CAPSTONE PROJECT-DATA SCIENCE

DEVELOPMENT OF PREDICTIVE MODEL

FOR

FAILURE PREDICTIONS

BATCH 15 (IITD ACDSDS-03)

Anupam Deep Verma

Anurag Kumar

BACKGROUND OF PROJECT

- This project pertains to **predictive maintenance** challenges encountered in the Metro train.

- The project focuses on analysing data from an ***Air Production Unit (APU),*** installed on the roof of metro train, to predict when the ***system is approaching failure***.

- Readings from multiple sensors of APU is available for analysis.

- This date set was taken from Kaggle.

- The objective of case study is to develop a machine learning algorithm to predict failures using the historical data, to avoid operational problems.

DATA SET DETAILS

- The system installed in the vehicle's APU collects data from Seven analog sensors and eight digital sensors.
- The dataset consists of 15169480 data points collected at 10 Hz from February 20 to August 2020 and is described by 15 features from 7 analogue (1-7) and 8 digital (8-15) sensors:

ANALOG SENSORS

1. TP2 (bar) – the measure of the pressure on the compressor.

2. TP3 (bar) – the measure of the pressure generated at the pneumatic panel.

3. H1 (bar) – the measure of the pressure generated due to pressure drop when the discharge of the cyclonic separator filter occurs.

4. DV pressure (bar) – the measure of the pressure drop generated when the towers discharge air dryers; a zero reading indicates that the compressor is operating under load.

5. Reservoirs (bar) – the measure of the downstream pressure of the reservoirs, which should be close to the pneumatic panel pressure (TP3).

6. Motor Current (A) – the measure of the current of one phase of the three-phase motor; it presents values close to 0A - when it turns off, 4A - when working offloaded, 7A - when working under load, and 9A - when it starts working.

7. Oil Temperature (°C) – the measure of the oil temperature on the compressor.

DIGITAL SENSORS

8. COMP - the electrical signal of the air intake valve on the compressor; it is active when there is no air intake, indicating that the compressor is either turned off or operating in an offloaded state.

9. DV electric – the electrical signal that controls the compressor outlet valve; it is active when the compressor is functioning under load and inactive when the compressor is either off or operating in an offloaded state.

10. TOWERS – the electrical signal that defines the tower responsible for drying the air and the tower responsible for draining the humidity removed from the air; when not active, it indicates that tower one is functioning; when active, it indicates that tower two is in operation.

11. MPG – the electrical signal responsible for starting the compressor under load by activating the intake valve when the pressure in the air production unit (APU) falls below 8.2 bar; it activates the COMP sensor, which assumes the same behaviour as the MPG sensor.

12. LPS – the electrical signal that detects and activates when the pressure drops below 7 bars.

13. Pressure Switch - the electrical signal that detects the discharge in the air-drying towers.

14. Oil Level – the electrical signal that detects the oil level on the compressor; it is active when the oil is below the expected values.

15. Caudal Impulse – the electrical signal that counts the pulse outputs generated by the absolute amount of air flowing from the APU to the reservoirs.

SAMPLE DATA

S. NO.	Data Recorded (10Hz)	date	time	TP2	TP3	H1	DV Pressure	Reservoirs	Oil_temp erature	Motor current	COMP	DV eletric	Towers	MPG	LPS	Pressure_ switch	Oil_level	Caudal impulses
1	0	2020-02-01	00:00:00	-0.012	9.358	9.340	-0.024	9.358	53.600	0.0400	1.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0
2	10	2020-02-01	00:00:10	-0.014	9.348	9.332	-0.022	9.348	53.675	0.0400	1.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0
3	20	2020-02-01	00:00:19	-0.012	9.338	9.322	-0.022	9.338	53.600	0.0425	1.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0
4	30	2020-02-01	00:00:29	-0.012	9.328	9.312	-0.022	9.328	53.425	0.0400	1.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0
5	40	2020-02-01	00:00:39	-0.012	9.318	9.302	-0.022	9.318	53.475	0.0400	1.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0

DATA POINTS

15169480

CONTINUOUS DATA

CATEGORICAL DATA

FAILURE INFORMATION

Nr.	Start Time	End Time	Failure	Severity	Report
#1	4/18/2020 0:00	4/18/2020 23:59	**Air leak**	High stress	
#1	5/29/2020 23:30	5/30/2020 6:00	**Air Leak**	High stress	Maintenance on 30Apr at 12:00
#3	6/5/2020 10:00	6/7/2020 14:30	**Air Leak**	High stress	Maintenance on 8Jun at 16:00
#4	7/15/2020 14:30	7/15/2020 19:00	**Air Leak**	High stress	Maintenance on 16Jul at 00:00

METHODOLOGY

Data Preprocessing

Cleaned data by removing missing values and unnecessary columns.

