Fake News Classifier

Project Proposal

WOO-CHAN KIM

I. Project title

Fake News Classifier

II. Project introduction

1) Objective

: The project objective is to detect fake news by analyzing news content truthfulness. It is a subtask of text classification and is defined as the task of classifying news as real or fake.

2) Background

: Recently, the advancement of artificial intelligence has led to an increase in the generation of false information, causing various social issues. In particular, fake news can be maliciously exploited in society, prompting active research using natural language processing (NLP) models to detect such misinformation.
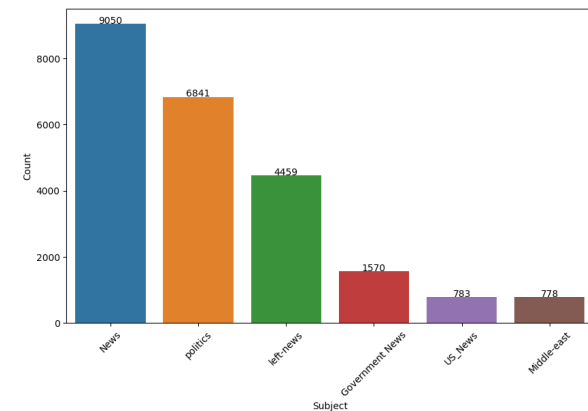
III. Dataset

Dataset separated in two files:
- Fake.csv (23,502 fake news article)
- True.csv (21,417 true news article)

Dataset columns:
- Title: title of news article
- Text: body text of news article
- Subject: subject of news article
- Date: publish date of news article

Word Cloud for True News Titles

Fake News subject Distribution

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

5 heads of True News Dataset

IV. Baseline

LSTM based Model

LSTM which is a natural language processing model designed is used to classify whether a news article text is real or fake. The input is an integer sequence consisting of up to 150 tokens, and the final output is the probability of being real or fake, computed using a softmax function.
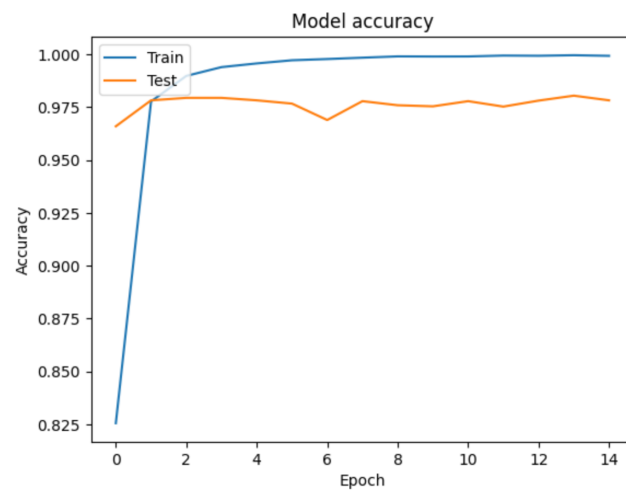
Model Overview

- **Task:** Fake News Classification
- **Architecture:** LSTM-based NLP model
- **Input:** Tokenized news text (maxlen = 150)
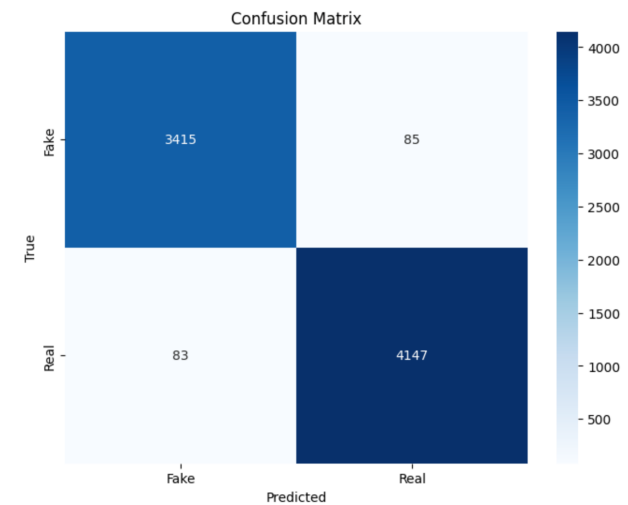- **Output:** Softmax – [Real, Fake] probabilities

LSTM Architecture
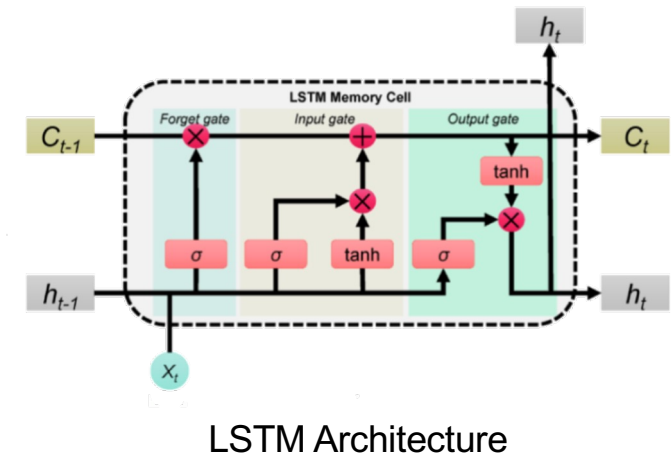
V. Results

- Test Loss: 0.059
- Test Accuracy: 0.9782

Train and Test set accuracy

Confusion matrix

Q & A

E-mail : kimwc620@korea.ac.kr

GitHub address :

https://github.com/SkyDreamer14/FakeNews