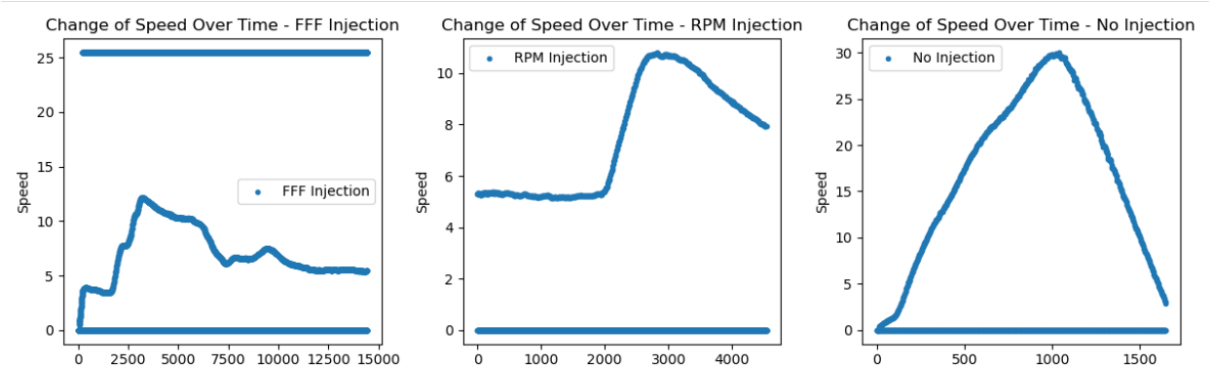
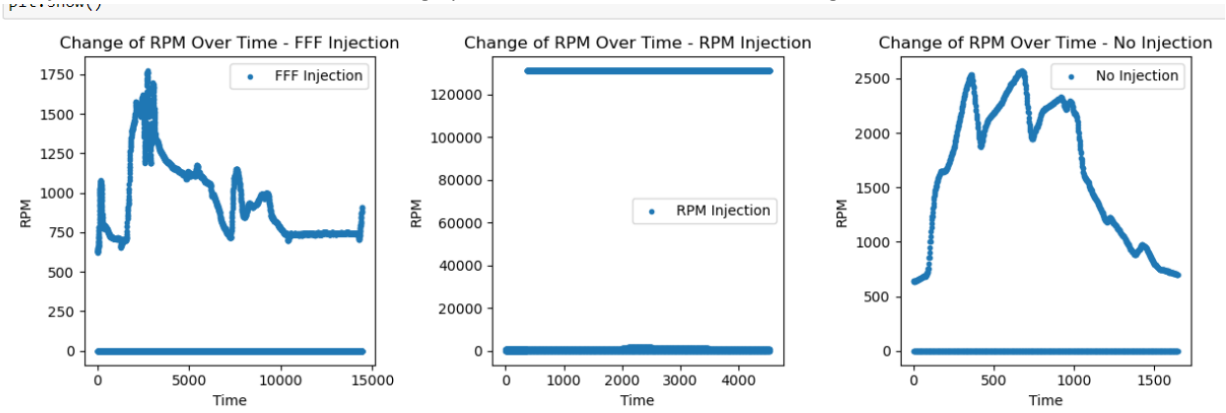
DEVELOP SUPERVISED MACHINE LEARNIG MODEL

#TASK 2
STEP 1:-
SPEED VS TIME:

FFF INJECTION:-
At initial from 0-2500 on X-axis the vehicle is slowly moving
2)As time is moving, there is increase and rapid decrease of speed. This might be attack
3)As there is rapidly changing indicates there is a potential attack on vehicles speed control system
RPM INJECTION:-
initially the speed is constant at low speed, indicating there was a attack on RPM data, causing engine to operate at abnormal Speed
2)Increasing significantly and suddenly decreased indicating the attacker is manipulating the engines RPM
NO INJECTION:-
There is smooth increase of speed and slowly the speed has decreased .Indicating there is no attack

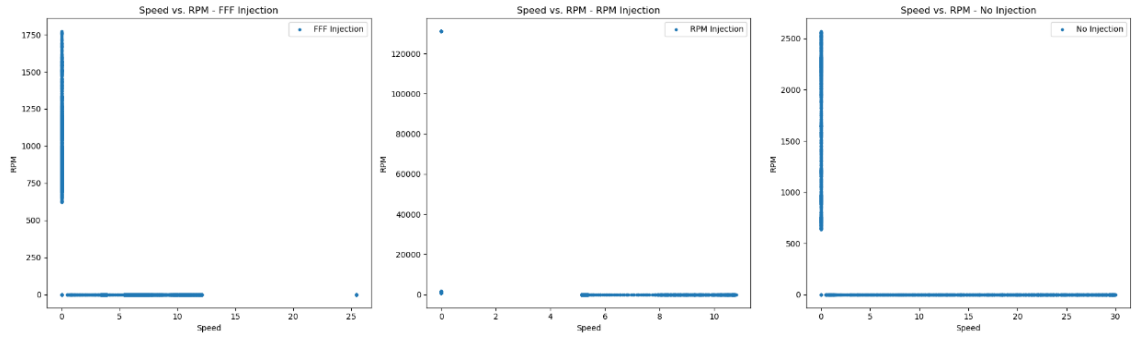


RPM VS TIME
FFF INJECTION:-
1)Rapid fluctuation in RPM may indicate the attack on vehicles engine .The attacker have injected fabricated RPM CAN message leading to unstable engine RPM behavior
2)RPM Injection:- The RPM is totally constant ,indicating the attacker has manipulated the RPM data leading to maintain a specific engine speed
3)No Injection:- There is no fluctuation in the graph ,the RPM is increases and decreasing little. Its a natural behavior of car



