**Detection of Bengali Online Fake News and Clickbait Headline on Social Media**

Proposal Prepared by Rakib & Shakib in part fulfillment of the degree requirements for the BSc in Computer Science and Telecommunication Engineering

Date:

**Abstract:**

Fake news is a phenomenon which is having a significant impact on our social life, in particular in the political and religious world. It can be responsible for hampering social peace and can create different types of anarchies. At 29 September 2012, one of the worst religious attacks in Bangladesh in Cox's Bazaar's Ramu Upazila was apparently triggered by a fake Facebook posting allegedly defaming the Holly Quran. Later on, many more political and religious anarchies had taken place due to spreading out of many fake news over time. Besides different political groups use them for their self-interest. On the other hand, clickbait headline, which is mainly used in SEO, befooled peoples and waste their valuable time and energy by tempting them to view their articles which isn’t literally the news expressed through the headline. Only by building a model based on a count vectorizer won’t work because these models do not consider the important qualities like word ordering and context. We propose in this paper, a ‘fake news’ and ‘clickbait headline’ detection model that use NLP and machine learning techniques. This experimental evaluation yields the best performance using Term Frequency-Inverted Document Frequency (TF-IDF) as feature extraction technique, and Bayesian Network/Neural Network as a classifier.

**Objective:**

The main objective is to detect the fake news and clickbait headline, which is a classic text classification problem with a straight forward proposition. It is needed to build a model that can differentiate between “Real” news and “Fake” news and can check if a headline is an overstatement or not.

**Introduction and Background:**

These days, fake news is creating different issues from sarcastic articles to a fabricated news and spread out different propaganda in some outlets.On the other hand, Clickbait is a form of false advertisement, designed to make readers want to click on a hyperlink, especially when the link leads to content of dubious value or interest. Online content has been playing a significant role in swaying user’s decisions and opinions. Opinion such as online reviews are the main source of information for the e-commerce customers to help with gaining insight into the product they are planning to buy. Fake news and clickbait headline are working as a form of opinion spam here. Facebook has been at the epicenter of much critique following media attention. They have already implemented a feature to flag fake news on the site when a user see’s it. They have also said publicly they are working on to distinguish these articles in an automated way. Certainly, it is not an easy task. A given algorithm must be politically unbiased – since fake news exists on both ends of the spectrum – and also give equal balance to legitimate news sources on either end of the spectrum. In addition, the question of legitimacy is a difficult one.

We present in this paper an n-gram features based approach to detect fake news,

which consists of using text analysis based on n-gram features and machine learning

classiﬁcation techniques. We study and compare six different supervised classiﬁcation

techniques, namely, K-Nearest Neighbor (KNN), Support Vector Machine (SVM),

Logistic Regression (LR), Linear Support Vector Machine (LSVM), Decision tree

(DT) and Stochastic Gradient Descent (SGD). Experimental evaluation is conducted

using a dataset compiled from real and fake news websites, yielding very encouraging

results

We present in this paper an n-gram features-based approach to detect fake news, which consists of using text analysis based on n-gram features and machine learning classiﬁcation techniques. We study and compare six/seven different supervised classiﬁcation techniques, namely, Naive-Bayes classification method, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Linear Support Vector Machine (LSVM), Decision tree (DT) and Stochastic Gradient Descent (SGD). Experimental evaluation is conducted using a dataset collected from real and fake news websites, yielding very encouraging results.