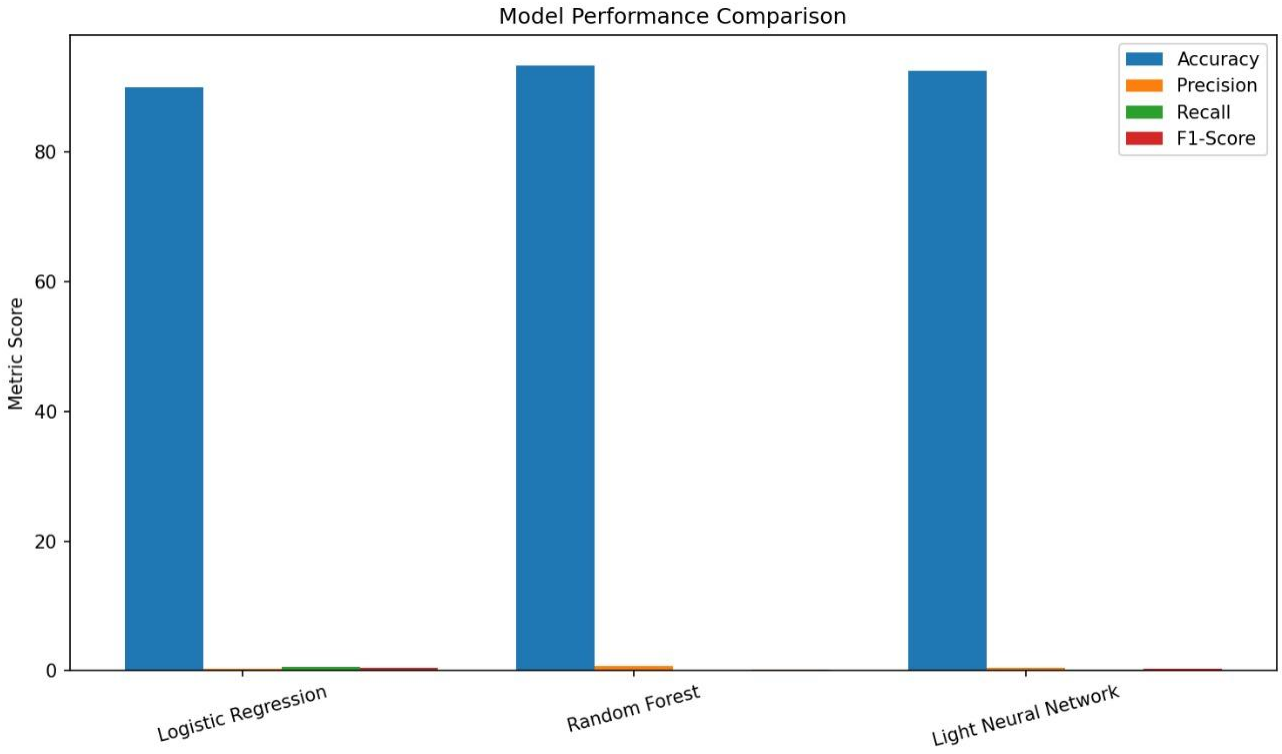
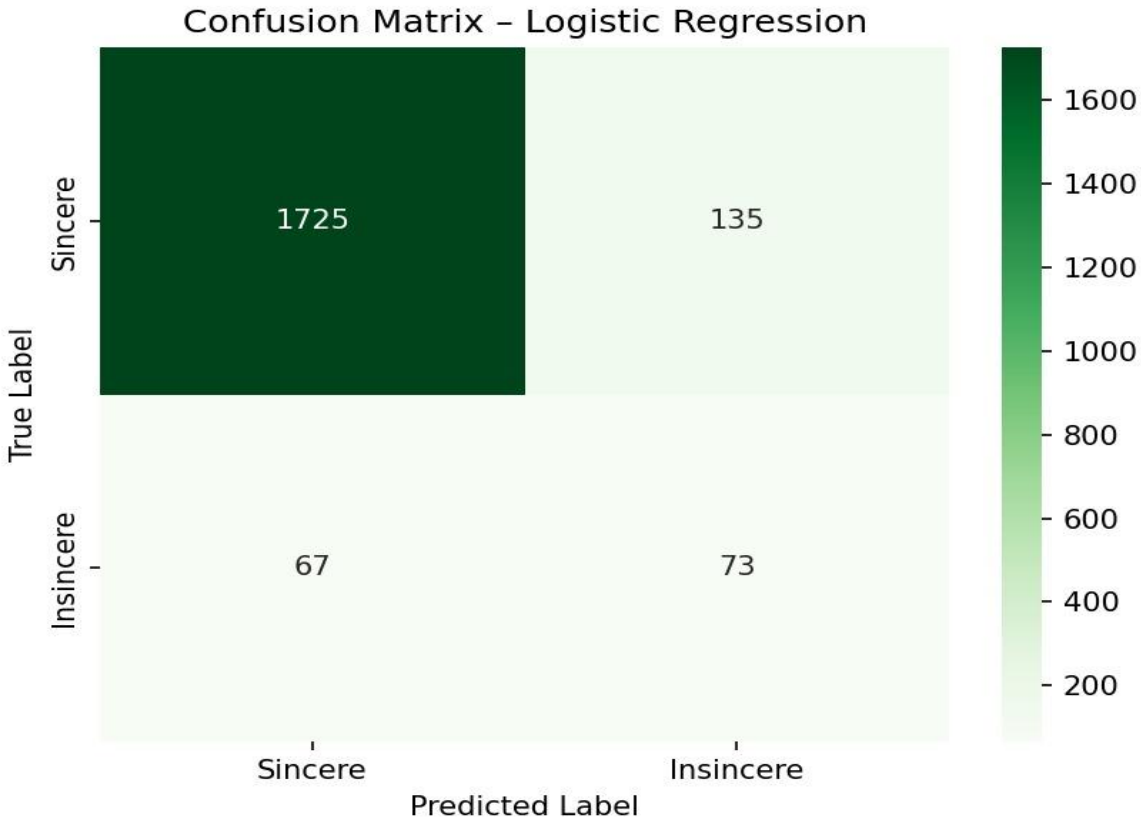
ML ASSIGNMENT OUTPUT SCREENSHOTS

