

Analyzing Numerical Patterns in Twitter Data: Unveiling Fake and Bot Accounts During Telangana State Elections

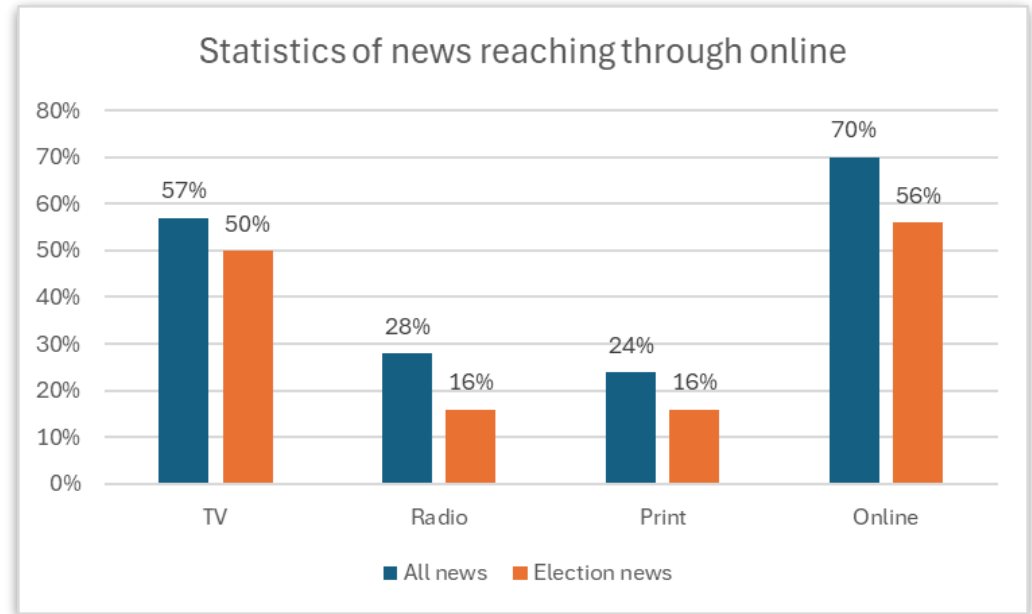
13th IEEE Conference CSNT 2024,
Paper ID: 1570997760



Introduction:

- Online Social Networks (OSNs) are pivotal platforms for disseminating diverse content.
- Proliferation of fake accounts poses a significant challenge, leading to the dissemination of misleading information.
- Fake accounts are rampant across online social networks, posing as legitimate users but created with deceitful intentions.
- These fraudulent profiles are utilized to disseminate deceptive content, ranging from false surveys to fabricated reports on critical events like elections and government actions.
- The presence of fake accounts undermines trust among users and compromises the authenticity of shared information, thereby jeopardizing the credibility of democratic processes and public discourse.

The graph illustrates the prevalence of social media and online communication channels for accessing news, particularly during election periods.



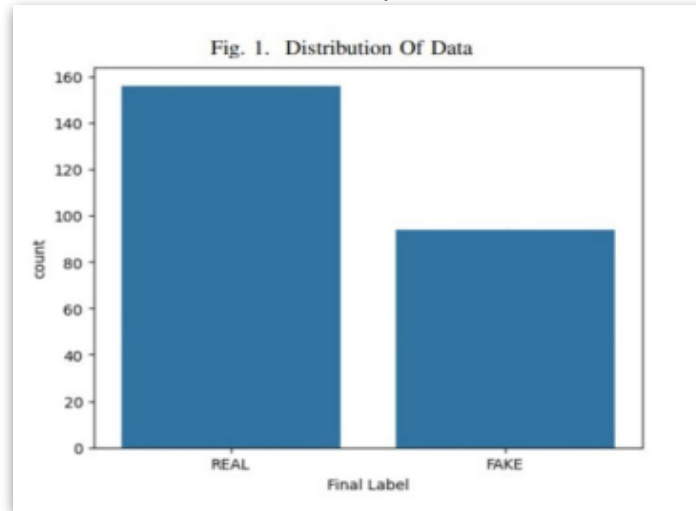
Dataset Description:

