

EARLY FAKE NEWS DETECTION

Presented by:

Abdelrahman Gamal

Mariam Khaled

Radwa Mostafa

John Gerges

Supervised by:

Dr. Daaa Salama

Assisted by:

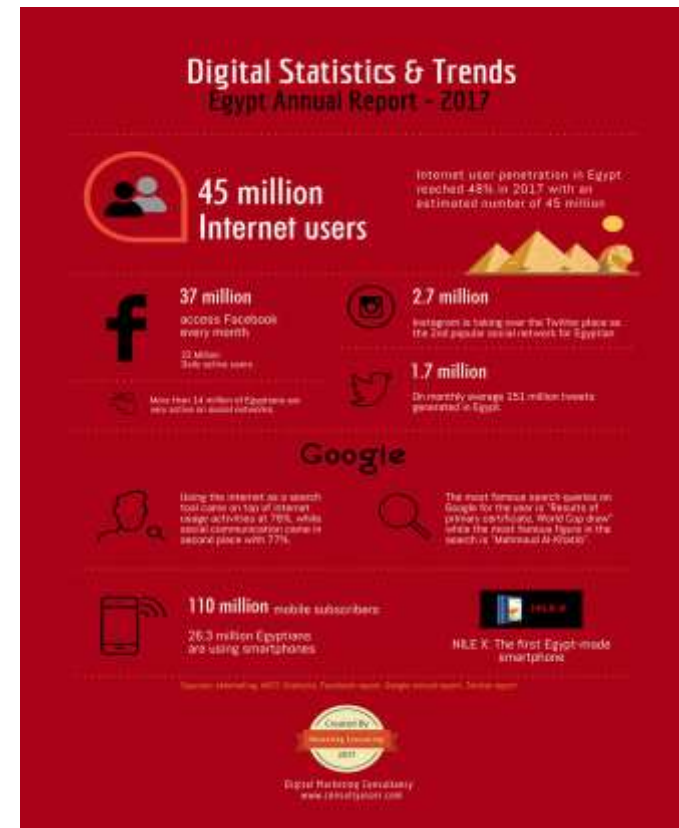
Eng.Youmna Ibrahim

AGENDA

- Introduction
- Market motivation
- Problem statement
- System overview
- Expected results

INTRODUCTION(1/2)

- Nowadays Social Media has been taking a huge part in people's life and is the most common place for fake news such as Facebook and Twitter As it has no restrictions on people's posts as anyone can edit their posts whenever they want.
- However, that didn't stop people from using Social Media as their source.
- Fake news is one of the most common critical types of deception.



INTRODUCTION(2/2)

- Studies have shown that human ability to differentiate between fake news and real ones are only 54%
- Machine Learning is the best way to enhance the accuracy of this detection.



MOTIVATION

- The spread of economical and political fake news can directly affect the stock market.
- Tourism income can also be affected by fake news, as spreading fake news in countries can make tourists not interested in visiting certain places which costs losing a lot of income.
- As reported by the NBC teenagers earned at least \$60,000 due to spreading fake news and their expected wage is \$4,800 which motivates them to increase spreading