SPEED VS RPM
FFM injection: Clearly the attack has happened as RPM has highest 1750 at low speed
RPM injection :- There is a value for speed and corresponding there is no value for RPM
NO INJECTION:- both speed and rpm are increasingly. Clearly showing no attack

```
plt.show()
```



STEP 2:-

Frequency vs SPEED OR RPM –

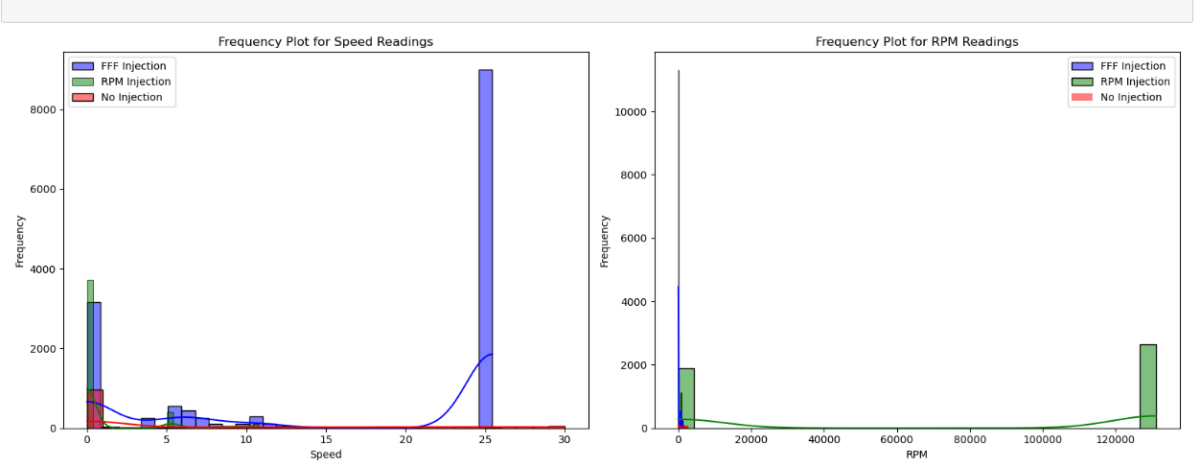
Clearly it is shown that FFM injection is more dominant i. Speed and RPM injection js more dominant in RPM Graph.

SPEED VS FRQUENCY

The speed has reached the highest frequency when it was FFM injection. For the FFM injection the speed as increased and became constant and increased suddenly and the RPM injection the speed almost constant .Indicating the speed is not too much effected by that data .

RPM VS FREQUENCY:-

The FFFM injection ja almost constant and is not affecting the RPM of car .RPM injection is slightly increased and became constant and increased .



STEP 3:- The correlation coefficient measures the linear relationship between two variables and P/value indicates the significance correlation

Correlation Coefficient:-Range is -1 to 1 .

FFM Injection :- It has value (-0.76) indicating strong negative correlation, as speed increases and Rpm decreases

RPM Injection ;- Negative correlation.

No Injection;- Negative Correlation

P-value :- All the three cases the p-value is extremely low so indicating highly statically significant, means the relationship between the variables is not random.

Task 3(Random Forest Classifier):-
Confusion matrix:-
True positive (2212):- The model has correctly identified 2212 cases as attacks
False positive: - There is no case where model incorrectly identified as attack
True negatives(2945):- Model has correctly identified the 2945 cases as not attack
False negative:- There is no case where model didn't missed any attack when the attack is present
Performance Metrics:- All the Accuracy, Precision, Recall and F1-Score are all 100% .The model is perfectly identify the attack .There is no case where model didn't missed the attack .The model is highly efficient

TASK 5:-
Both the models are highly efficient, the random forest has 100% accuracy and logistic had 99.9% .The False negative rates are 0% for both the models as they didn't miss any attack .In the case of logistic regression model the 5 cases where classified as attacks but actually they are not attacks (false positives) .The confusion matrix differ TP(True Positive) FOR Random Forest IS 2212 and for Logistic is 2207 and the FP(False Positive) is 0 for Random forest and 5 for logistic. but the both the model performance is high accuracy and high precision.

Task 6:-
Drawbacks of Using Supervised Learning Model:-

1. Limited feature engineering was used in the training of the models. Model performance might be enhanced by investigating and developing more characteristics, such as time-based feature etc.

2.Usage of Under sampling or Oversampling is necessary as there is imbalance of attacks vs non-attacks data .

3. Domain Expertise required as we are dealing with the vehicle data

Improve Efficiency:-

1. More Features to be added so that the model can able to predict any type of attack

2. To identify the unseen attacks, add the un-supervised or Semo-Supervised model

3. Learn from the previous model based on vehicle attack and build the model

Solve the problem better:-

1. To respond to changing attacks, it is crucial to continuously evaluate model performance and feedback loops to change models based on new attack patterns or data.

2. *Collaboration between experts in the automobile industry, cybersecurity experts, and data scientists can produce more reliable and efficient solutions. Bringing together topic knowledge and experience can aid in creating thorough defenses.

3. While tackling security issues is important, any solution design must put ethical considerations, such as privacy and data usage, first.