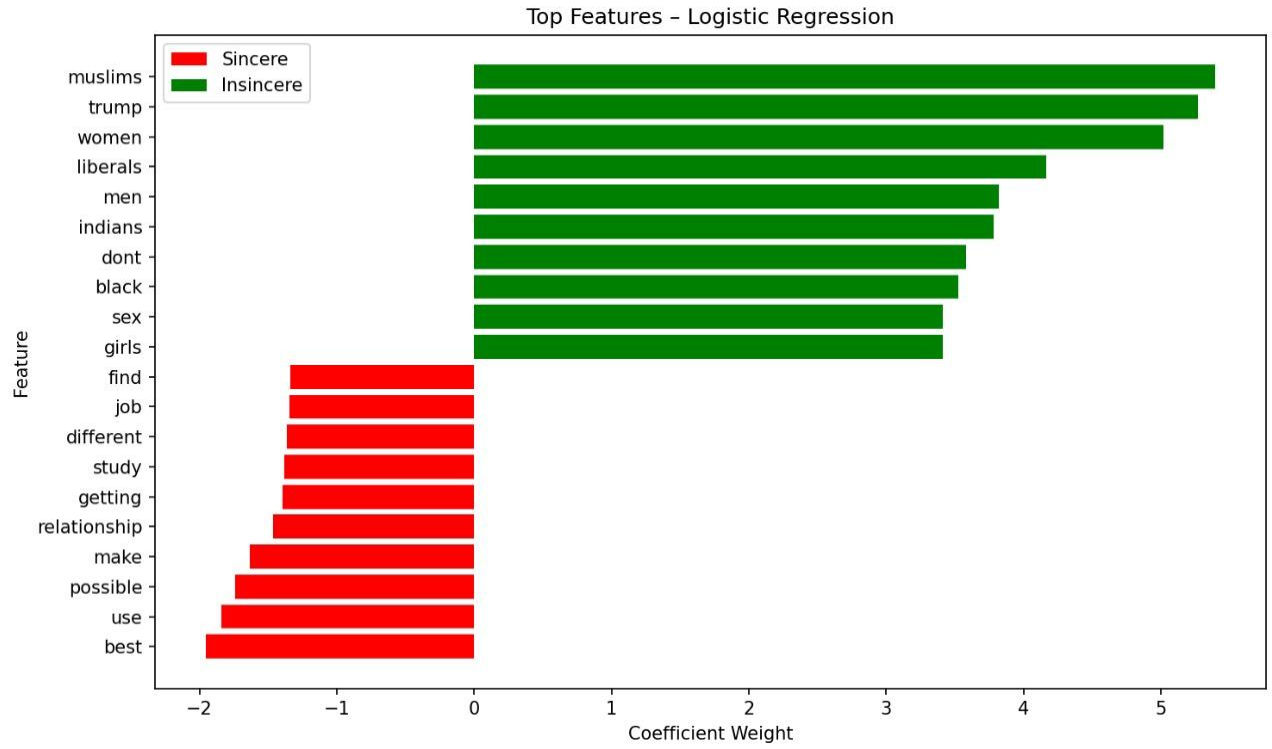
Team ID : 16

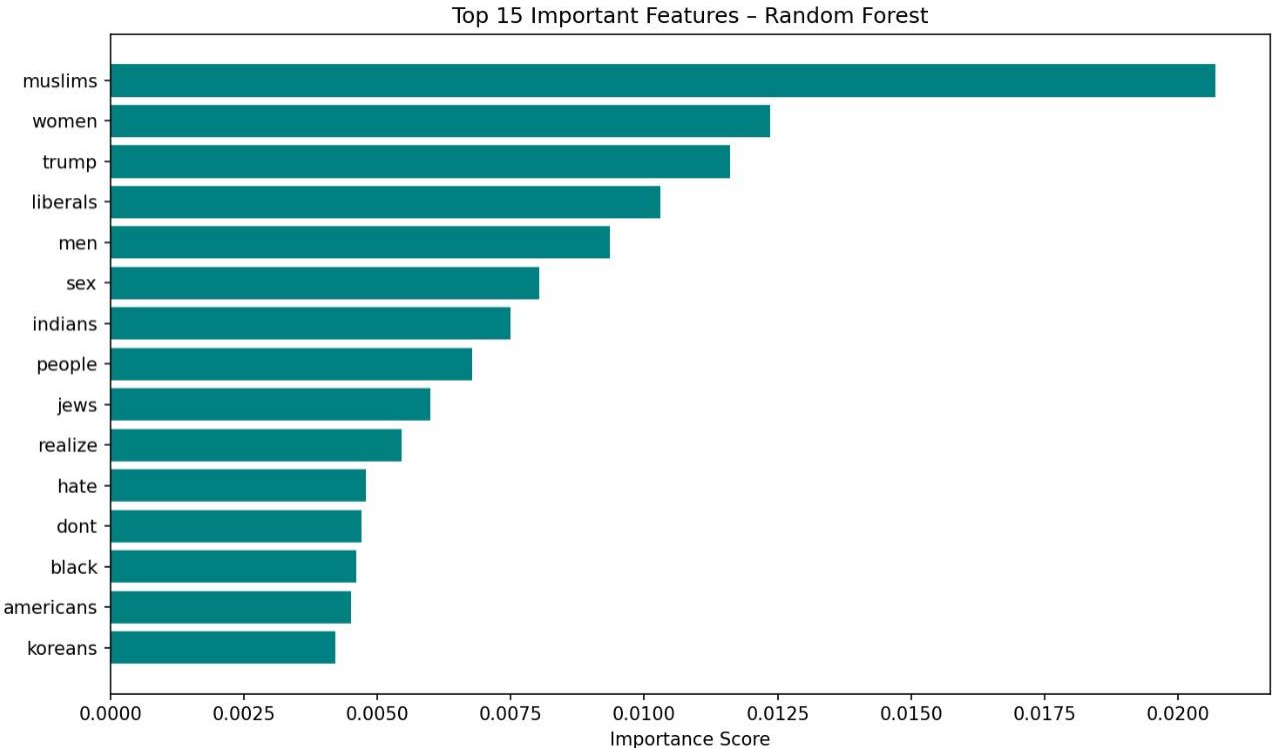
```
PS C:\Users\Chetan\Downloads\ML_MINI_PROJECT\Insincere_Question> python main.py
Loading dataset from train.csv...
 Dataset Loaded: 10000 samples
 Cleaning text data...
 Applying TF-IDF vectorization...

 Training Logistic Regression...
Logistic Regression done → Accuracy: 89.9%, Time: 0.16s

 Training Random Forest...
Random Forest done → Accuracy: 93.35%, Time: 4.78s

 Training Light Neural Network...
C:\Users\Chetan\AppData\Local\Programs\Python\Python312\Lib\site-packages\sklearn\neural_network\_multilay
er_perceptron.py:781: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (20) reached and the op
timization hasn't converged yet.
  warnings.warn(
Light Neural Network done → Accuracy: 92.55%, Time: 17.63s

Model Performance Summary:

| Model               | Accuracy | Precision | Recall | F1 | FP | FN | Train Time (s) |
|:--------------------|:---------|:----------|:------:|:---:|:---:|:---:|:---------------:|
| Logistic Regression | 89.9%    | 0.35      | 0.52   | 0.42 | 135 | 67 | 0.16           |
| Random Forest       | 93.35%   | 0.71      | 0.09   | 0.15 | 5   | 128 | 4.78           |
| Light Neural Network | 92.55%   | 0.43      | 0.21   | 0.28 | 38 | 111 | 17.63          |

Best Performing Model: Logistic Regression with F1-Score = 0.42
```