Data Analysis

Explore Pre-Failure Indicators

Feature Selection

Selected important features and dropped remaining features

Data Labeling

Labelled data by defining weekly, daily, and hourly pre-failure periods for detailed analysis.

Segmented data into normal operation and failure-specific periods to avoid data contamination.

Predictive Modeling

Compared different models like Logistic Regression, XG Boost, XG Boost Rolling Statistics, K-Means and Neural Network to classify healthy and impending failure states.

DATA ANALYSIS

SOME PRE FAILURE INDICATORS

```
Low Oil Level ratio during Normal Operation:7.301412917092421

Low Oil Level ratio during Week Pre Failure: 88.1465666474322

 week_df  Oil_level
0        1.0          1036107
         0.0           141905
1        1.0           305516
         0.0             3466
```

OIL LEVEL RATIO

Oil_temperature

week_df	
0	61.888252
1	64.270368

OIL TEMPERATURE

DV_pressure

hourly_df	
0	0.020033
1	-0.001444

DV PRESSURE

```
Pressure Switch ratio during Normal Operation:0.00750920262189197

Pressure Switch ratio during Week Pre Failure: 0.013284972665947378

 week_df  Pressure_switch
0        1.0          1169232
         0.0             8780
1        1.0           304931
         0.0             4051
```

PRESSURE SWITCH

FEATURE SELECTION

- CORRELATION ANALYSIS WAS DONE

- H1, COMP, DV_ELECTRIC WERE DROPPED

  - H1 - THIS COLUMN IS HIGHLY CORRELATED WITH MULTIPLE FEATURES, AND AS EVALUATED PREVIOUSLY, IS NOT CORRELATED WITH ANY OF THE PRE FAILURE INDICATORS.

  - COMP - THIS SENSOR IS ACTIVE WHEN THE MPG SENSOR IS ACTIVE, WE WILL DROP THIS FEATURE AS THE MPG ACTIVATES BASED ON AN ACTUAL MEASUREMENT OF PRESSURE, WHICH THEN ACTIVATES THE COMP SENSOR.

  - DV_ELECTRIC - THE BEHAVIOR OF THIS FEATURE CAN BE QUANTIFIED BY THE MPG FEATURE THAT WE ARE KEEPING, SO WE WILL DROP THIS COLUMN.

