Team Name - Ctrl Alt Elite
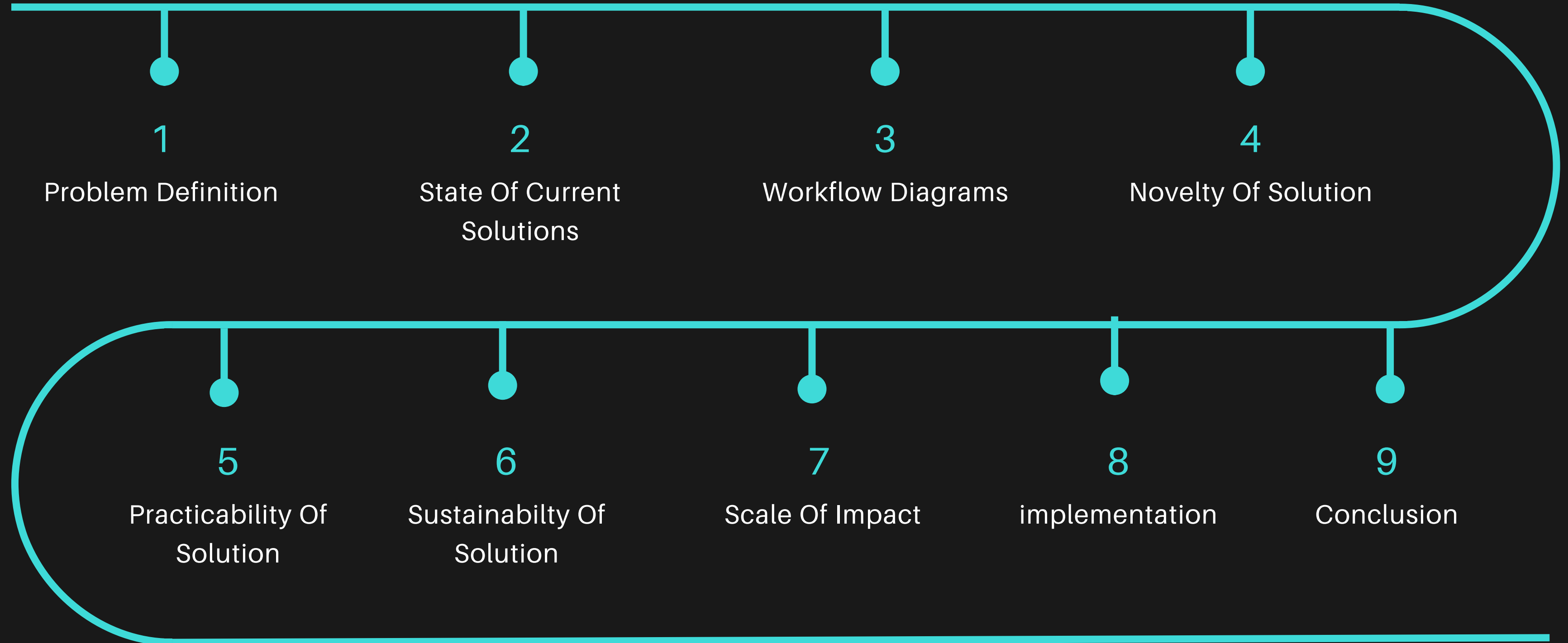
Problem Statement

Fake Social Media Profile Detection and Reporting

Problem Statement No. 1364

PPT Timeline

1

Problem Definition

2

State Of Current Solutions

3

Workflow Diagrams

4

Novelty Of Solution

5

Practicability Of Solution

6

Sustainabilty Of Solution

7

Scale Of Impact

8

implementation

9

Conclusion

Problem Definition

In the age of pervasive online social networking, the prevalence of fake social media profiles has grown exponentially, posing significant risks to individuals, businesses, and society as a whole. These fraudulent profiles engage in activities such as online impersonation, spamming, and dissemination of inappropriate or illegal content. Detecting and mitigating the presence of fake social media profiles is imperative for ensuring a safe and trustworthy digital environment.

