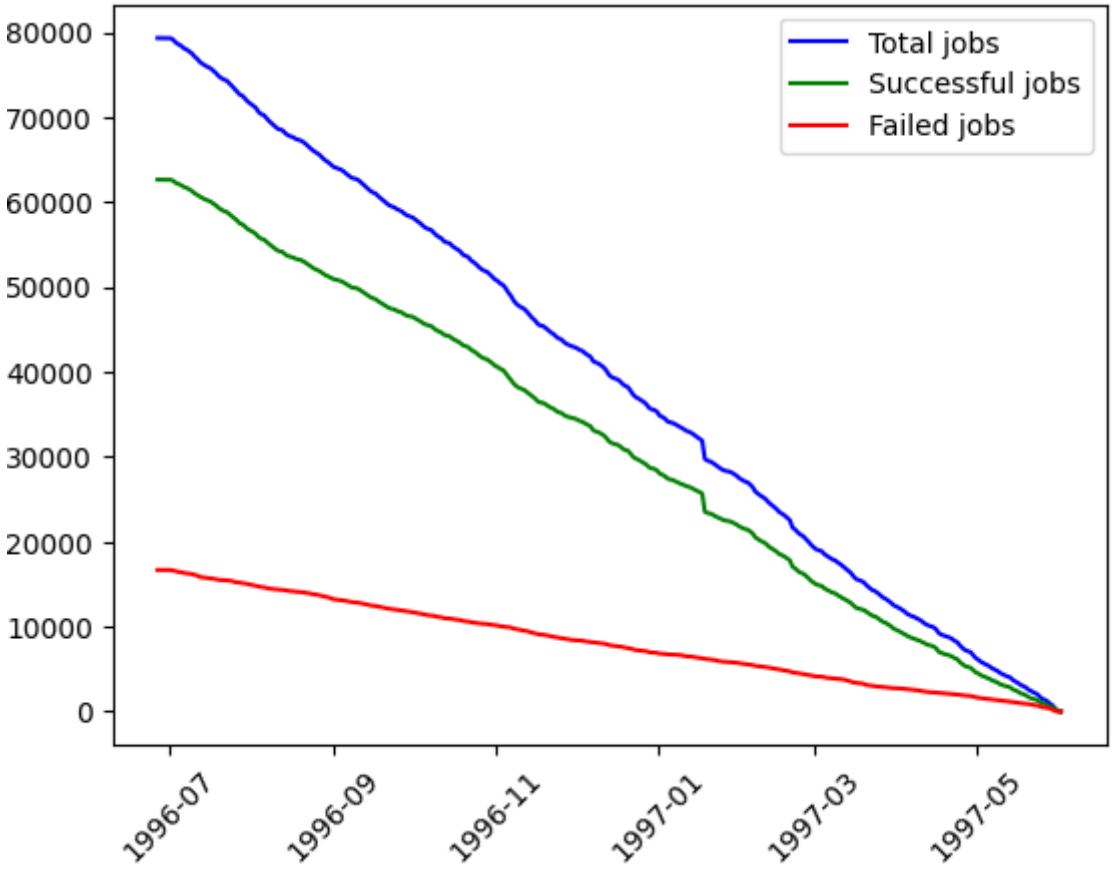
Difference Negative: -3442

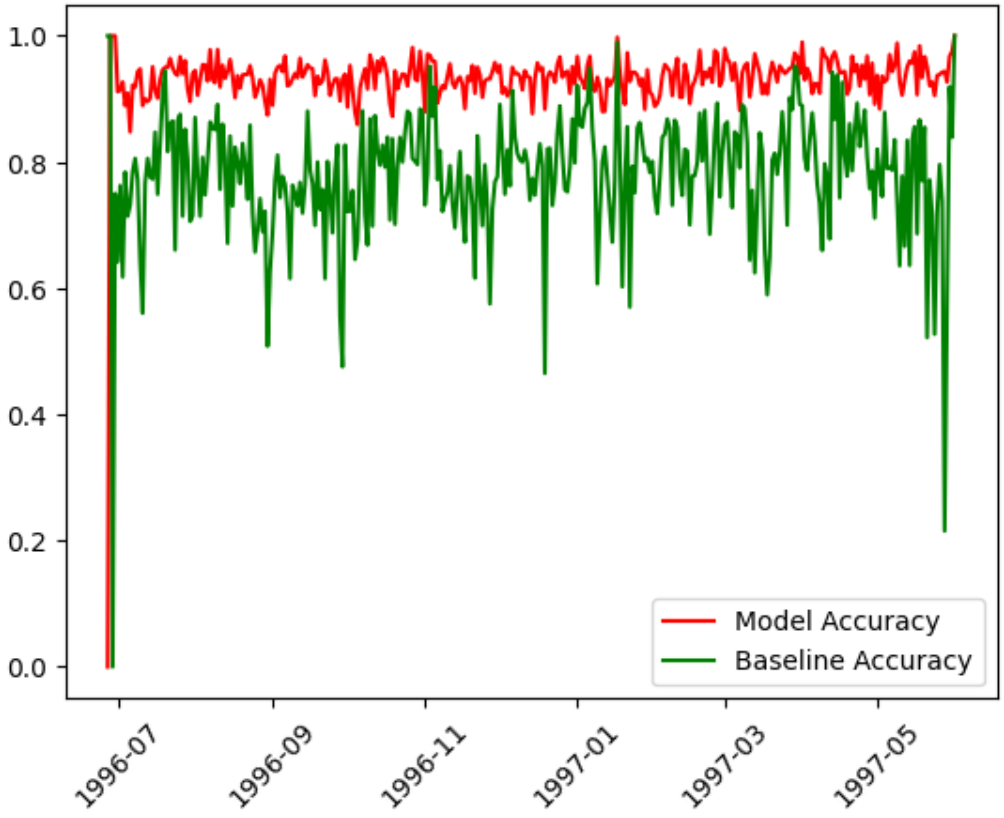
Confusion Matrix (XGBoost Model):
[[61863 767]
 [4209 12463]]

Mismatch Counts:
False Positives: 767
False Negatives: 4209

Total TP+FN, which is defference in Test data count : 4976

Records where Predicted does not match Binary Exit Code: 4976, of Train data: 79302
Percentage of records where Predicted does not match Binary Exit Code: 6.27%




1.0
0.8
0.6
0.4
0.2
0.0
1996-07
1996-09
1996-11
1997-01
1997-03
1997-05
Model Accuracy
Baseline Accuracy