1. Training/Testing Dataset Features:

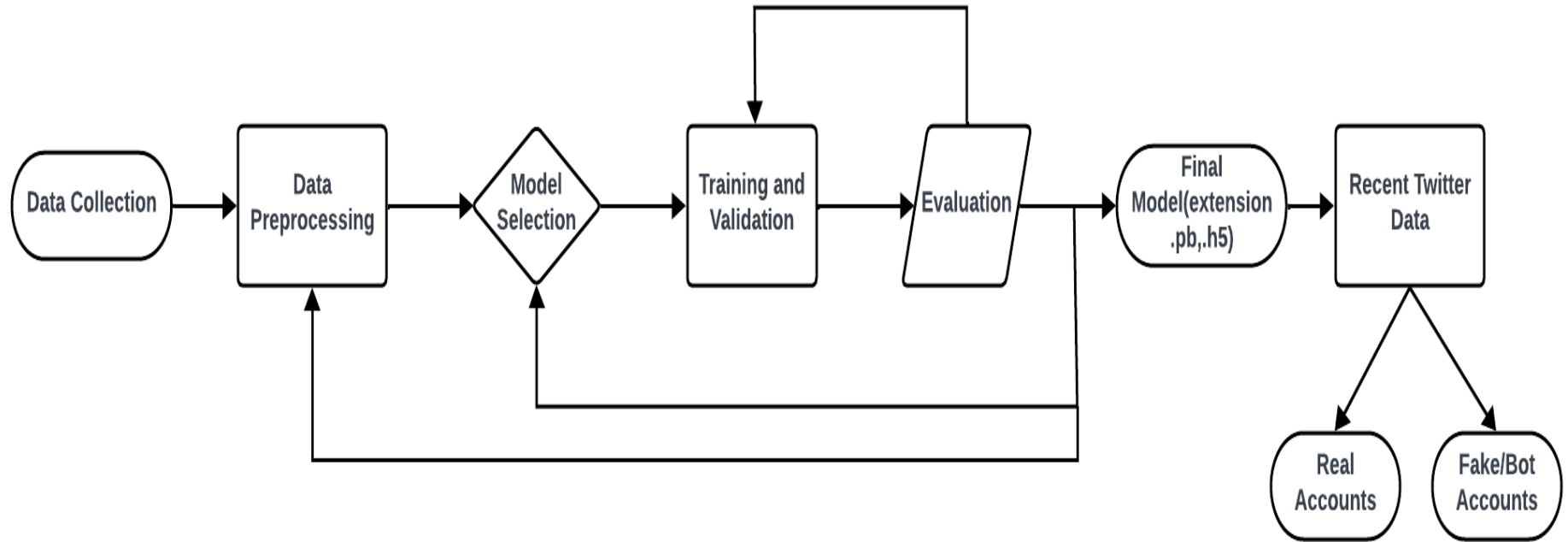
- a. Contains 22 features.
- b. Includes tweet details such as text content, creation timestamp, retweet and favorite counts etc.

2. Dataset Sizes:

- a. Training dataset consists of 250 labelled samples.
- b. Testing dataset comprises 2049 unlabelled samples.