Rapid Growth

Sophistication

Legal and Ethical Considerations

State Of Current Solutions

The real worlds' problems depends on the solution's sophistication and data quality. Depending on the type and nature of data, the solution may lack transparency in decision-making

Key Aspects

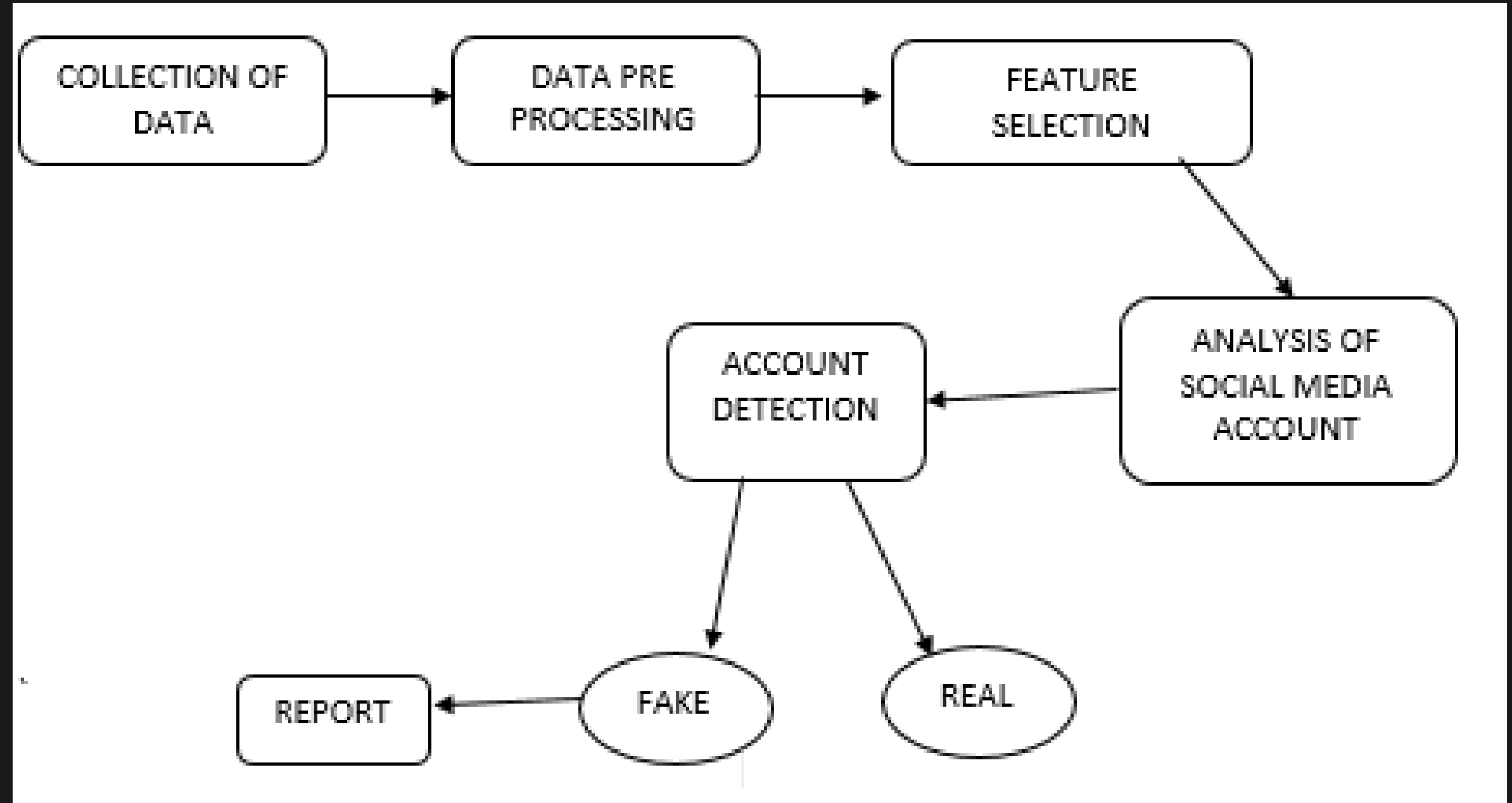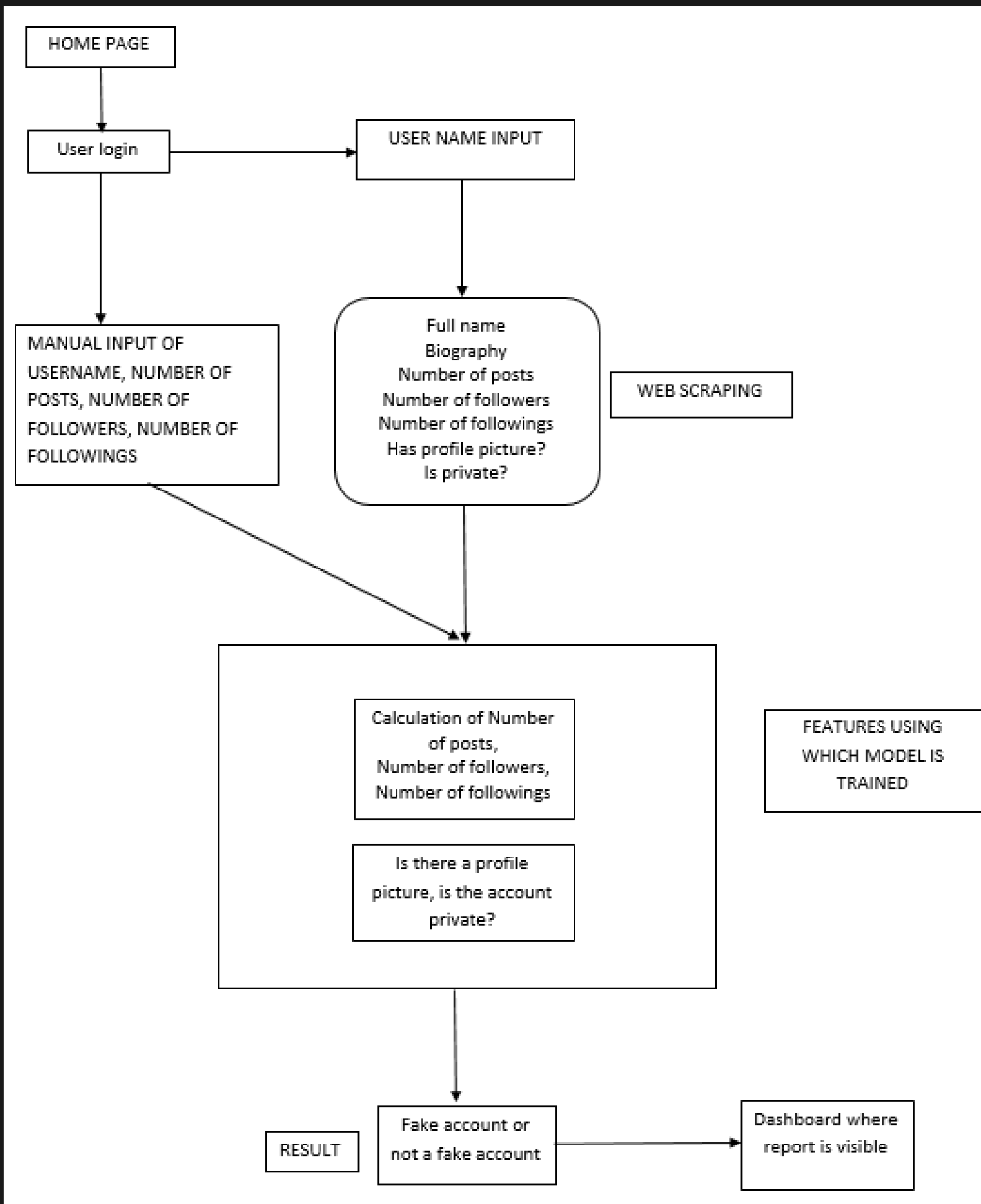Transparency