- REMAINING FEATURES WERE USED FOR FURTHER ANALYSIS

Correaltion Matrix
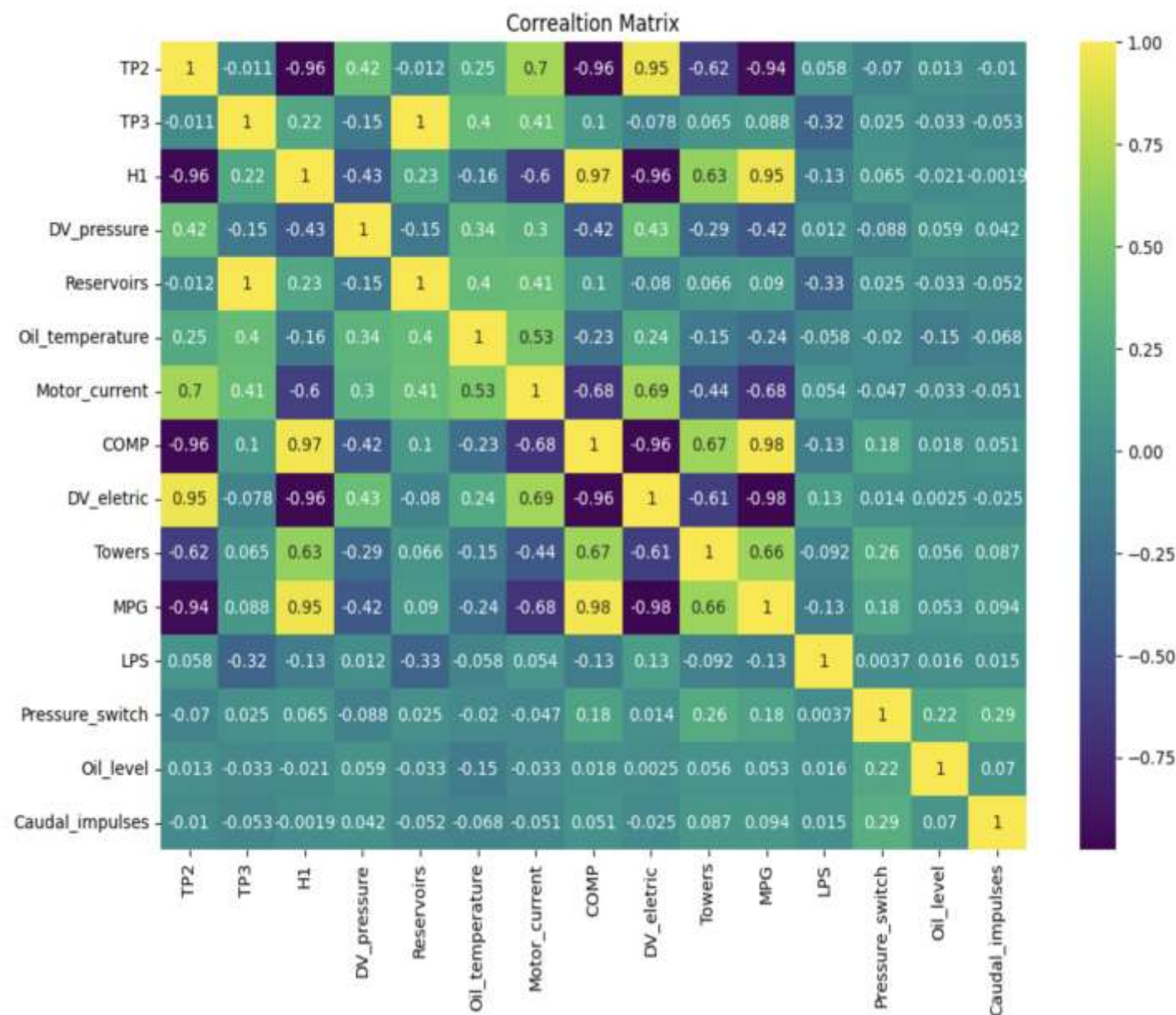
DATA LABELING

- THREE COLUMNS WERE ADDED USING THE FAILURE INFORMATION GIVEN
  - ONE WEEK BEFORE FAILURE – DAY INTERVAL
  - ONE DAY BEFOR FAILURE – HOURS INTERVAL
  - ONE HOUR BEFOR FAILURE – 10 SEC INTERVAL
- LABELING OF DATA FOR SPECIFIC TIME PERIODS WAS ALSO DONE
  - BEFOR FAILURE
  - BETWEEN FAILURE

Nr.	Start Time	End Time	Failure	Severity	Report
#1	4/18/2020 0:00	4/18/2020 23:59	Air leak	High stress	
#1	5/29/2020 23:30	5/30/2020 6:00	Air Leak	High stress	Maintenance on 30Apr at 12:00
#3	6/5/2020 10:00	6/7/2020 14:30	Air Leak	High stress	Maintenance on 8Jun at 16:00
#4	7/15/2020 14:30	7/15/2020 19:00	Air Leak	High stress	Maintenance on 16Jul at 00:00

Failure Information

week_df	hourly_df	second_wise_df	failure_1	failure_2	failure_3	failure_4
0	0	0	True	False	False	False
0	0	0	True	False	False	False
0	0	0	True	False	False	False
0	0	0	True	False	False	False
0	0	0	True	False	False	False