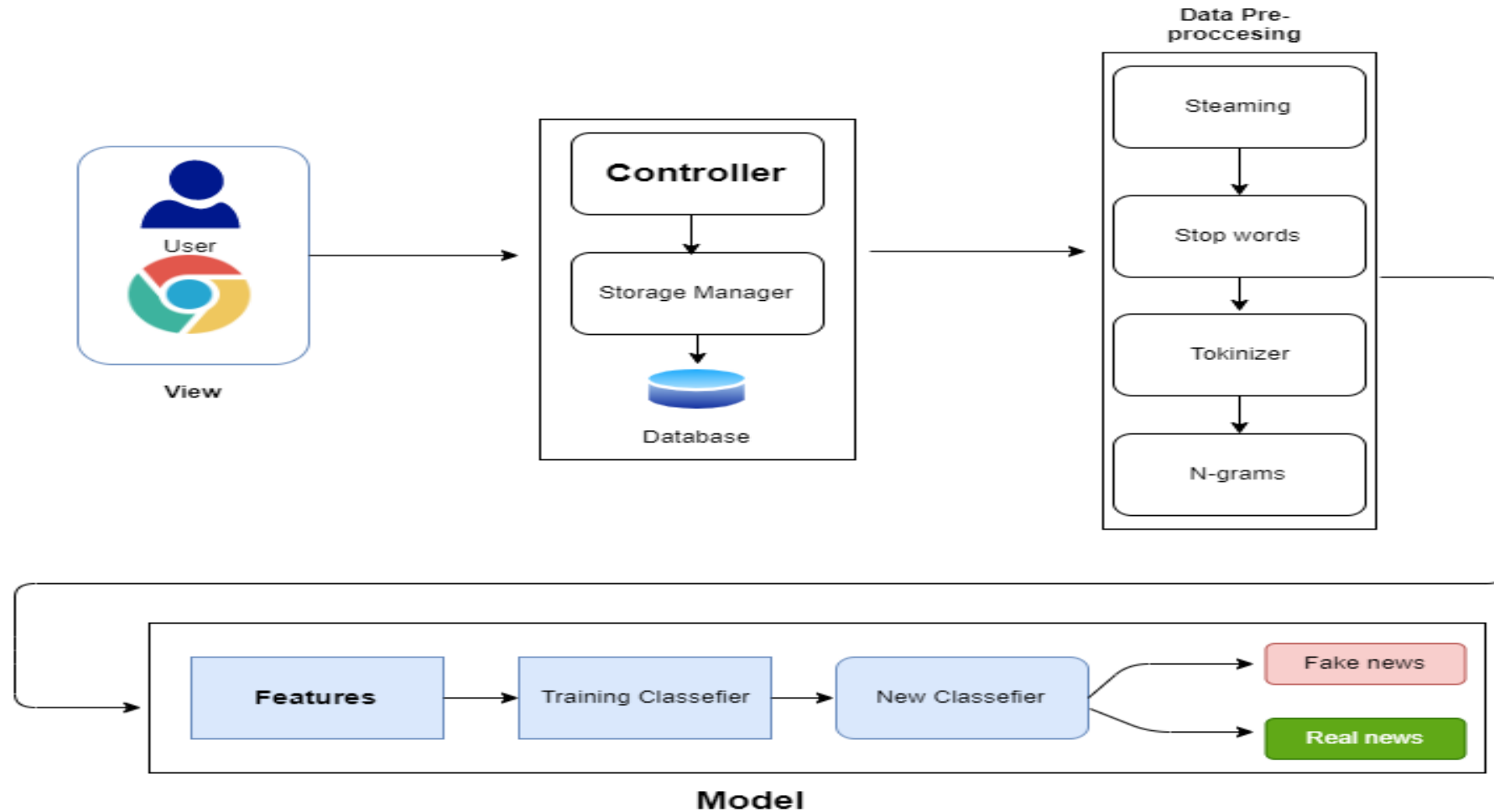
PROBLEM STATEMENT

Our aim is to build a website to keep track of fake news on twitter in an interval of time because social media became a reliable source for news and it has no restrictions so it affects the economy of the country . The output on this website will help the user differentiate between fake and real news according to a certain category and in a specific interval of time using machine learning and a combination of the best two algorithms.

