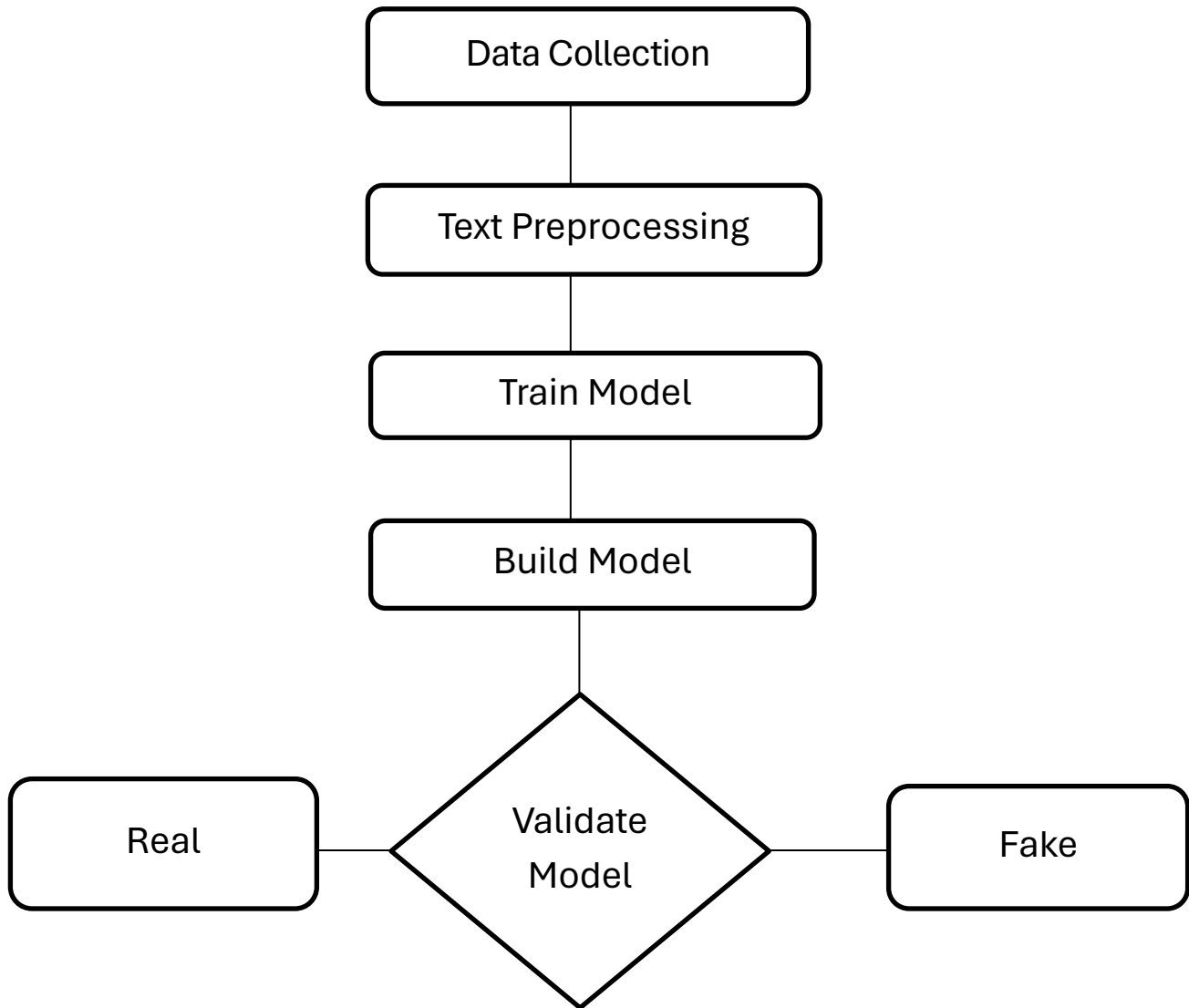


Fake News Detection Using NLP

Phase 2: INNOVATION

To implement this, collecting the fake news data set from the kaggle .



1. **Data Collection:** In the first step, you gather a diverse and representative dataset of news articles. This dataset should include both reliable, authentic news articles and fake, deceptive ones. The accuracy and performance of your fake news detection model heavily depend on the quality and quantity of the data collected. A larger and more balanced dataset helps the model learn the distinguishing features between real and fake news.
2. **Text Preprocessing:** Text preprocessing is essential to clean and prepare the data for analysis. This step involves tasks like removing HTML tags, special characters, and punctuation, tokenization (splitting the text into words or tokens), converting text to lowercase, and stemming or lemmatization to standardize words. Additionally, stop words (common words like "the" or "and") may be removed to reduce noise in the dataset. Text preprocessing ensures that the text data is in a suitable format for NLP analysis.
3. **Train Model:** In this step, you split your preprocessed dataset into training and testing sets. The training data is used to teach the machine learning model the patterns and characteristics of both real and fake news articles. Various NLP techniques are applied here, such as feature extraction, vectorization, and the creation of word embeddings using methods like Word2Vec or TF-IDF. The model, which can be a deep neural network, a support vector machine, or other algorithms, learns to distinguish between real and fake news based on these features.
4. **Building Model:** The model-building phase involves selecting and implementing a machine learning algorithm. This algorithm is trained on the preprocessed data and learns to classify news articles as real or fake based on the features extracted during the training phase. You can fine-tune the model's parameters and architecture to improve its performance, and you may also consider techniques like ensembling or deep learning architectures for enhanced accuracy.
5. **Validate Model:** Once the model is built, it needs to be evaluated for its performance. This is done by testing the model on a separate validation dataset (different from the training data). You can use various evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to assess how well the model is at detecting fake news. If the model performs well on the validation dataset, it can then be deployed for real-time fake news detection.

Real News Sample Head

	title	text	subject	date
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017

Fake News Sample Head

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017

Team:

MOHAMED ABDUL RAHMAN S (au951221104023)

YABESH JESILEN V (au951221104059)

SAM SELIN D (au951221104039)

RAHUL B (au951221104034)

NAGA MARI MUTHU A (au951221104028)