



SAPIENZA
UNIVERSITÀ DI ROMA

Big Data Computing 2021-2022

Fake News Detection using Pyspark

A presentation by: Albiona Guri



Outline

1. Introduction
2. Dataset
3. Models
4. Experimental results
5. Conclusion
6. A functional Application



Introduction

- ❖ The main objective of our project is to determine if a given text data by the user is fake or real.
- ❖ It consists of a binary classification problem.
- ❖ In order to automate the news qualifying procedure we implemented 4 machine learning algorithms to predict the news.



Dataset

- ❖ The [dataset](#) we used is derived from Kaggle resource and is composed by text data.
- ❖ It consists of around 44900 training examples.
- ❖ The ratio between minority class and majority is particularly small.



Feature Engineering

Pre-Processing





Models

1. Logistic Regression
 - Logistic Regression
2. Random Forest
 - Random Forest Classifier
3. Multilayer Perceptron
 - MultilayerPerceptronClassifier
4. Transformers
 - Hugging Face Transformer

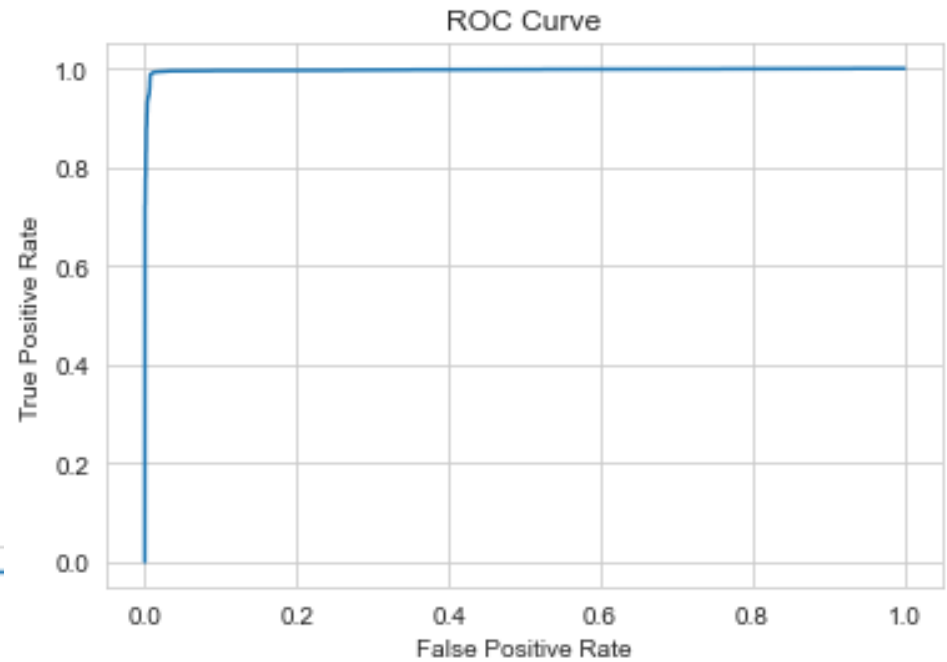
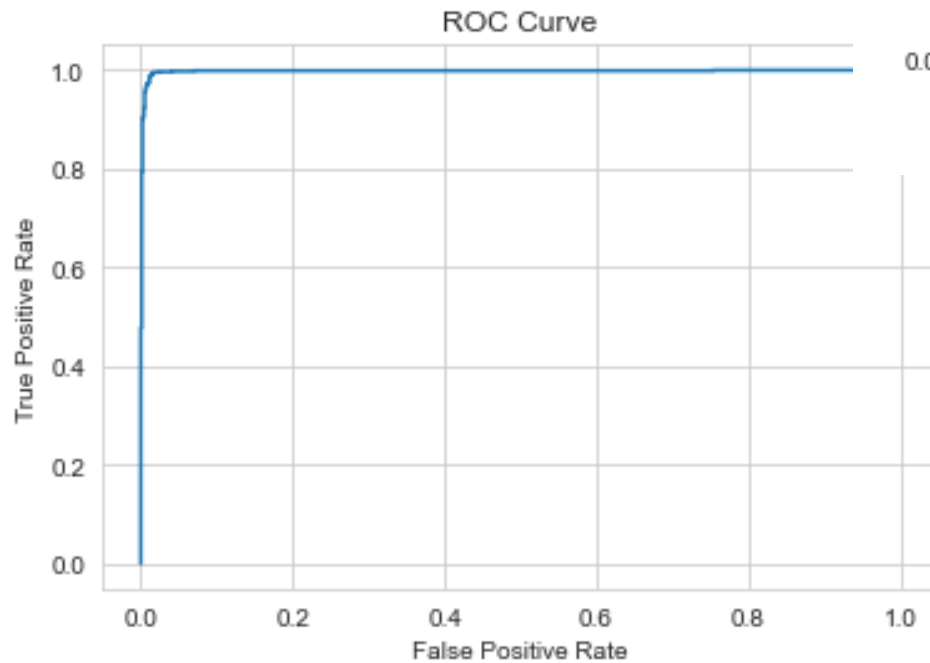