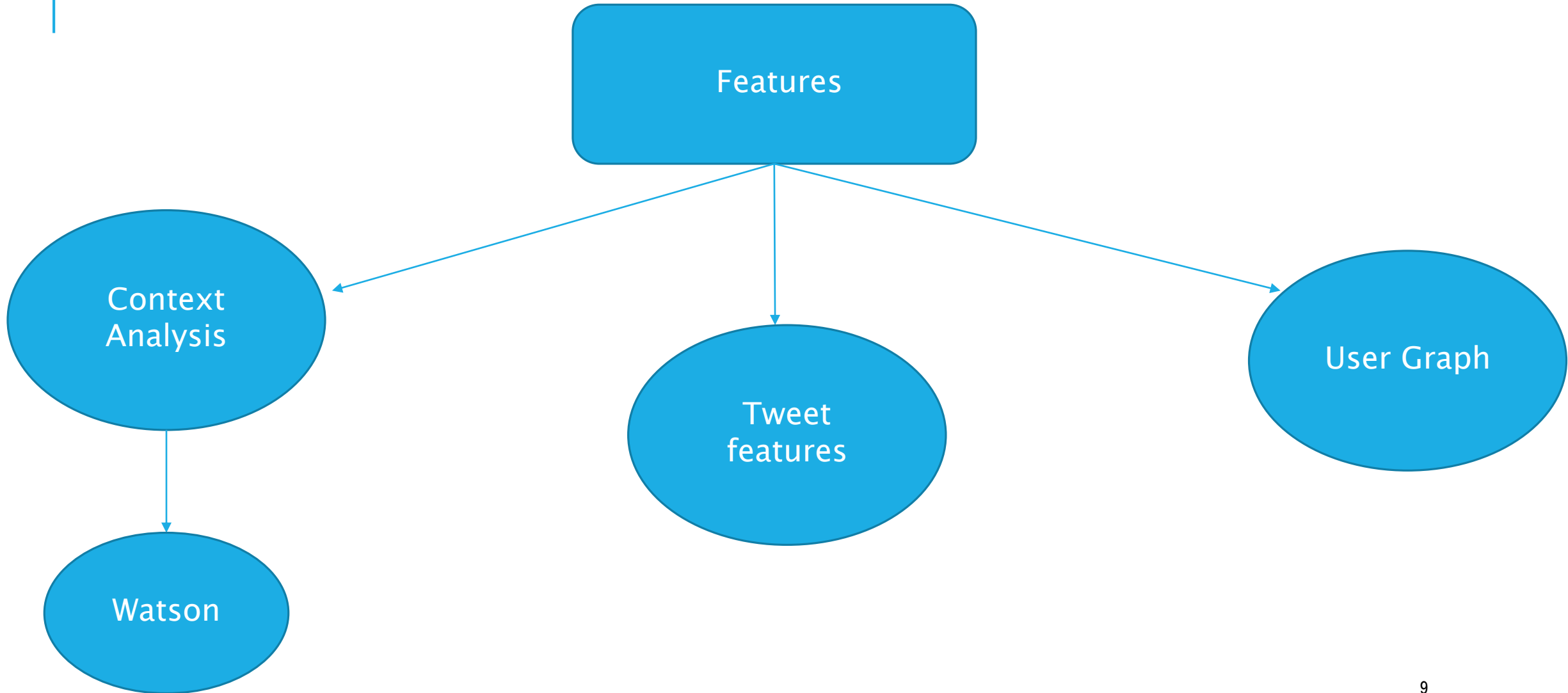
CHALLENGES

1. Detecting news from social media (Twitter)
2. Finding trained dataset.
3. Enhancing accuracy in detecting the Fake news.
4. Combining multi- classifiers.