Data Labeling

PREDICTIVE MODELING

- LOGISTIC REGRESSION

- X-G BOOST METHOD

- IMPROVEMENT IN X-G BOOST USING ROLLING SATISTICS

- K-MEANS

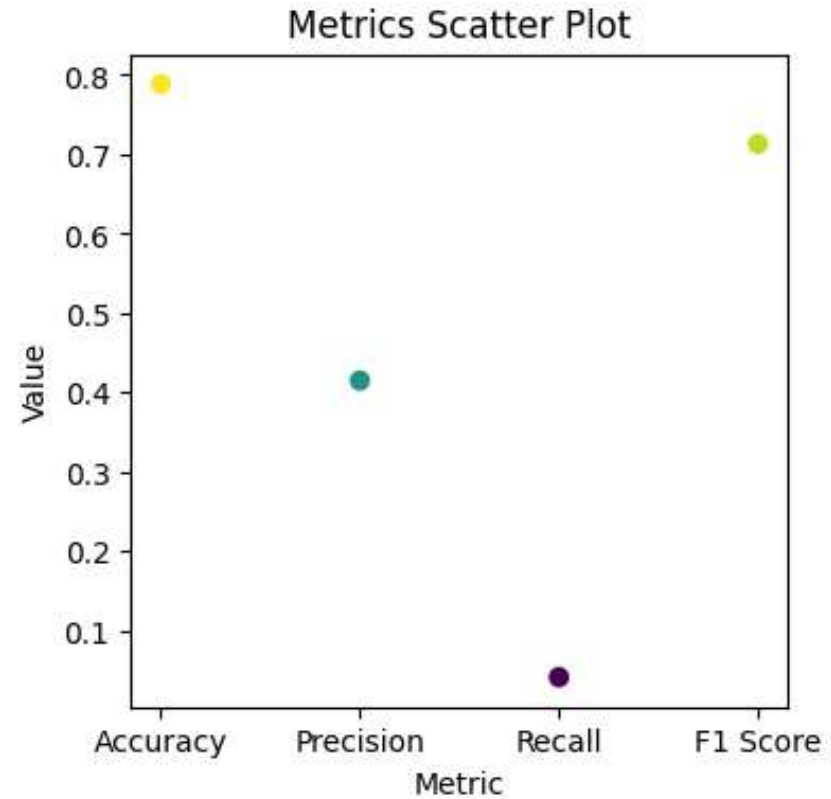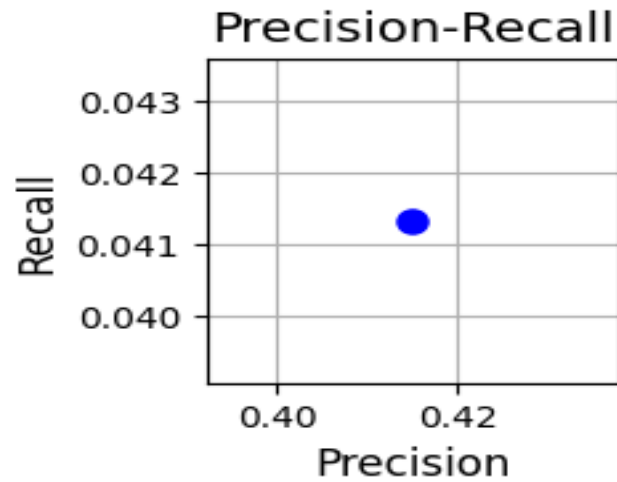- NEURAL NETWORK - MLP CLASSIFIER

LOGISTIC REGRESSION

RESULTS

Accuracy: 0.7886948861127239

Precision: 0.4150595882990249
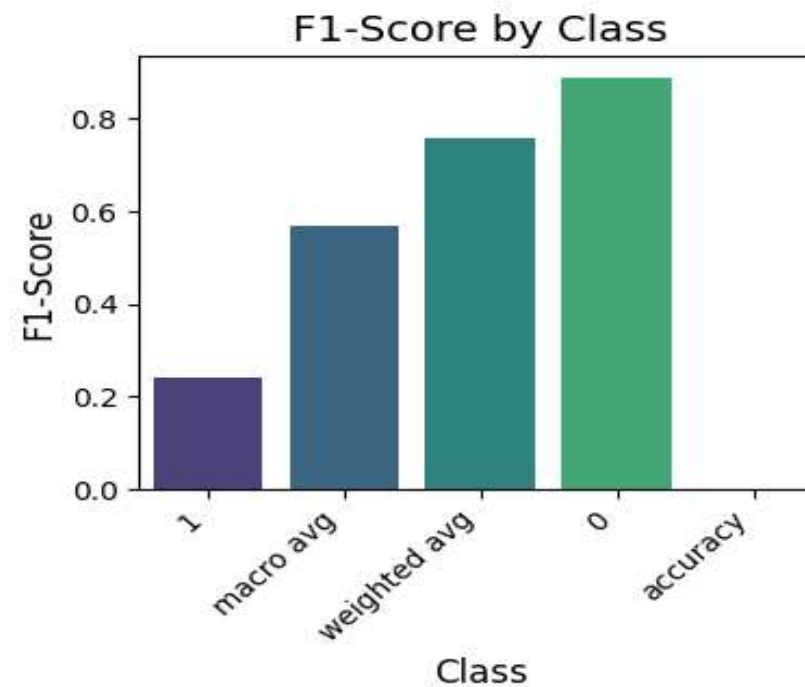
Recall: 0.04132909002b430766

