Fake News Detection
using ML tools

Arkaprabha Majumdar   &   Ankit Singh

pravaa8@gmail.com   ,   gahlonankit@yahoo.in
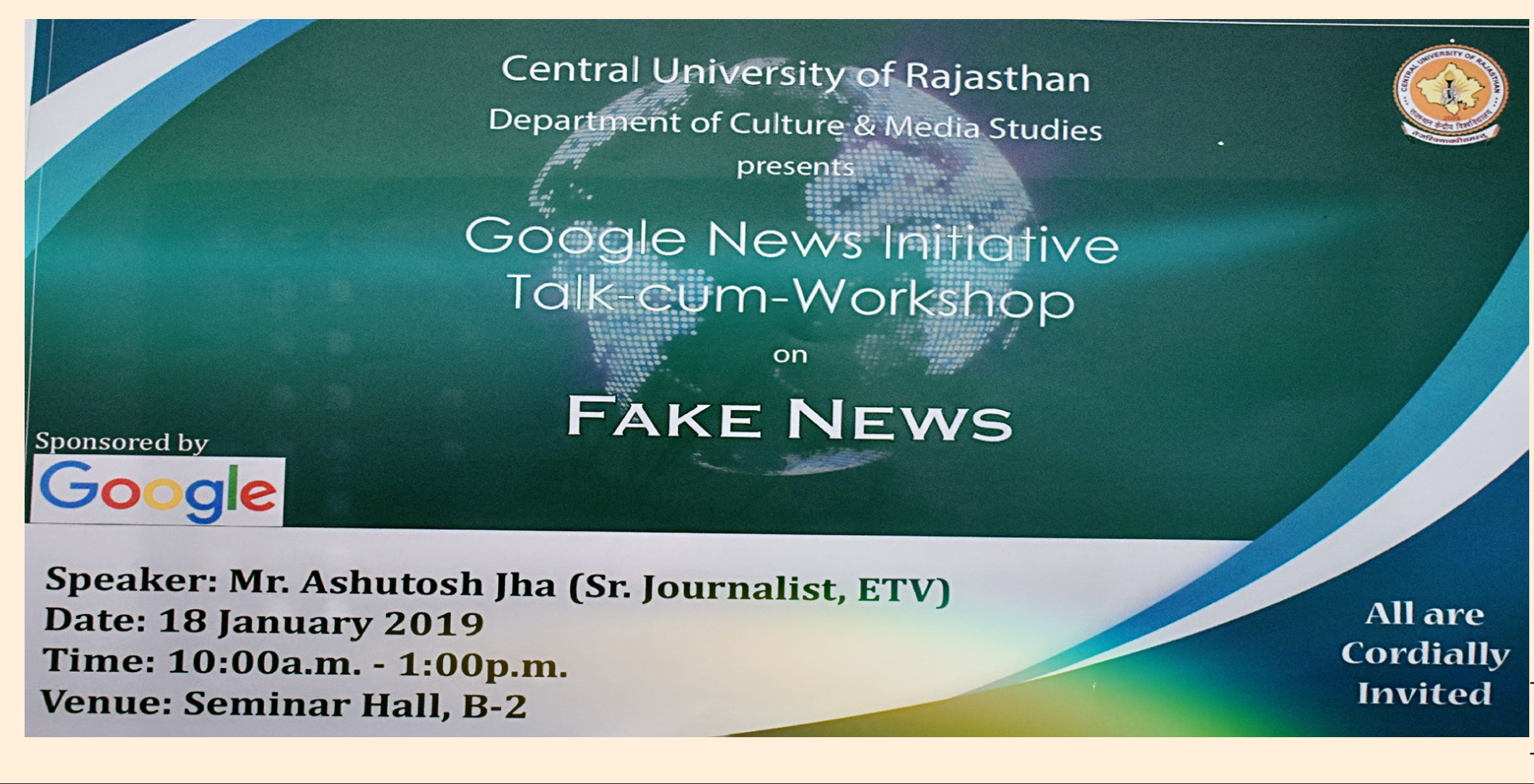
Supervised by: SK Choudhury (Asst. Prof, Dept of Data Science)

DATASCIENCE

CENTRAL UNIVERSITY OF RAJASTHAN

Abstract

Owing to the large influx of social media and data from around the globe, it is proving very difficult for the ordinary layman to differentiate between real news and fake news. Many studies by institutions like Stanford have shown that students were only able to judge fake news with poor accuracy. Hence, in the coming times we feel it is a necessity to create an algorithm that can identify fake news with a high accuracy, and Machine Learning tools have started making this a possibility.

