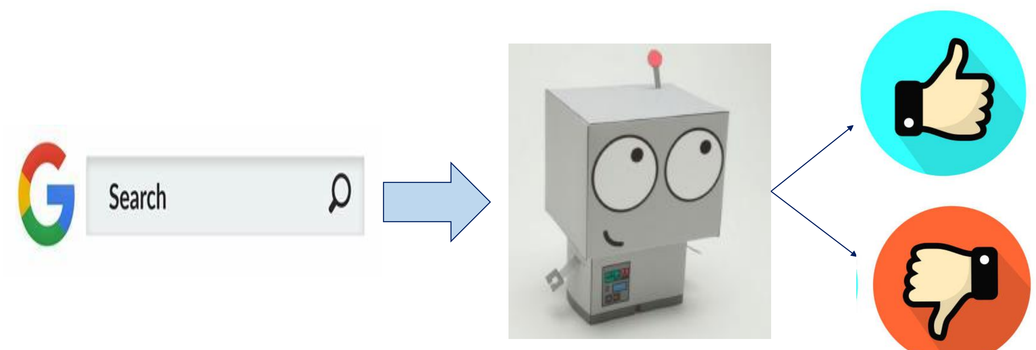
A Node2vec-based Bangla Fake News Detection Model Using Web Search Results

Mosharat Jahan
ID: 18701040
Session: 2017-2018
Supervisor: Associate Professor Mr. Rezaul Karim
Department of Computer Science & Engineering, University of Chittagong

Problem Statement & Motivation

Articles that can potentially mislead or deceive readers by providing fabricated information are known as fake news.

- Even-increasing amount of information available for consumption
- Creation of fake news is very easy with modern technology
- Can we determine whether or not an article is unreliable using only the URL link the news was obtained from?



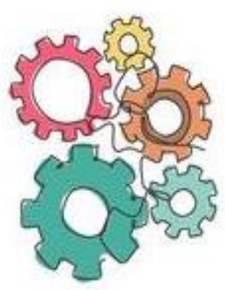
Fake News Detection Model

Related Works

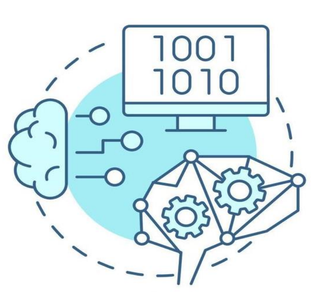
Authors	Used Dataset	Used Models	Best Results	
			Accuracy	F1-Score
Jae-Seung Shim Yunju Lee Hyunchul Ahn	English Dataset & Korean Dataset	Link2vec, SVM,Logistic Regresssion, ANN	93.1% (SVM)	--
Mahmuda Rahman Sakib Al Hasan Md Gulzar Hussain	Bangla News Datset	SVM,MNB	96.64% (SVM)	98%
Md Zobaer Hossain Md Ashraful Rahman Md Saiful Islam	BanFakeNews Dataset	CNN,LSTM, BERT	98% (CNN)	91%

Methodology

Data Collection
- BanFakeNews Dataset



Preprocessing
- Collect Domain from news source
- Feature Extraction using Node2vec
- Text processing for comparison model

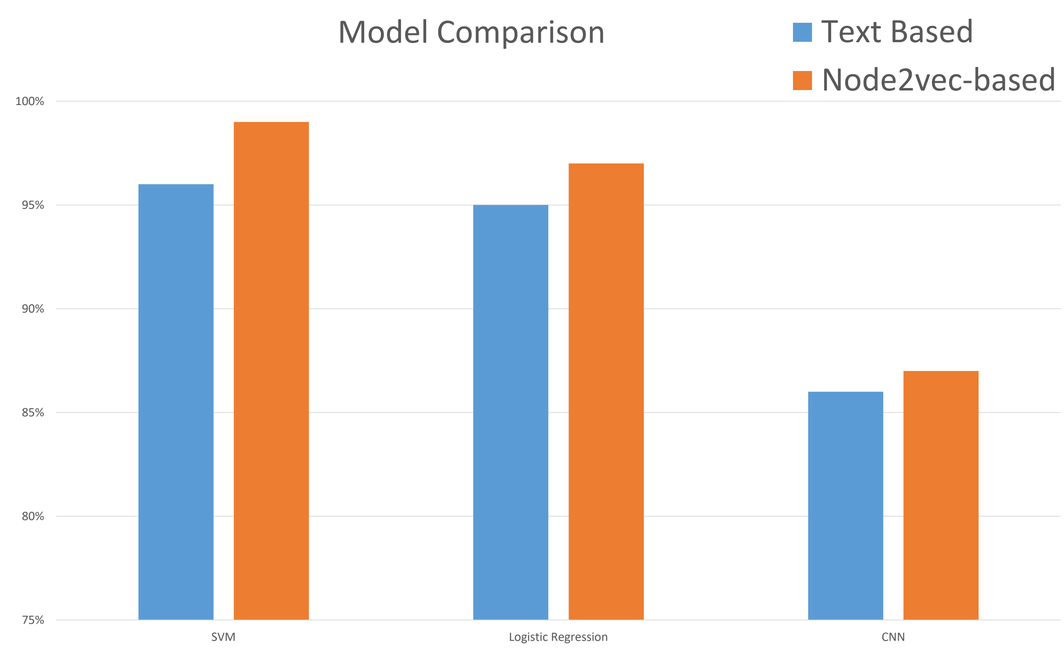


Model Training
- SVM
- Logistic Regression
- CNN

Evaluation
- Accuracy
- Precision
- Recall
- F1-Score

Results

- The Node2vec model has shown better accuracy compared to the text based model



Conclusion & Future Work

In this research we have proposed a model using Node2vec which has:
- Low and better execution time than text based model
- More than 90% accuracy for both ML model
- Ability to work with data even if its unbalanced
- Ability to work with dataset of different language

Our Future research could continue to:
- Work with datasets of different language
- Collect more than one URL and use them together for classification
- Apply different model to compare accuracy with Node2vec