F1 Score: 0.713335977487S245

XGB CLASSIFIER MODEL FROM THE XGBOOST

```
              precision    recall  f1-score   support

           0       0.82      0.99      0.89    353404
           1       0.82      0.14      0.24     92695

    accuracy                           0.82    446099
   macro avg       0.82      0.57      0.57    446099
weighted avg       0.82      0.82      0.76    446099

0.3998747822574337
0.7966753352709482
```
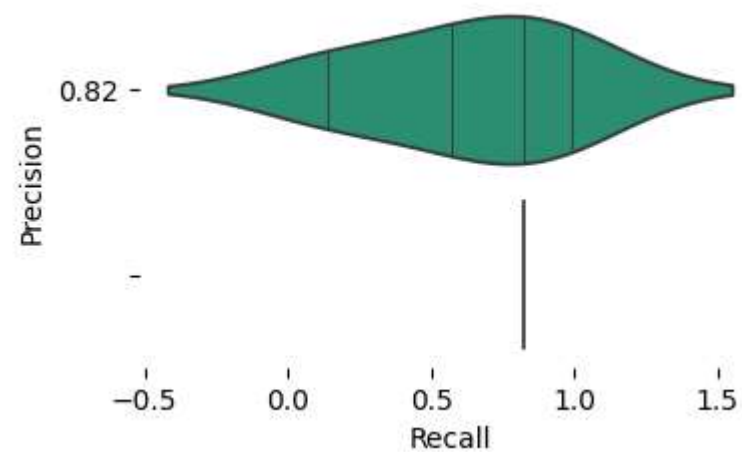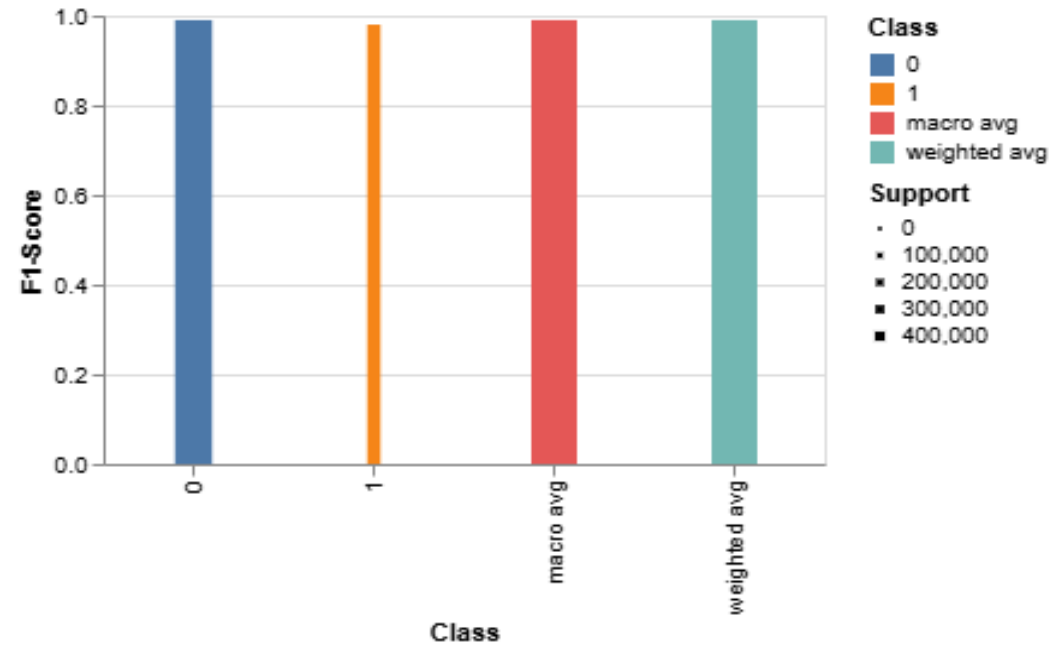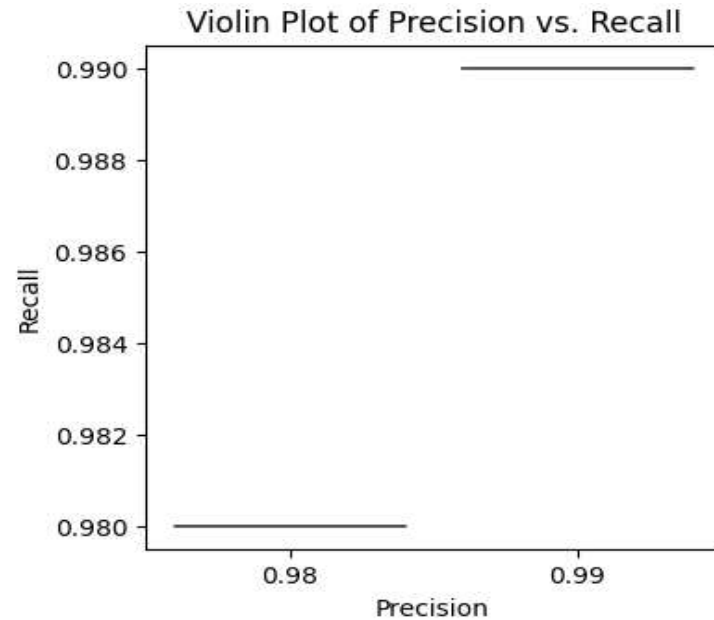