Motivation

A Google Initiative seminar by the Dept. Of Cultural & Media Studies,CURAJ on the 18th of January.
Why we need the fake news detection technology?

Central University of Rajasthan
Department of Culture & Media Studies
presents
Google News Initiative
Talk cum-Workshop
on
FAKE NEWS
Sponsored by
Google
Speaker: Mr. Ashutosh Jha (Sr. Journalist, ETV)
Date: 18 January 2019
Time: 10:00a.m. - 1:00p.m.
Venue: Seminar Hall, B-2
All are Cordially Invited

Models

- **Multinomial NB**
- **PassiveAgressiveClassifier**
- **Hash Vectorizer + PassiveAggressive**
- **Logistic Reg**
- **Random Forest**
- **XgBoost**

References:

[1] Julio Amador Diaz Lopez , Axel Oehmichen, Miguel Molina-Solana *Characterizing Political Fake News in Twitter by its Meta-data* Imperial College, London

[2] Manisha Gahirwal, Sanjana Moghe, Tanvi Kulkarni, Devash Khakhar, Jayesh Bhatia *Fake News Detection*, International Journal of Advanced Research

Technologies Used

matplotlib

Natural Language Analysis with Python NLTK

python

scikit learn

python Natural Language Analyses with NLTK

Data & Pre-processing

Dataset = Demonetization (KAGGLE.com)   FakeNews Twitter Data (arxiv.com)
Data pre-processing is an important step in the data mining process. It involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data pre-processing is a proven method of resolving such issues.
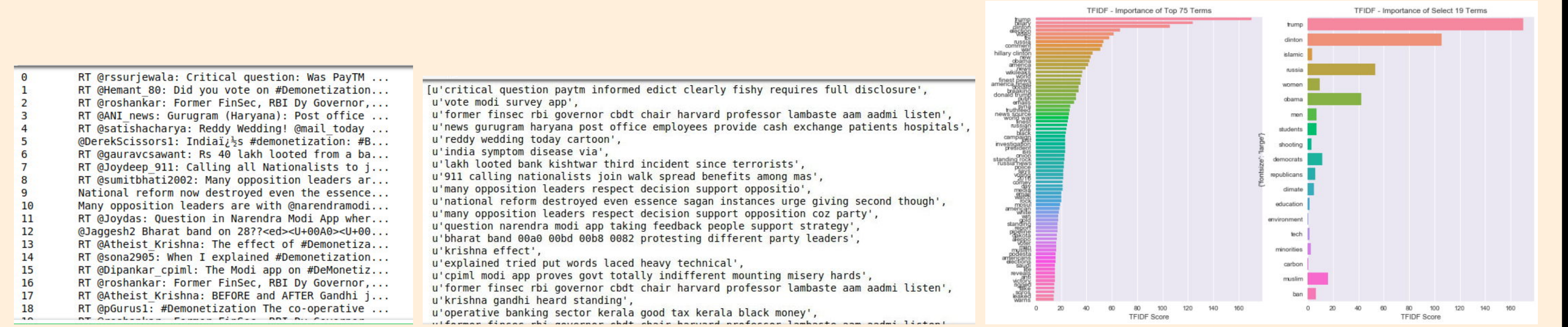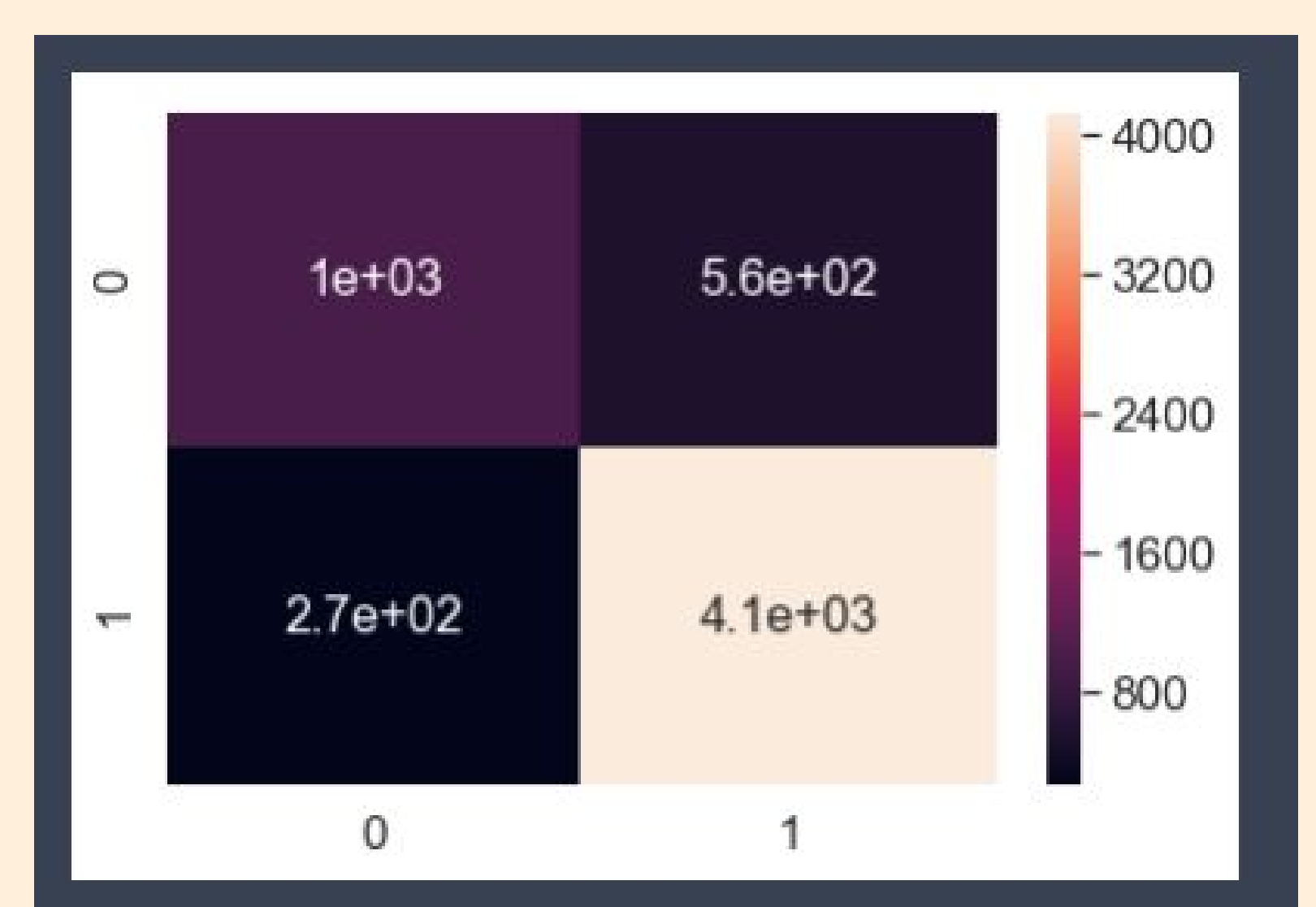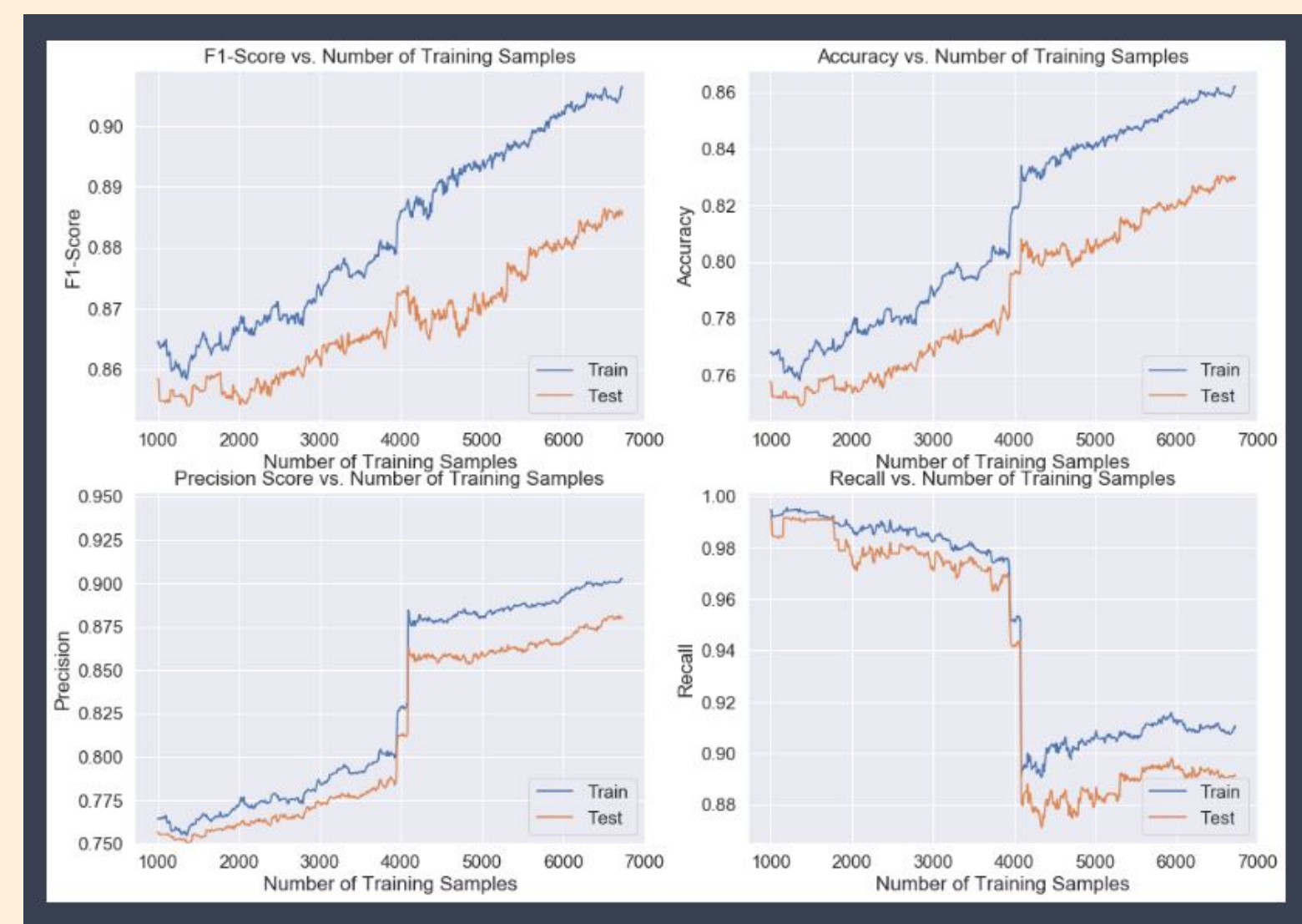

