FAKE NEWS DETECTION

- ARCITECTURE DOCUMENT

Contents

Document Version Control

Date	Version	Description	Author

1 INTRODUCTION

1.1 PROJECT DESCRIPTION

The goal of this project is to create a graphical user interface (GUI) web application that could predict whether a news article is fake or not using trained models that have been trained over open-source datasets with the utilisation of Natural Language Processing (NLP) concepts based on the article text content.

1.2 PURPOSE OF THE DOCUMENT

The aim of this document is to provide a comprehensive architectural framework for the fake news automatic detection using machine learning and deep learning algorithms.

1.3 SCOPE OF THE DOCUMENT

This document presents the technical architecture which helps the readers to get an overall understanding of the tools and technologies, system level and web application frameworks, model workflow etc. for the development of web application.

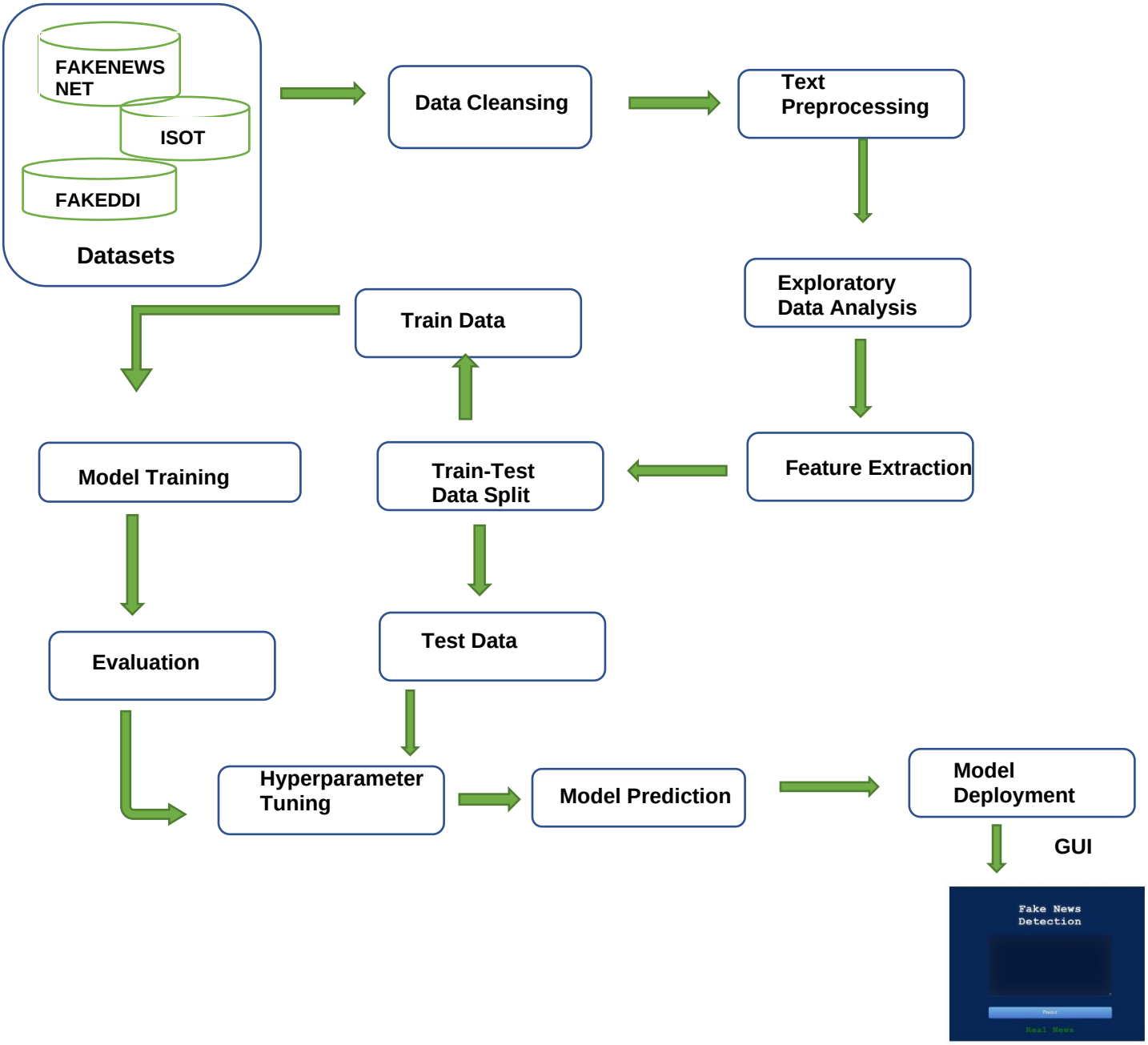
1.4 INTENDED AUDIENCE

The present document is intended to be read by the following people:

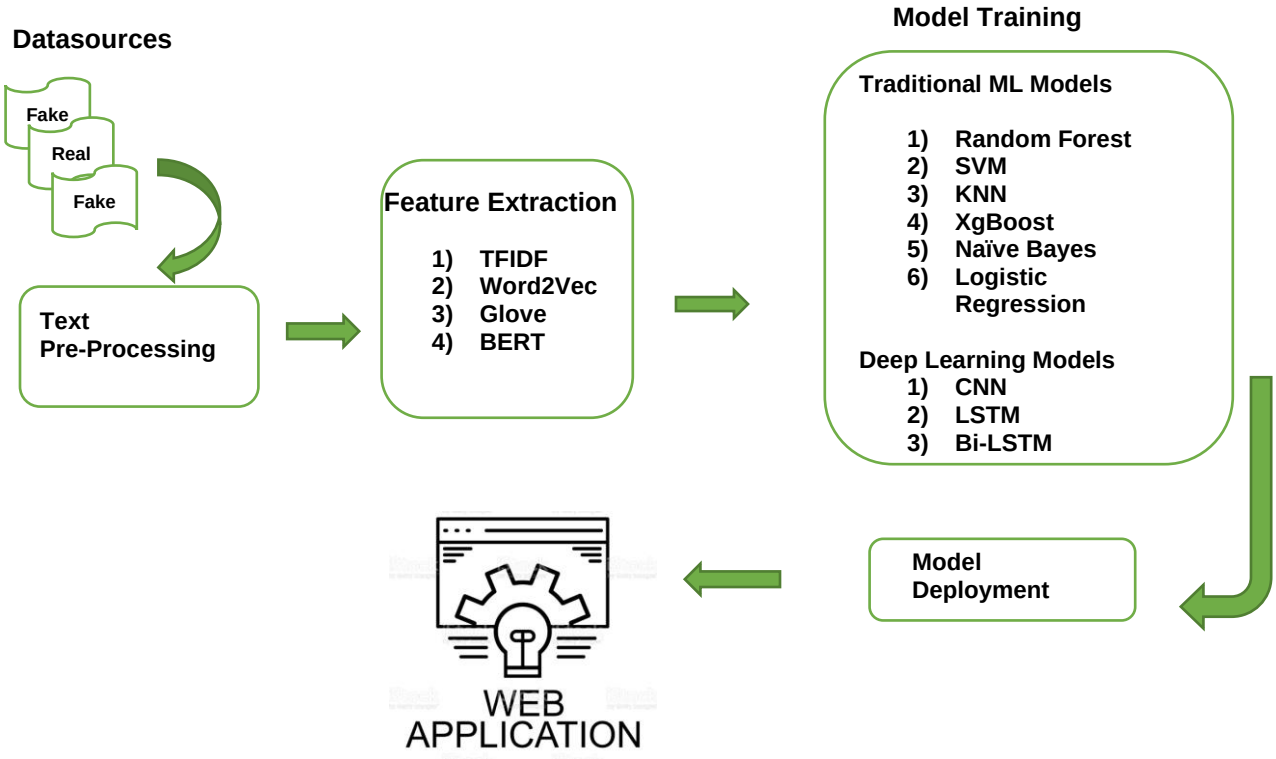
- Researchers.
- Students.
- Software Development Team.