F1-Score by Class

IMPROVEMENT IN X-G BOOST USING ROLLING SATISTICS

	precision	recall	f1-score	support
0	0.99	0.99	0.99	353404
1	0.98	0.98	0.98	92695
accuracy			0.99	446099
macro avg	0.99	0.99	0.99	446099
weighted avg	0.99	0.99	0.99	446099

0.06694035539112968
0.9774390541491587
0.991533269520891

Violin Plot of Precision vs. Recall

K-MEANS MODEL

```
KMeans Accuracy: 0.4775195640429591
KMeans Precision: 0.25142804528665375
KMeans Recall: 0.7659312800043152
KMeans F1 Score: 0.3785812932914572
              precision    recall  f1-score   support

           0       0.87      0.40      0.55    353404
           1       0.25      0.77      0.38     92695

    accuracy                           0.48    446099
   macro avg       0.56      0.58      0.46    446099
weighted avg       0.74      0.48      0.51    446099
```


F1-Score per Class

Violin Plot of Precision by Class
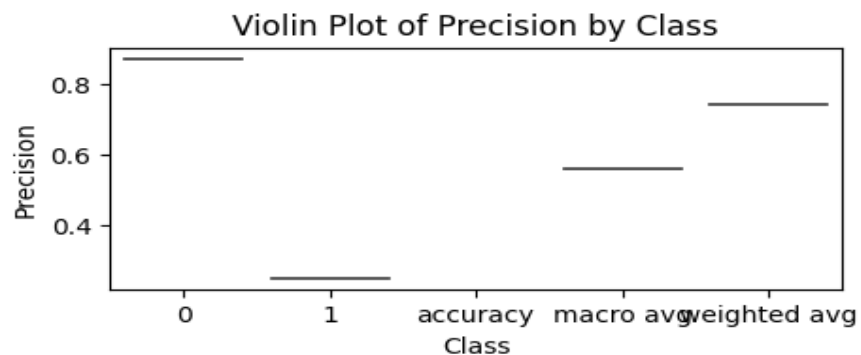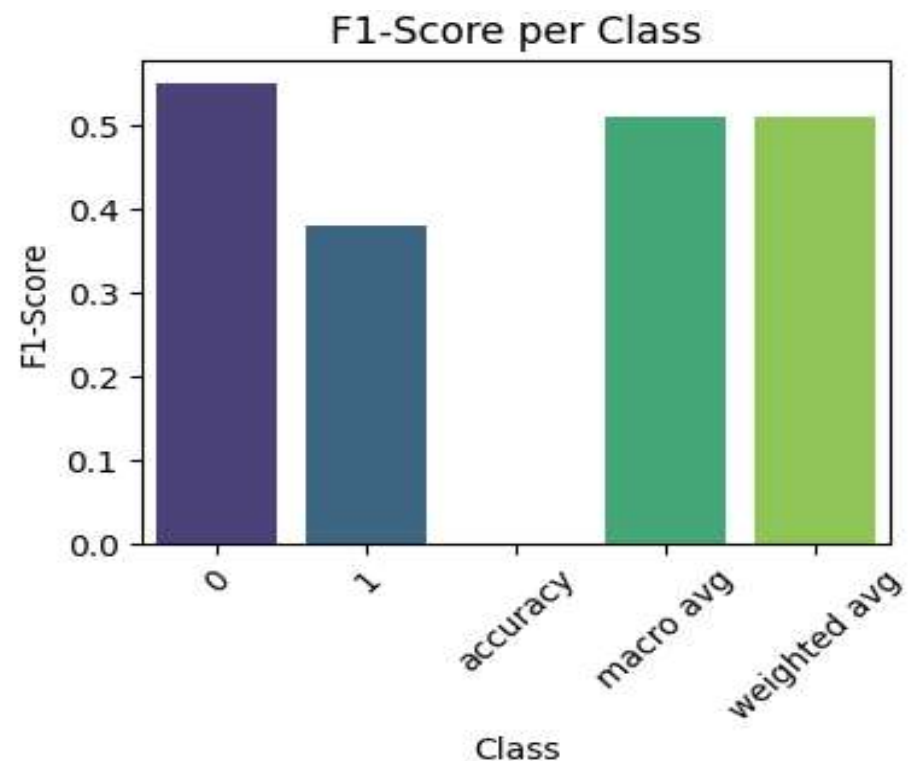
# NEURAL NETWORK MLP CLASSIFIER
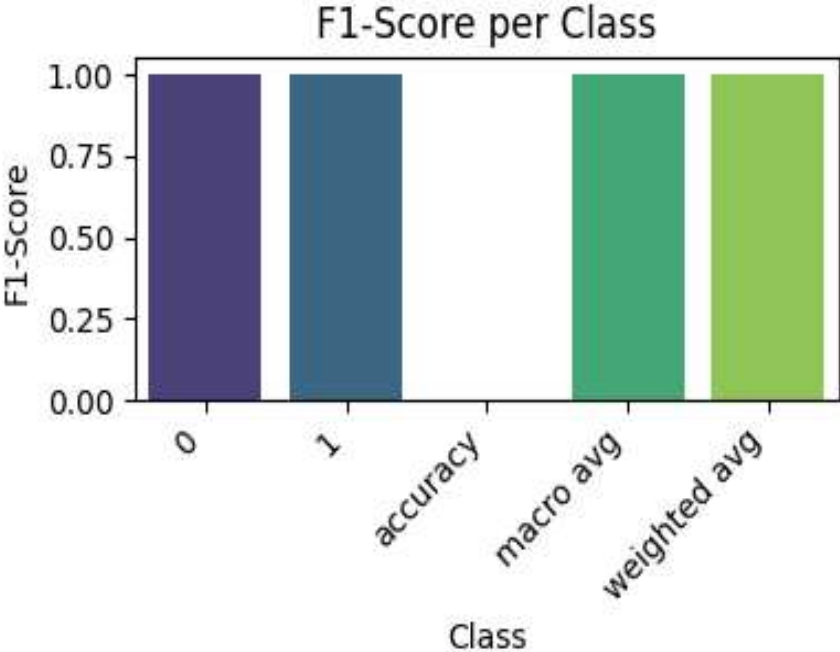
```
MLP Accuracy: 0.9998386008486906
MLP Precision: 0.9997949514898393
MLP Recall: 0.9994282323749932
MLP F1 Score: 0.9996115582986254
              precision    recall   f1-score    support

           0       1.00      1.00       1.00     353404
           1       1.00      1.00       1.00      92695

    accuracy                            1.00     446099
   macro avg       1.00      1.00       1.00     446099
weighted avg       1.00      1.00       1.00     446099
```
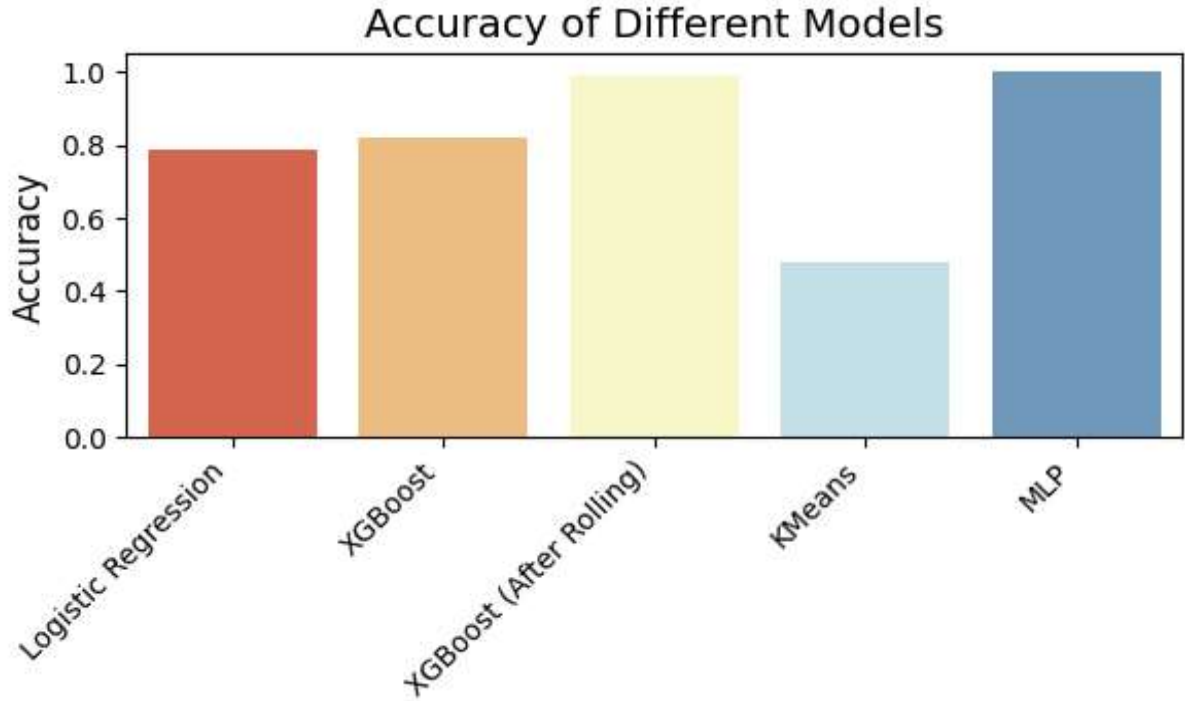