Scalability

Regularization

Model Accuracy

Continous Learning
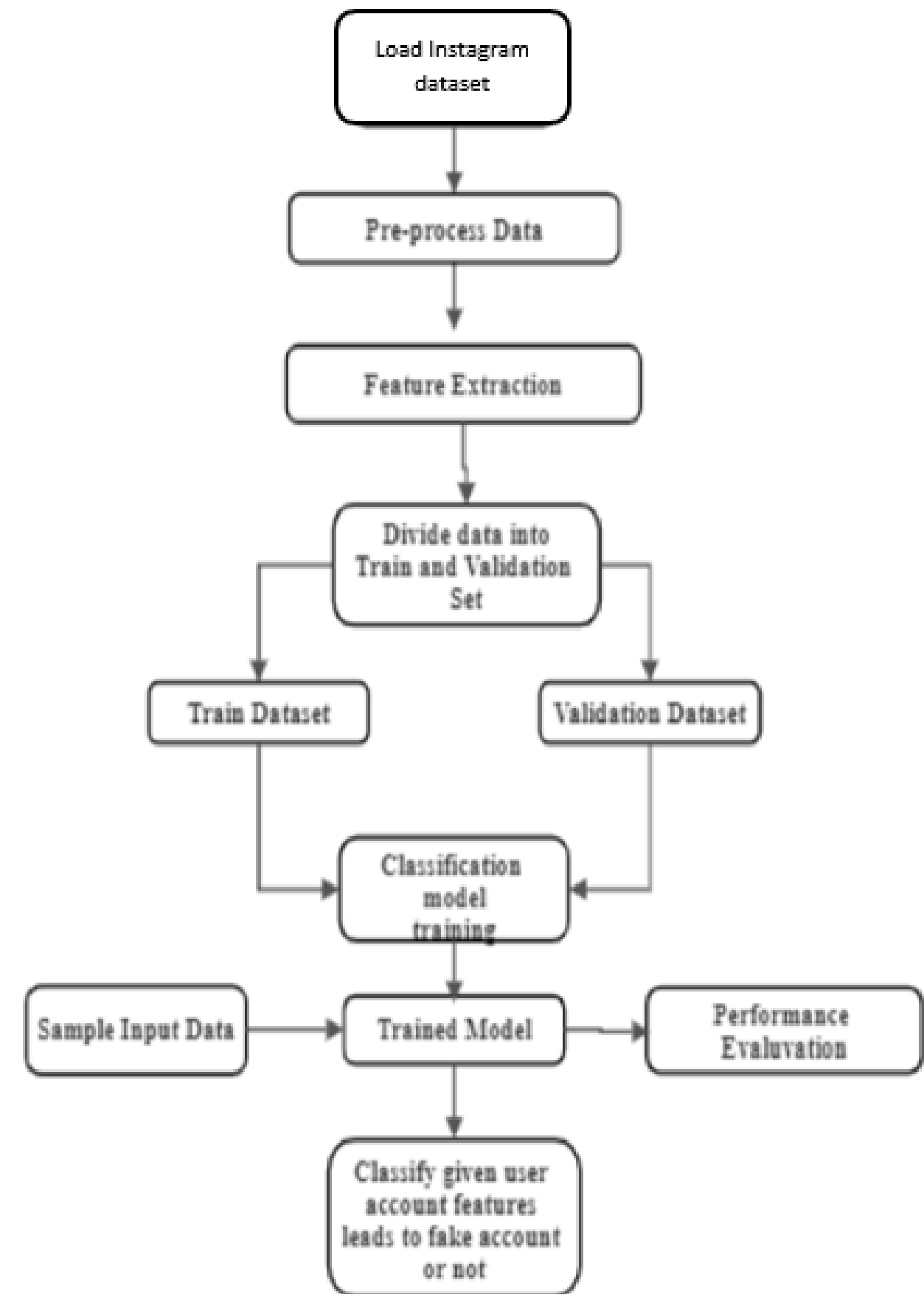
Cost Effectiveness

WorkFlow Diagrams

Flowchart

System Architecture

Overall Detection Process

Novelty Of Solution

We made use of the **XGboost Algorithm** as proved to give the most accuracy as compared to logistic regression and SVM (Support Vector Machine).