Result

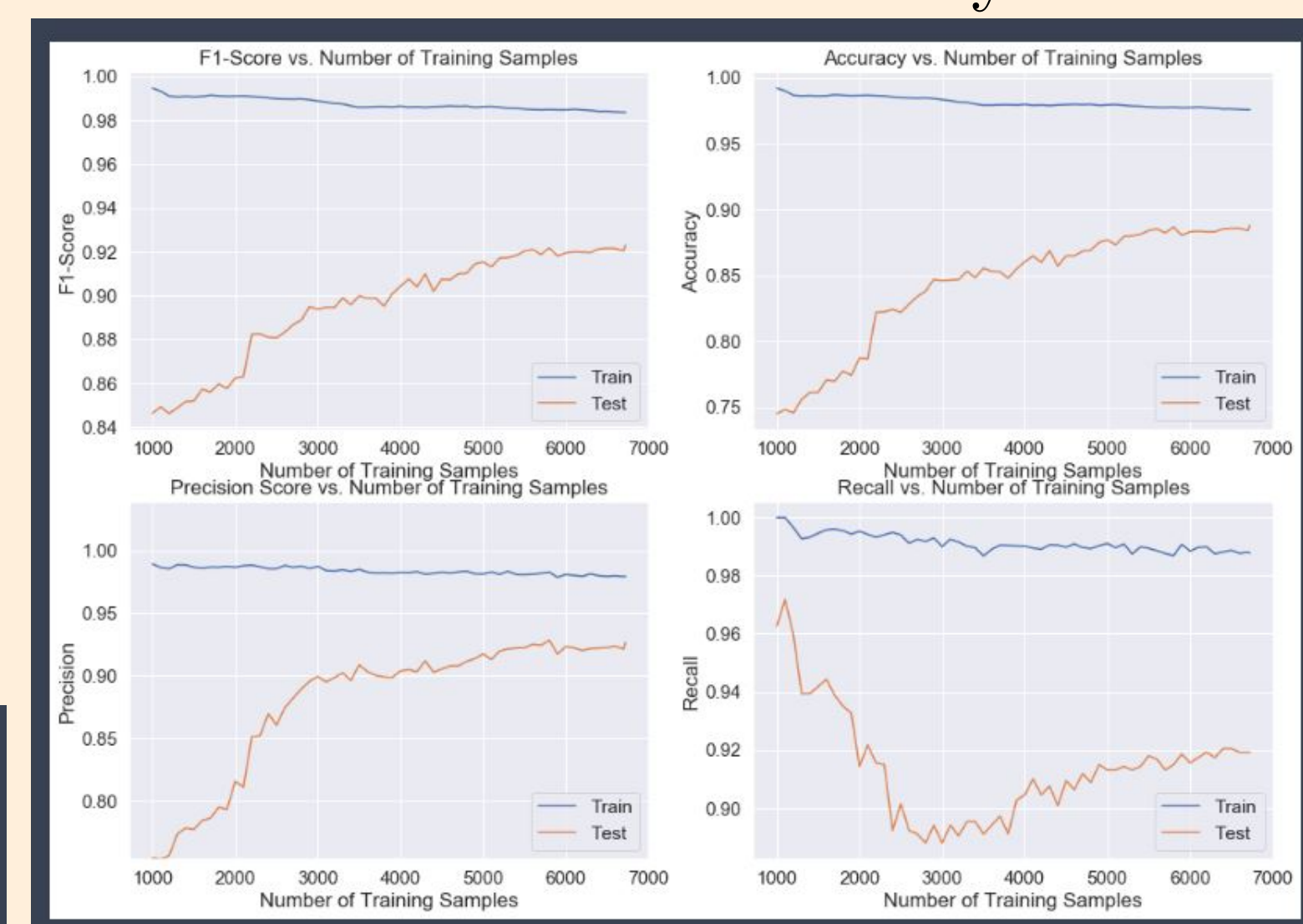- **Multinomial _NB**: Accuracy=81.058

PassiveAggressive: Accuracy=86.496

- **LogisticRegression**: Accuracy=84.62

XgBoost: Accuracy=77.79

- **RandomForest**: Accuracy=90.16

Logistig Regression F1 and Accuracy Scores :
F1 score 79.64%
Accuracy score 84.67%

Random Forest F1 and Accuracy Scores :
F1 score 86.55%
Accuracy score 89.68%

final: 90.16 % Accuracy

Conclusion

In the 21st century, the majority of tasks are done online. Newspapers who were earlier preferred as hard copies are now being substituted by applications like FB, Twitter and news articles to be read online. The growing problem of fake news only makes things more complicateda dn tries to change or hamper the opinion and attitude of people towards use of digital technology. When a peraon is deceived by the real news, two possible things happen: people start believing that their perception about a particular topic are true as assumed. Another problem is that if there is any news article available which contradict a supposedly fake one, people belive in the words which just support their thinking. Without taking in the measure of the fact involved.
Thus, in order to curb the phenomenon, Google and fb are taking their step towards preventing the fake news