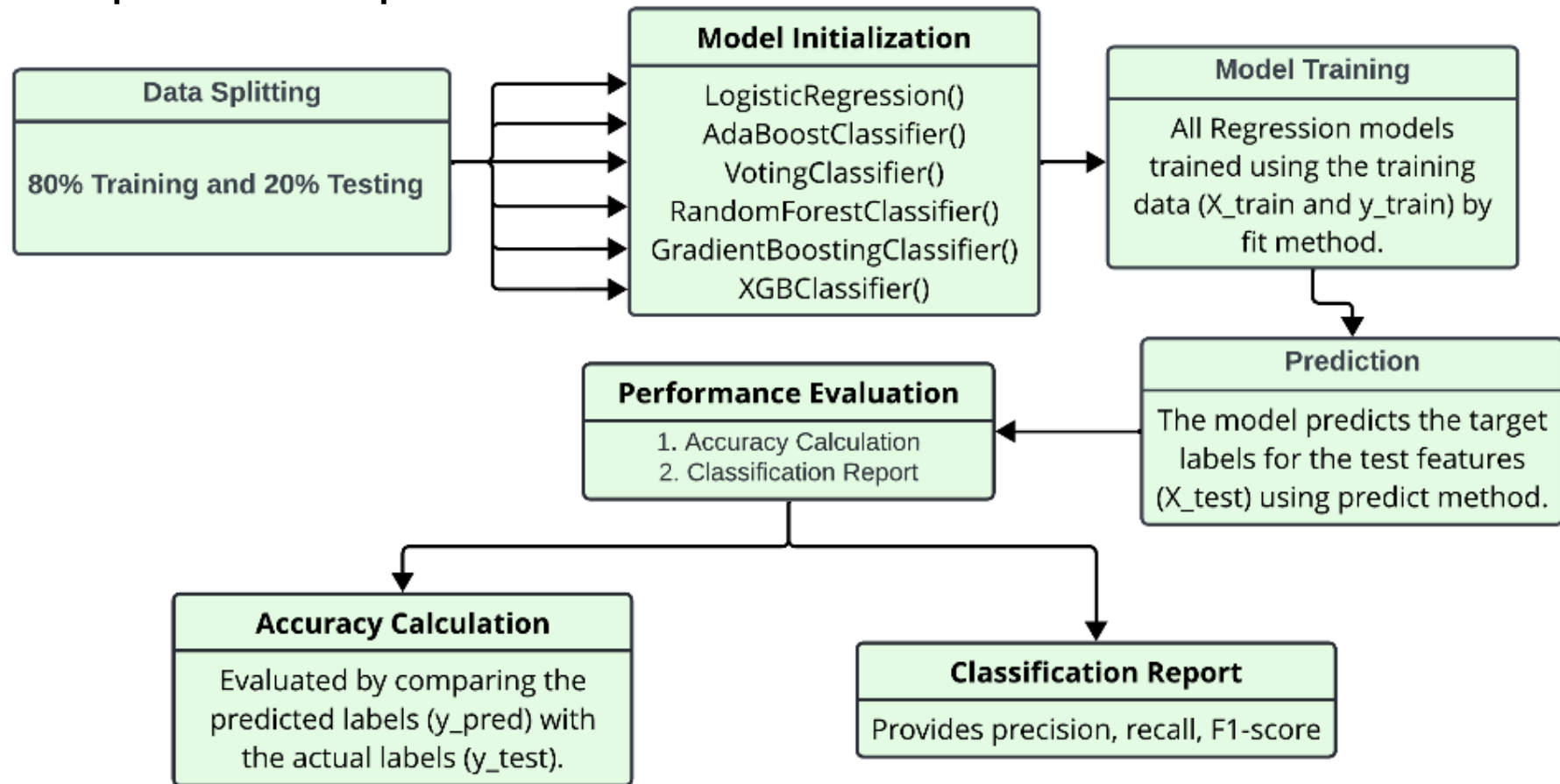


Approach/Methodology:



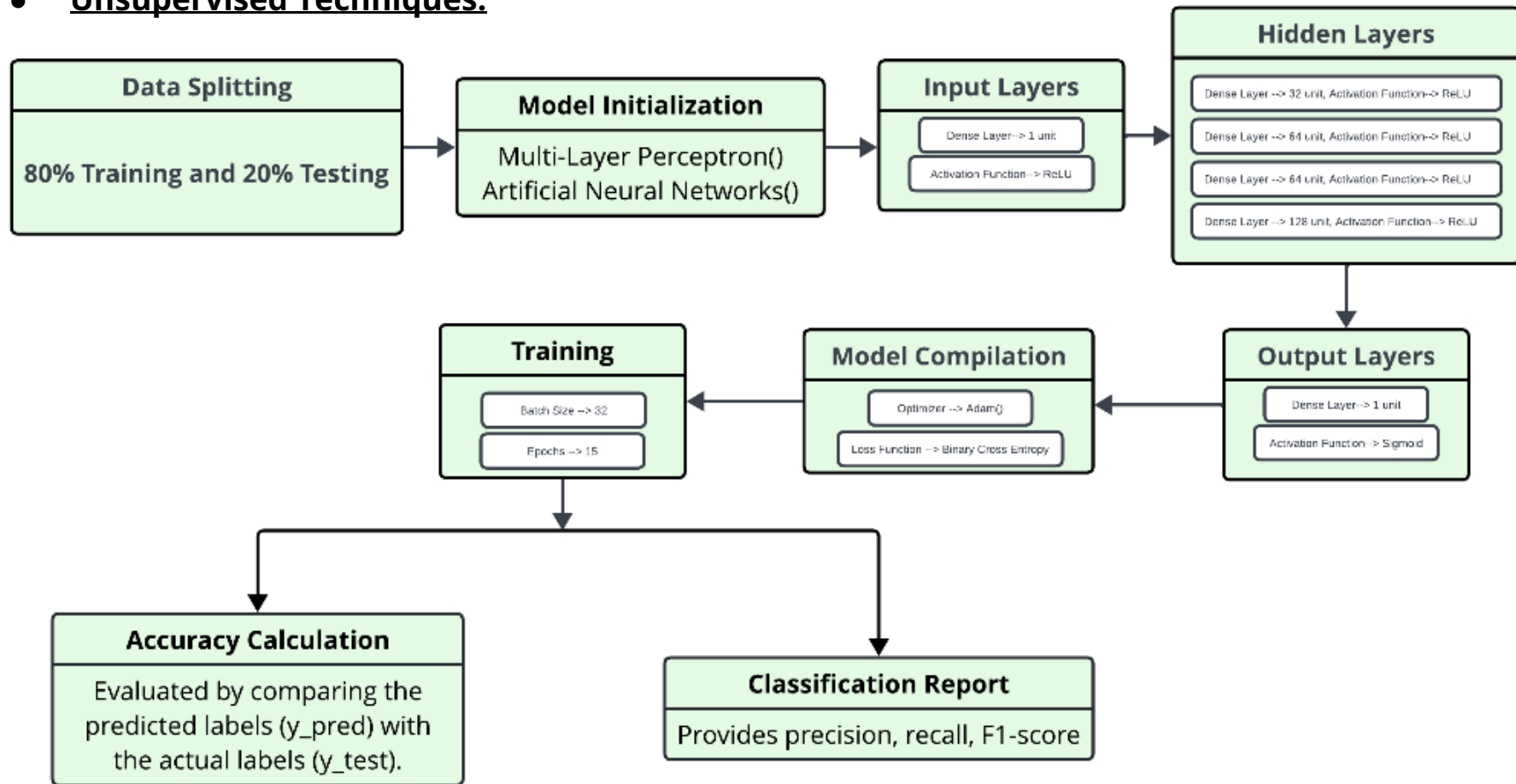
Model Description:

- Supervised Techniques:



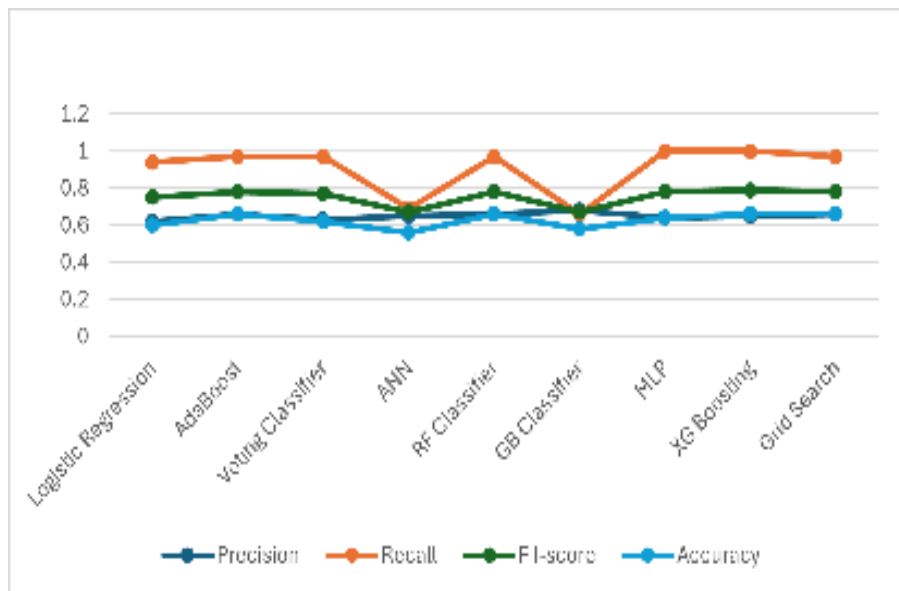
Model Description:

- Unsupervised Techniques:



Model Classification Report

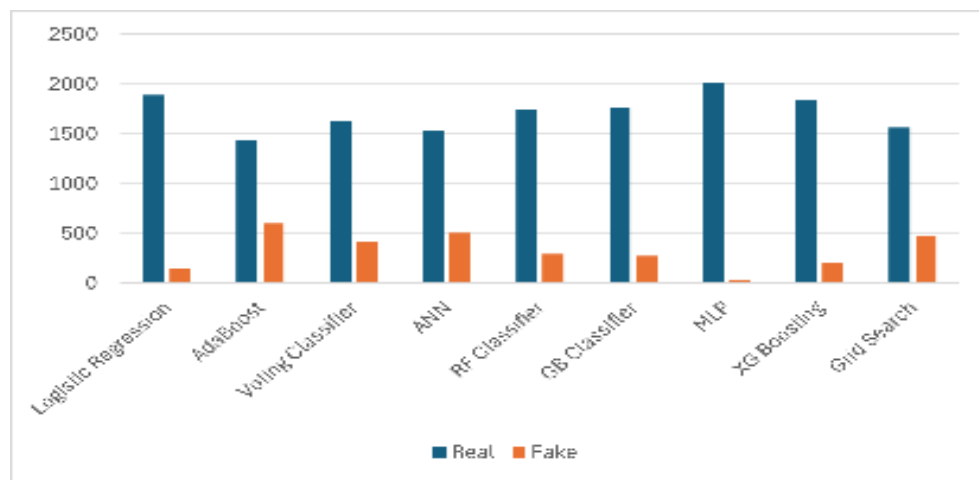
Models	Precision	Recall	F1-score	Accuracy
Logistic Regression	0.62	0.94	0.75	0.6
AdaBoost	0.66	0.97	0.78	0.66
Voting Classifier	0.63	0.97	0.77	0.62
ANN	0.65	0.69	0.67	0.56
RF Classifier	0.66	0.97	0.78	0.66
GB Classifier	0.68	0.66	0.67	0.58
MLP	0.64	1	0.78	0.64
XG Boosting	0.65	1	0.79	0.66
Grid Search	0.66	0.97	0.78	0.66



Model Implementation and Performance

The Model has been tested on the recent Twitter dataset and these are the performance results that we achieved.

Models	Real	Fake
Logistic Regression	1894	154
AdaBoost	1444	604
Voting Classifier	1633	415
ANN	1543	505
RF Classifier	1752	296
GB Classifier	1773	275
MLP	2020	28
XG Boosting	1837	211
Grid Search	1567	481



Conclusion

- Our study delved into Twitter conversations during the Telangana State Elections, employing diverse machine learning models to discern real from fake accounts.
- Despite the array of algorithms utilized, achieving optimal accuracy in distinguishing between genuine and deceptive accounts proved to be a nuanced challenge.
- Analysis fueled by engagement metrics like likes and retweets uncovered patterns hinting at the elusive nature of digital deception.
- Precision and recall metrics shed light on the complexities of maintaining authentic discourse while identifying fraudulent actors.
- Recognizing the dynamic nature of phony accounts, our study highlights the ongoing pursuit of accuracy and invites further exploration into the evolving landscape of social media authenticity.

Future Scope

1. Explore tailored deep learning models for text analysis to capture intricate patterns in social media data.
2. Develop systems for real-time detection of fake and bot accounts to enable timely intervention during live events or trending topics.
3. Extend the analysis beyond Twitter to understand user behavior patterns across various social media platforms.
4. Incorporating user engagement patterns, language analysis, sentiment analysis, and network characteristics to gain richer insights into Twitter account authenticity.



Thank You