A reason being XGBoost is more efficient for working with smaller datasets.

The results showing XGBoost is more accurate are displayed on the next slide.

Future scope :

- Including **Face Detection** to verify the user's profile picture and the person using the app.
- For larger datasets, we will use **ANN** as it proves to be more efficient.

```
Classification Report:
              precision    recall  f1-score   support

           0       0.90      0.88      0.89        60
           1       0.89      0.90      0.89        60

    accuracy                           0.89       120
   macro avg       0.89      0.89      0.89       120
weighted avg       0.89      0.89      0.89       120

Accuracy: 89.17%
```

SVM

```
              precision    recall  f1-score   support

           0       0.93      0.90      0.92        60
           1       0.90      0.93      0.92        60

    accuracy                           0.92       120
   macro avg       0.92      0.92      0.92       120
weighted avg       0.92      0.92      0.92       120

Accuracy: 91.67%
```

XGboost

```
Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.87      0.90        60
           1       0.88      0.95      0.91        60

    accuracy                           0.91       120
   macro avg       0.91      0.91      0.91       120
weighted avg       0.91      0.91      0.91       120

Accuracy: 90.83%
```
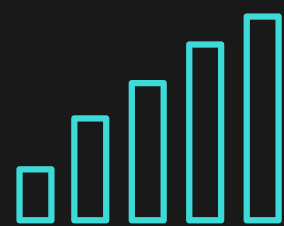
Logistic Regression

XGboost algorithm accuracy output compared to other training models

Practicability Of the Solution

MODEL PERFORMANCE
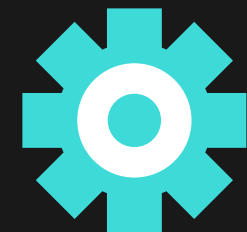XGBoost excels in binary classification tasks.

FEATURE ENGINEERING
Handles diverse feature types effectively
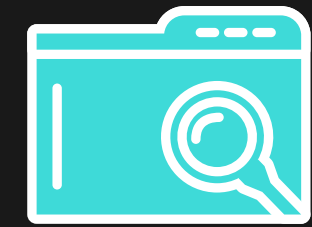
DATA AVAILABILITY
Abundant labeled data exists for model training.

HYPERPARAMTER TUNING
Allows fine-tuning for desired performance.

Sustainability Of the Solution

MONITORING AND OPTIMIZATION

Continuously monitor system performance and resource utilization while making adjustments to optimize efficiency.

EFFICIENT RESOURCE USAGE

The application minimizes waste and maximizes the efficient utilization of computational and storage resources, reducing environmental impact and operational costs

COMPLIANCE WITH LEGAL AND ETHICAL STANDARDS

The application adheres to data privacy, security, and other regulatory requirements, reducing legal risks and building user trust.

Scale Of Impact

- **User Safety**: Enhancing user safety by detecting and preventing interactions with fake profiles.

- **Platform Trustworthiness**: Increasing trust in online platforms by reducing the presence of fake accounts.

- **Reduced Fraud**: Minimizing fraudulent activities and scams that fake profiles may be associated with.

- **Enhanced User Experience**: Improving the overall user experience by reducing encounters with deceptive profiles.

- **Protecting Personal Information**: Safeguarding user data from misuse or exploitation by fake profiles.

Technology Stack

Frontend

- HTML
- CSS
- JavaScript

Backend

- Python
- Flask (Web Framework)

AI/ML

- XGBoost ML Model (Machine Learning)
- Pandas (Data Cleaning)
- Dataset for Training the Model
- Instaloader Library (Attribute Scraping)
- OpenCV (Face Detection)

Implementation

Result
Prediction: Fake

Instagram Profile Data

- **Username:** ranadesamihan
- **Full Name:** Samihan Ranade
- **Biography:**
- **Followers Count:** 127
- **Following Count:** 1863
- **Has Profile Picture:** 1
- **Username Length:** 13
- **Ratio Numerical Chars in Username:** 0.0
- **Full Name Tokens:** 2
- **Numerical Chars in Full Name:** 0.0
- **Full Name Length:** 14
- **Same Username and Full Name:** 0
- **Description Length:** 0
- **External URL:** 0
- **Private:** 1
- **Number of Posts:** 0