2. ARCHITECTURE - FAKE NEWS DETECTION FRAMEWORK

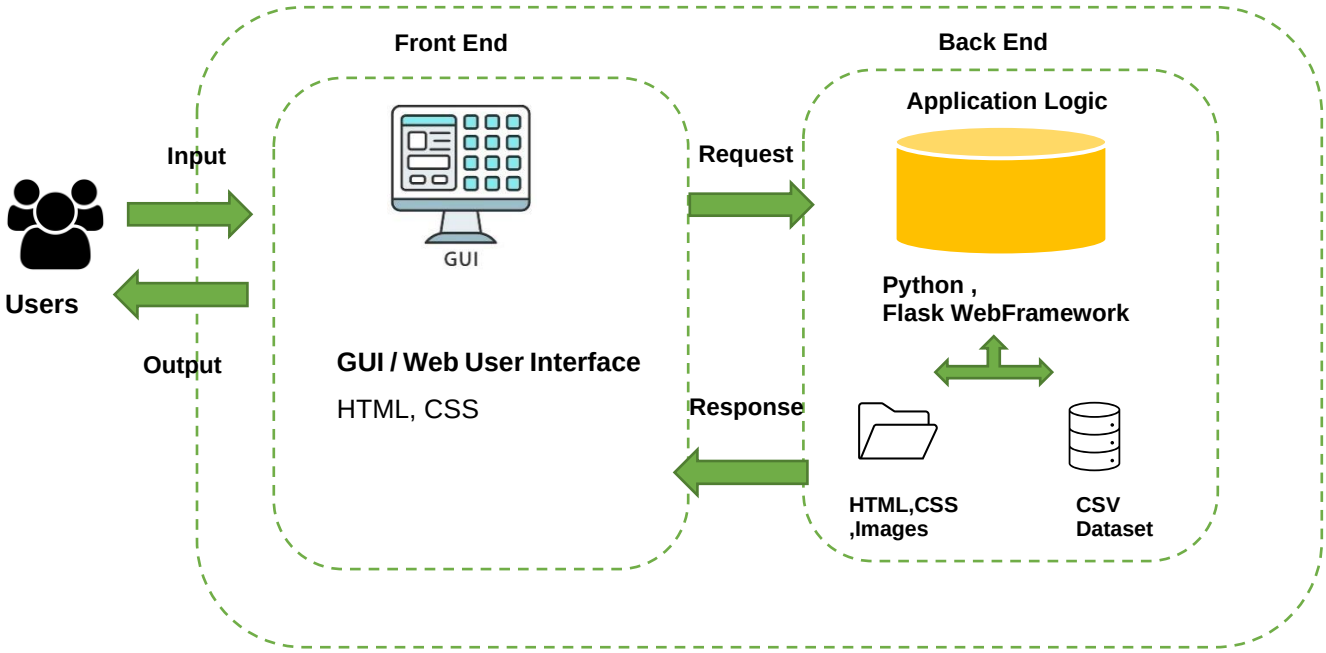
2.1 MODEL PIPELINE WORKFLOW



2.2 SYSTEM ARCHITECTURE OF PROPOSED FAKE NEWS DETECTION



2.3 WEB APPLICATION FRAMEWORK - GUI



3. SERVER CONFIGURATION DETAILS

3.1 Model Training Server

Type	Value
RAM	50 GB
GPU	RTX 6000
Hard Disk Space Capacity	50 GB

3.2 Deployment Server

Type	Value
RAM	16 GB
GPU	Not Needed
Hard Disk Space Capacity	10 GB

4. DATASET DETAILS

4.1 FAKENEWSNET

Type	Value
Data Size	267 MB
Row Records	21k
Labels	Fake and Real

4.2 ISOT

Type	Value
Data Size	110 MB
Row Records	32k
Labels	Fake and Real

4.3 FAKEDDIT

Type	Value
Data Size	2 GB

Row Records	800k
Labels	Fake and Real

5. TECHNOLOGY STACK DETAILS

5.1 Technology Versions

Technology	Versions
Python	3.7
Flask	2.1.2
TensorFlow Keras	2.9.0
HTML	HTML5
CSS	CSS3

5.2 List of Python Libraries

Libraries	Versions
Numpy	1.21.5
Pandas	1.3.4
Matplotlib	3.5.1
Seaborn	0.11.2
sklearn	1.0.2
Xgboost	1.6.1
NLTK	3.7
Wordcloud	1.8.1
Gensim	4.1.2
bert	0.14.9
tldextract	3.2.0
BeautifulSoup	4.11.1

5.3 Proposed Models

Machine Learning Models	Deep Learning Models
Random Forest	1 D Convolutional Neural Network
SVM	Long Short-Term Memory
XgBoost	Bidirectional Long Short Term Memory
KNN	
Naïve Bayes	
Logistic Regression	

5.4 Pretrained Models

Pretrained Model Name	Download URL
bert_en_uncased_L-12_H-768_A-12/1	https://tfhub.dev/tensorflow/bert_en_uncased_L-12_H-768_A-12/1
Word2Vec - **GoogleNews-vectors-negative300.bin**	https://s3.amazonaws.com/dl4j-distribution/GoogleNews-vectors-negative300.bin.gz
Glove - **glove.6B.300d.txt**	http://nlp.stanford.edu/data/glove.6B.zip