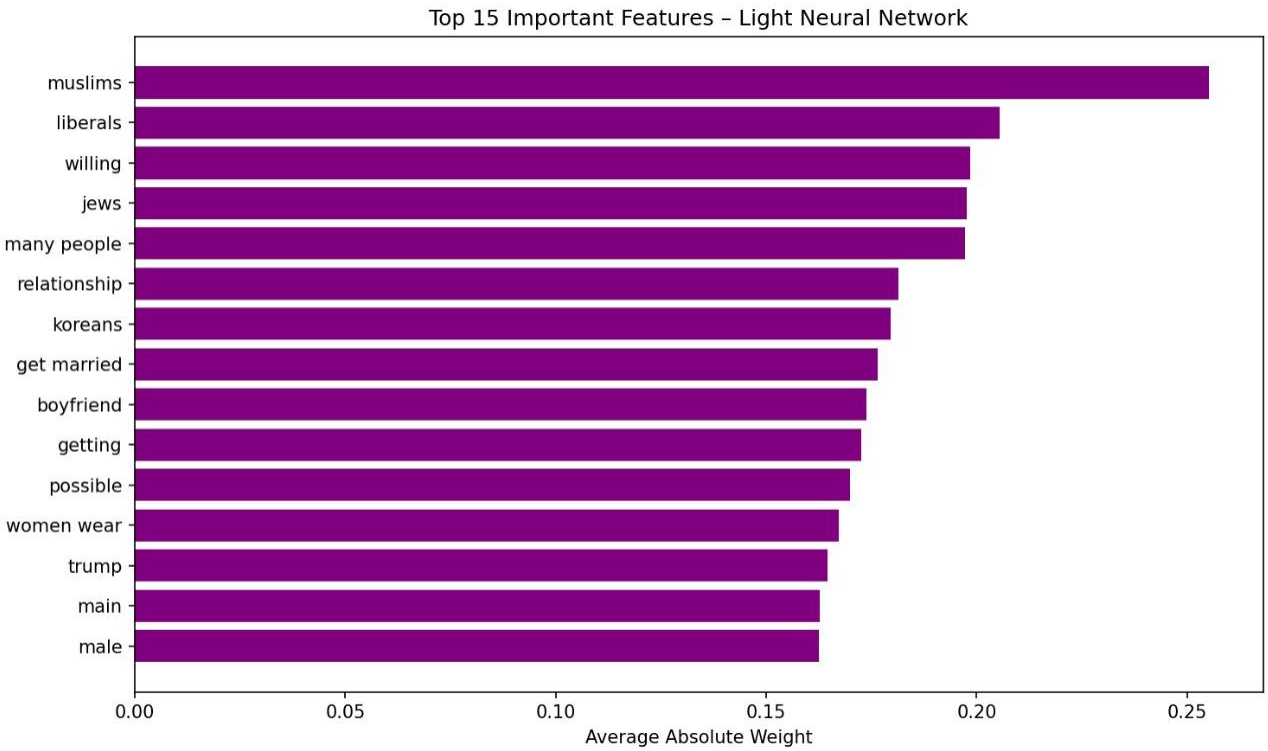
```
 Detailed Classification Report:

              precision    recall  f1-score   support

     Sincere       0.96      0.93      0.94      1860
   Insincere       0.35      0.52      0.42       140

    accuracy                           0.90      2000
   macro avg       0.66      0.72      0.68      2000
weighted avg       0.92      0.90      0.91      2000


Model and vectorizer saved successfully → Logistic Regression

Training, Evaluation, and Visualization Completed Successfully!
```


Model Performance Comparison
Accuracy
Precision
Recall
F1-Score
Metric Score
0
20
40
60
80
Logistic Regression
Random Forest
Light Neural Network


Confusion Matrix – Logistic Regression
1725
135
67
73
Sincere
Insincere
True Label
Sincere
Insincere
Predicted Label
1600
1400
1200
1000
800
600
400
200


Top Features – Logistic Regression
Sincere
Insincere
muslims
trump
women
liberals
men
indians
dont
black
sex
girls
find
job
different
study
getting
relationship
make
possible
use
best
Feature
−2
−1
0
1
2
3
4
5
Coefficient Weight


Top 15 Important Features – Random Forest
muslims
women
trump
liberals
men
sex
indians
people
jews
realize
hate
dont
black
americans
koreans
0.0000
0.0025
0.0050
0.0075
0.0100
0.0125
0.0150
0.0175
0.0200
Importance Score


Top 15 Important Features – Light Neural Network
muslims
liberals
willing
jews
many people
relationship
koreans
get married
boyfriend
getting
possible
women wear
trump
main
male
0.00
0.05
0.10
0.15
0.20
0.25
Average Absolute Weight

Thank You