SYSTEM OVERVIEW



FEATURES



RELATED WORK(1/3)

Technique	ACC	PRE	REC	F-M
LSTM	82.29	44.35	40.55	40.59
LSTMDrop	73.78	39.67	29.71	30.93
LSTM-CNN	80.38	43.94	39.53	39.70

- ❖ **Aim** : Identifies relevant features associated with fake news stories without previous knowledge of the domain.
- ❖ **Dataset** : CharlieHebdo • SydneySiege • Ottawa Shooting • Germanwings–Crash • Ferguson Shooting
- ❖ **Classifiers**: LSTM ,LSTMDrop, LSTM-CNN

[3] Oluwaseun Ajao, Deepayan Bhowmik, and Shahrzad Zargari. 2018. Fake News Identification on Twitter with Hybrid CNN and RNN Models. In Proceedings of the International Conference on Social Media & Society, Copenhagen, Denmark (SMSociety). 1 DOI: <https://doi.org/10.1145/3217804.3217917>

RELATED WORKS(2/3)

- **Entitled by:** Identifying Tweets With fake news
- **Aims:** This papers aims to identify fake news by making users analysis and context analysis by using NLP.
- **Dataset:** Dataset containg husseican candy and another one containing
- **Classifier:** SVM.
- **Accuracy:**62%

[4]Krishnan, Saranya, and Min Chen. "Identifying Tweets with Fake News." *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, 2018.
<https://doi.org/10.1109/iri.2018.00073>.

RELATED WORK(3/3)

- **Aim:** The main idea is to explore applications for NLP techniques to detect fake news that has misleading events from non-reputable sources.
- **Dataset:** Signal Media, OpenSource.co
- The dataset contains about 1 million articles from a variety of news sources these sources include major news outlets, local news sources and blogs
- **Classifiers:** SVM, Bounded Decision tree, Random Forests and Gradient Boosting.

[5] Gilda, Shlok. "Evaluating Machine Learning Algorithms for Fake News Detection." *2017 IEEE 15th Student Conference on Research and Development (SCORED)*, 2017.
<https://doi.org/10.1109/scored.2017.8305411>.

TABLE III: Average model performance with both PCFG and TF-IDF bi-gram features at 0.7 score threshold for categorization.