SAPIENZA
UNIVERSITÀ DI ROMA

Experimental Results



❖ Logistic Regression

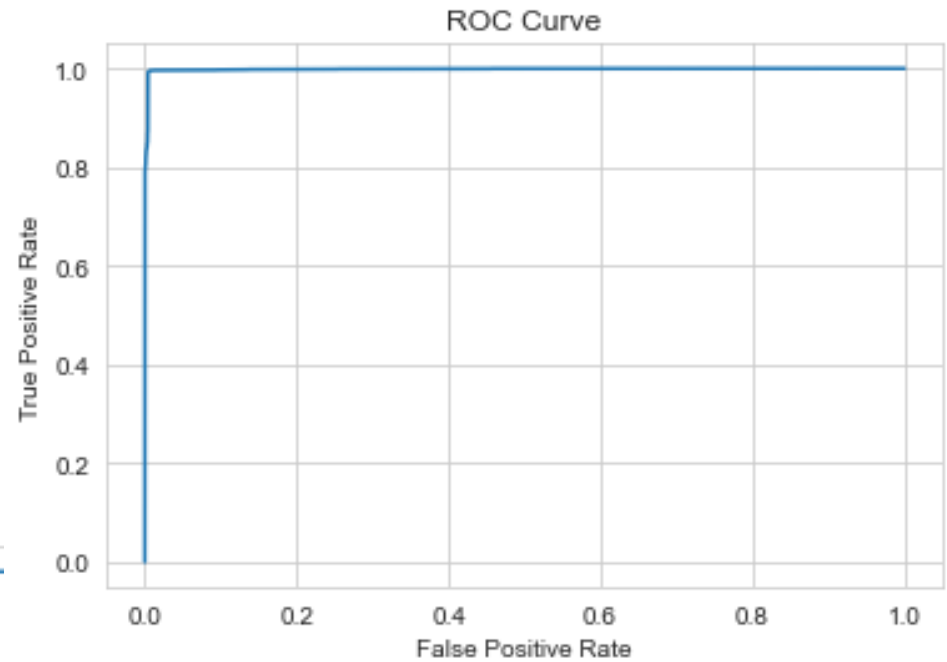
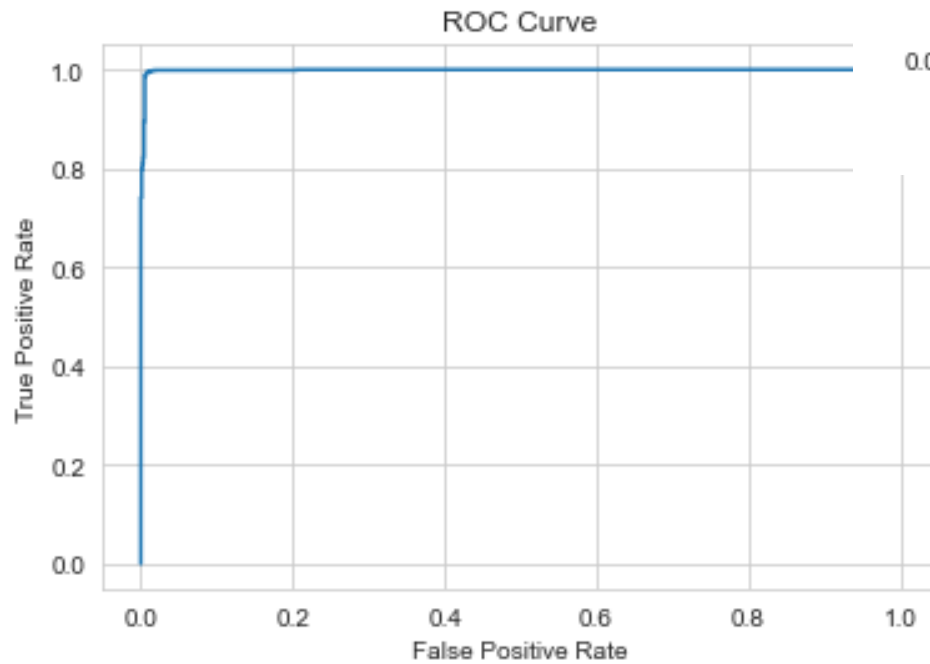


