

Fake News Detection using Deep Learning

Final Report

Srihitha Pulapa, CS22B2009

Rikitha Ravi, CS22B2045

1 Introduction

This project presents a deep learning model using LSTM to classify news articles as real or fake based on their textual content. The pipeline includes preprocessing, exploratory analysis, model building, training, and evaluation.

2 Dataset and Analysis

The dataset consists of 44,898 articles (21,417 real and 23,481 fake), evenly distributed across both classes. We examined topic and year-wise distribution.

Preprocessing

- No null or duplicate entries found
- Merged `title` and `text` fields
- Lowercased, removed punctuation, digits, and stopwords
- Tokenized and padded to 100 tokens

Exploratory Analysis

- Word clouds showed fake news used emotional language; real news was more factual
- Most articles had fewer than 100 tokens

Key Observations

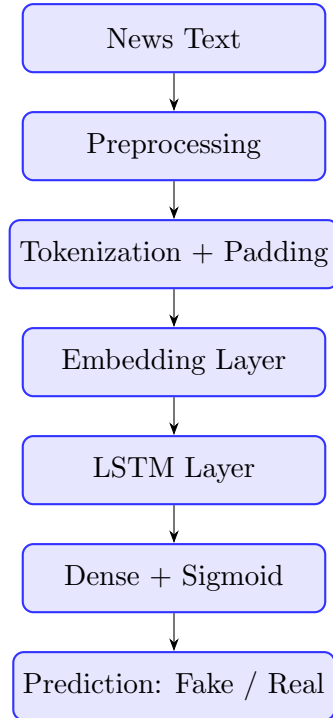
- Real news had structured, topic-specific terms
- Fake news was more sensational
- Balanced classes reduced bias

3 Model Architecture

- **Embedding Layer:** Vocabulary size = 20000, Embedding dimension = 50
- **LSTM Layer:** 15 units
- **Dense Output Layer:** 1 neuron with sigmoid activation
- **Loss Function:** Binary Crossentropy

- **Optimizer:** Adam
- **Epochs:** 10, **Batch Size:** 32

4 Block Diagram



5 Results and Evaluation

The model was tested on 4,696 fake and 4,284 real news samples, showing balanced performance across both classes. Key metrics:

- **Accuracy:** 0.9823
- **AUC:** 0.9984
- **Precision:** 0.9764
- **EER:** 0.0162

These results confirm strong and balanced classification with minimal errors.