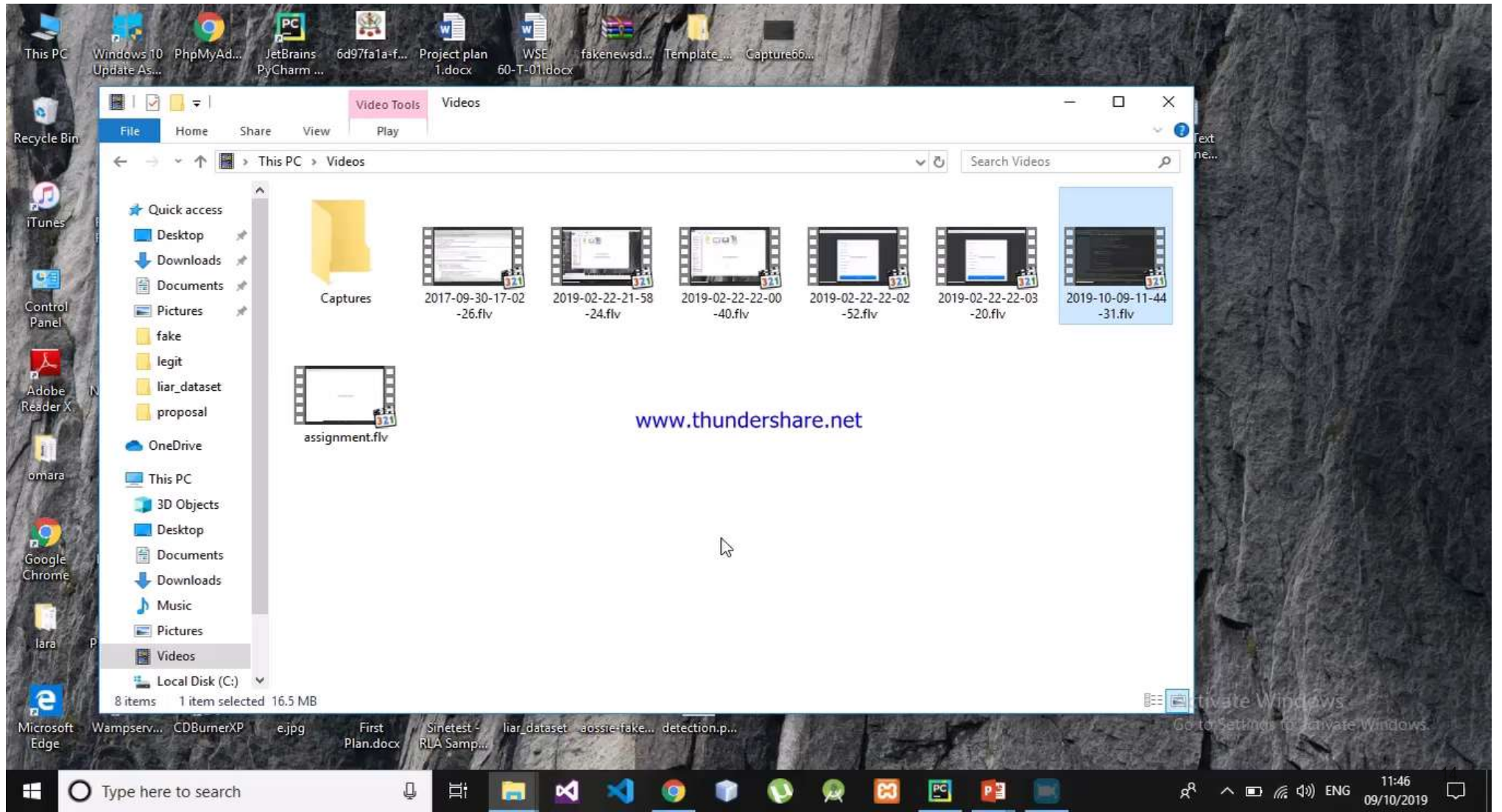
Model	Area Under Curve	Precision	Recall	Accuracy
Bounded Decision Trees	65.9%	66.9%	37.9%	67.6%
Gradient Boosting	75.6%	40.2%	16.1%	65.7%
Random Forests	80.0%	84.2%	18.4%	64.8%
Stochastic Gradient Descent	87.5%	74.1%	71.7%	65.7%
Support Vector Machine	84.3%	80.9%	44.5%	73.6%
Baseline	-	32.18%	32.18%	67.89%

TABLE OF COMPARISON

According to the information in this table we will try to combine every two classifier together then we're going to compare their accuracy and f1-score to develop a hybrid classification that can reach the optimum results.

Paper name	Classifier	Dataset	Accuracy	Reference
Fake News Identification on Twitter with Hybrid CNN and RNN Models	LSTM, LSTM-Drop, LSTM-CNN	CharlieHebdo • SydneySiege • Ottawa Shooting • GermanWings-Crash • Ferguson Shooting	82.29%	Oluwaseun Ajao, Deepayan Bhowmik, and Shahrzad Zargari. 2018. Fake News Identification on Twitter with Hybrid CNN and RNN Models. In Proceedings of the International Conference on Social Media & Society, Copenhagen, Denmark (SMSociety). 1 DOI: https://doi.org/10.1145/3217804.3217917
Identifying Tweets with Fake News."	SVM.	Dataset containing husseican candy and another one containing	62%	[4] Krishnan, Saranya, and Min Chen. "Identifying Tweets with Fake News." 2018 IEEE International Conference on Information Reuse and Integration (IRI), 2018. https://doi.org/10.1109/iri.2018.00073
Evaluating Machine Learning Algorithms for Fake News Detection.	SVM, Bounded Decision tree, Random Forests and Gradient Boosting.	Signal Media, OpenSource.c o	73.6%	[5] Gilda, Shlok. "Evaluating Machine Learning Algorithms for Fake News Detection." 2017 IEEE 15th Student Conference on Research and Development (SCORED), 2017. https://doi.org/10.1109/scored.2017.8305411

DEMO





ANY QUESTIONS ?