Train set

Test set



❖ Random Forest

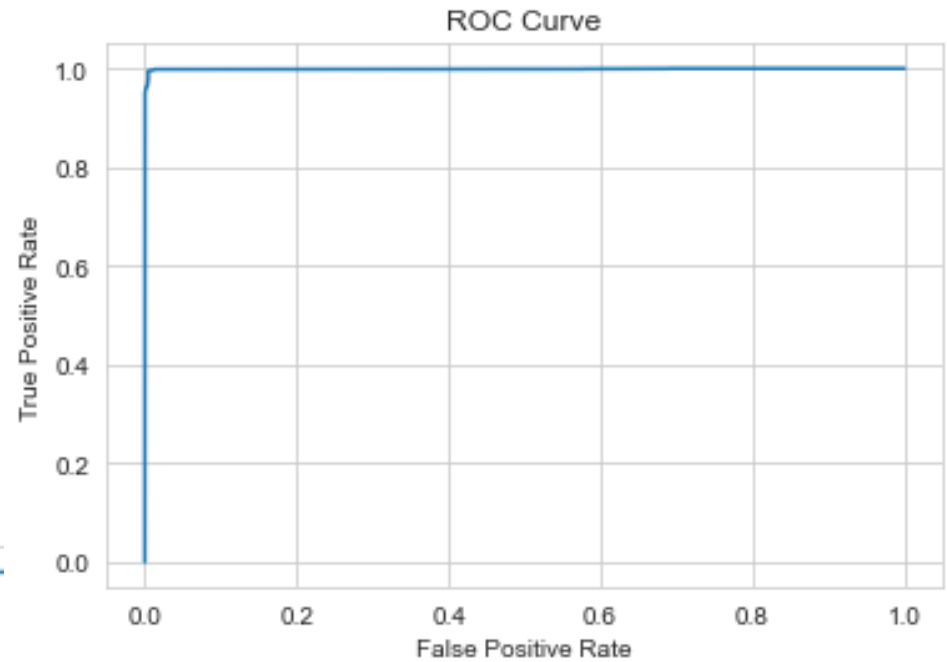
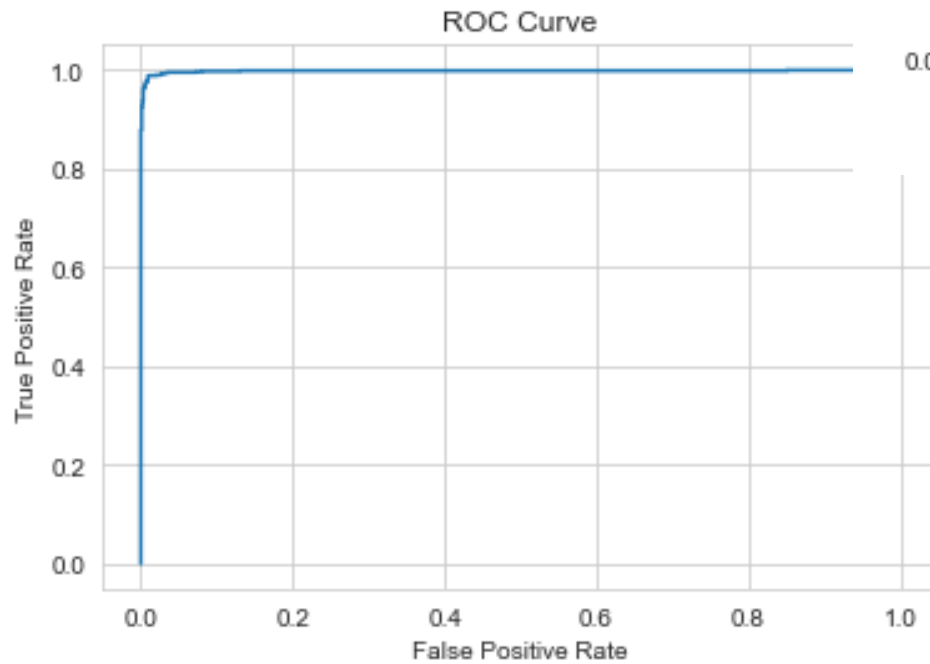


Train set

Test set



❖ MultiLayer Perceptron



Train set

Test set



Results Summary

Metrics	Logistic Regression	Random Forest	Multilayer Perceptron	Transformers
Accuracy	0.98831	0.99309	0.98353	0.99867
F1-Score Class 0	0.98851	0.99323	0.98381	-
F1-Score Class 1	0.98810	0.99295	0.98325	0.99587
AUC	0.99678	0.99876	0.99717	0.99852



Conclusion

- ❖ Among all tested models in our dataset, we noticed that the best performance comes from:

Random Forest and Transformers

- ❖ We built a small Application that can be used to get informed if the news are real or fake. The used model in our specific case is Transformers.
- ❖ We have saved all the models, so we can adapt the Application for using each of them.



A functional Application

- ❖ To put into practice our model, we built a small application where the user can enter the news and receive a feedback whether it is FAKE or REAL
- ❖ Technologies used:
 - Flask
 - React.js
 - Docker
- ❖ Links to the docker images:
 - <https://hub.docker.com/repository/docker/elvinalika/fake-news-detector-fe>
 - <https://hub.docker.com/repository/docker/elvinalika/fake-news-detector-be>



Fake or Real News

Enter your NEWS

CHECK NEWS