F1-Score per Class

Violin Plot of Precision vs. Recall
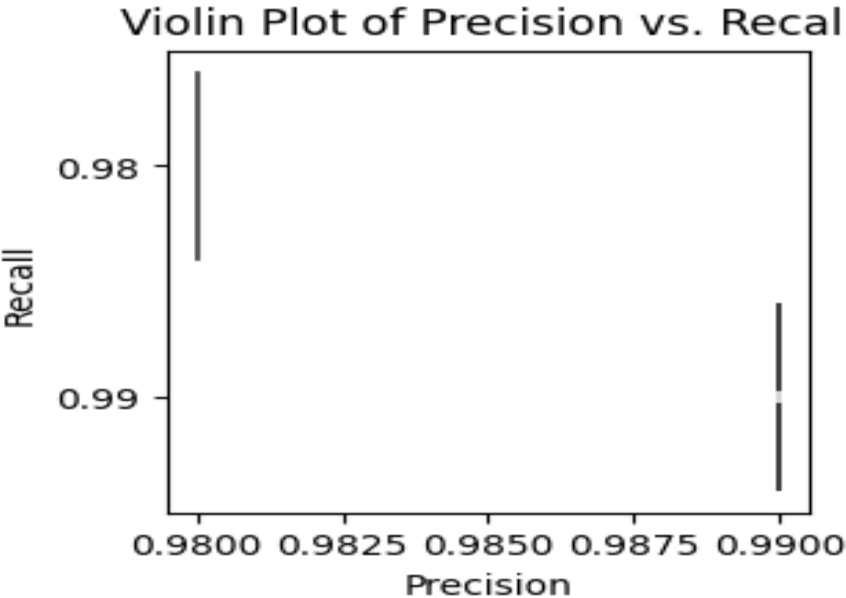
COMPARISON OF VARIOUS MODELS

	Model	Accuracy
0	Logistic Regression	0.788695
1	XGBoost	0.820000
2	XGBoost (After Rolling)	0.990000
3	KMeans	0.477520
4	MLP	0.999839

Accuracy of Different Models

	ACCURACY	PRECISION	RECALL	F1 SCORE
LOGISTIC	0.78	0.41	0.041	0.71
XG BOOST	0.82			
CLASS 0		0.82	0.99	0.89
CLASS 1		0.82	0.14	0.24
XGBOOST + ROLL OVER	0.99			
CLASS 0		0.99	0.99	0.99
CLASS 1		0.98	0.98	0.98
K-MEANS	0.48			
CLASS 0		0.87	0.40	0.55
CLASS 1		0.25	0.77	0.38
NEURAL NETWORK	0.99			
CLASS 0		0.99	0.99	0.99
CLASS 1		0.99	0.99	0.99

THANK YOU