FLAG ACCOUNT

Profile Picture (0 or 1):

Numbers/Length of Username:

Fullname Words:

Numbers/Length of Fullname:
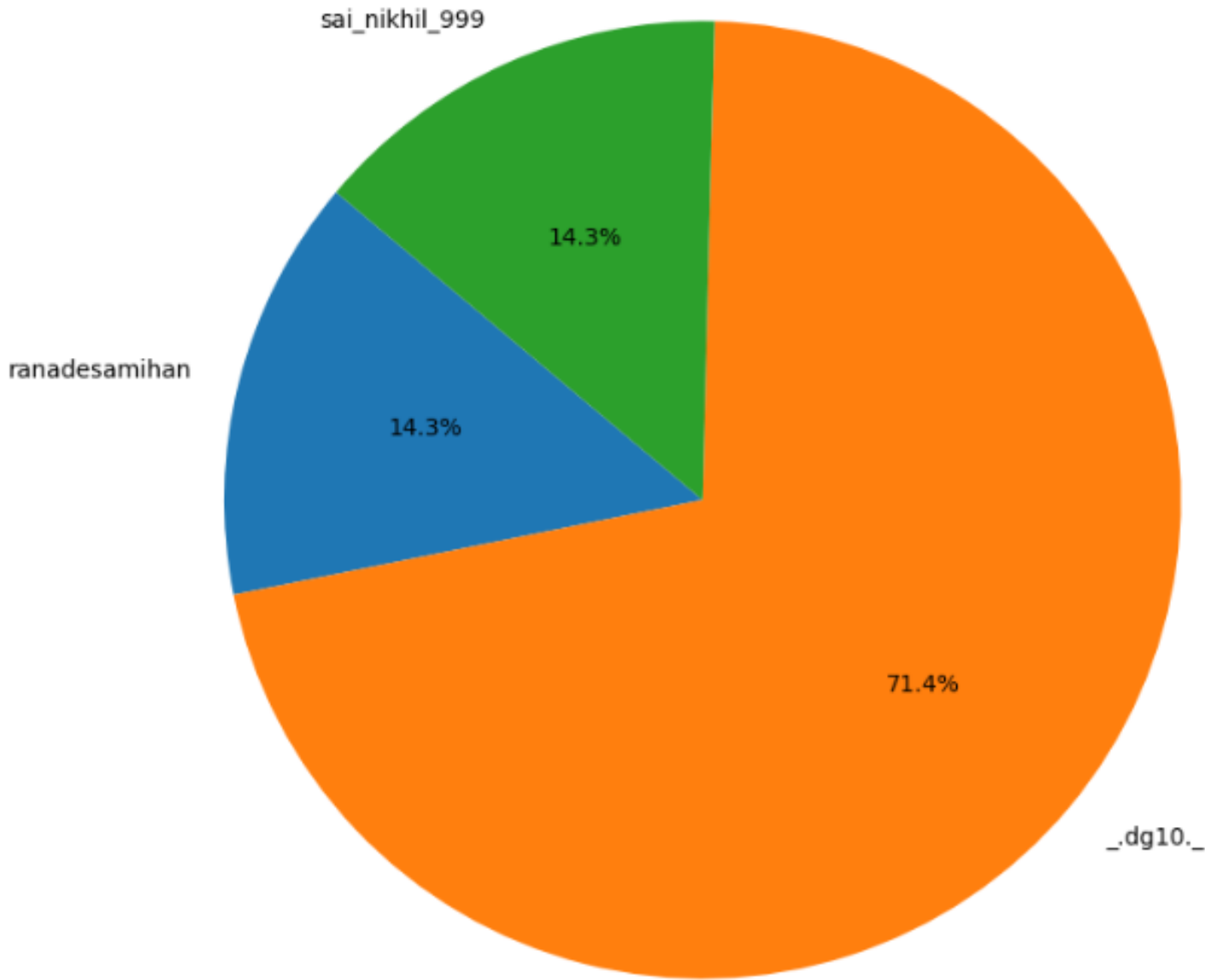
Name Equals Username (0 or 1):

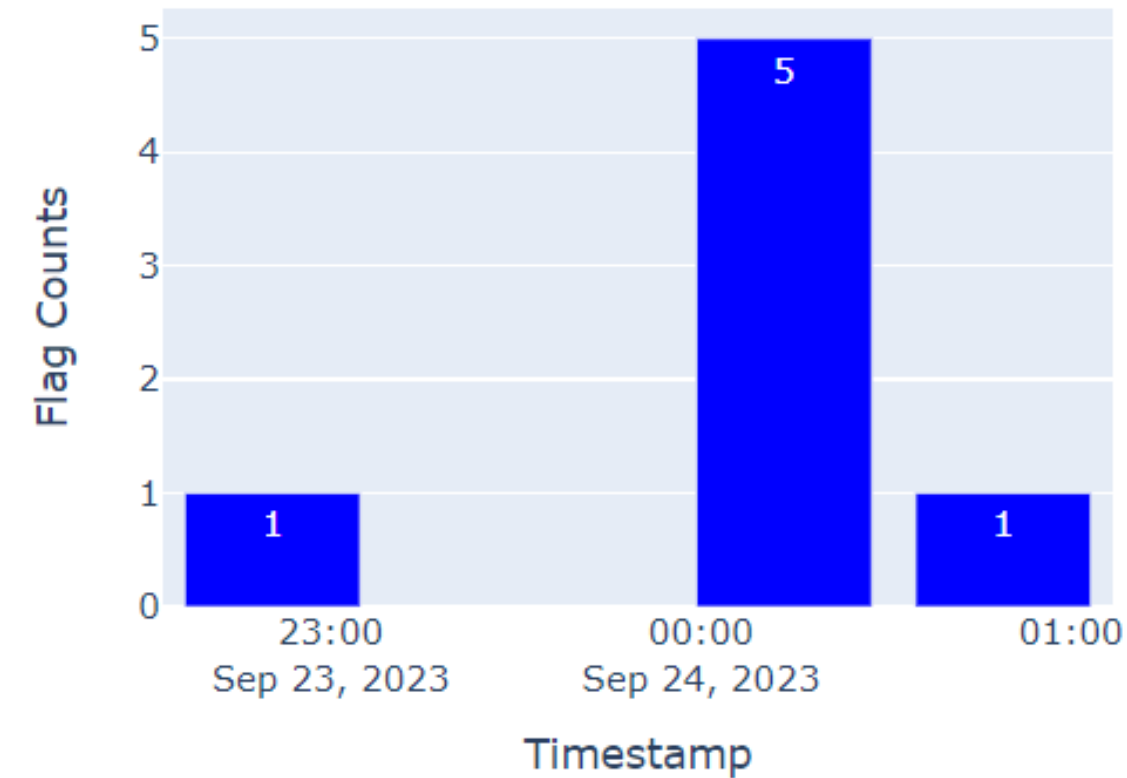Description Length:

External URL (0 or 1):

Private (0 or 1):

Number of Posts:

Number of Followers:

Flagged Usernames

sai_nikhil_999

ranadesamihan

14.3%

14.3%

71.4%

_.dg10._

Flagged Users Over Time

Username	Flag Count	Timestamp
ranadesamihan	1	2023-09-23 22:50:34.514168
_.dg10._	5	2023-09-24 00:13:30.755718
sai_nikhil_999	1	2023-09-24 00:49:01.584297

Conclusion

In conclusion, detection of the fake and automated accounts which leads to fake engagement in Instagram is studied as a binary classification problem . To our knowledge, this is the first time for such an analysis over Instagram accounts. Our contributions with this work are: collection of datasets for fake account detection, proposing derived features for fake classification and proposing a cost sensitive feature reduction technique based on XGboost algorithms.

Thank You!