SCHOOL OF COMPUTING
Faculty of Engineering

Project Proposal Form MCSD 6215
Sem:…1……. Session:……

SECTION A: Project Information.

Program Name: **Masters of Science (Data Science)**

Subject Name: **Project 1 (MCSD 6215)**

Student Name: Alexander Tan Ka Jin

Metric Number: MSCS241002

Student Email & Phone: tanka@graduate.utm.my 0167778810

Project Title: Sentiment Analysis of news articles using Bidirectional Recurrent Neural Networks

Supervisor 1:

Supervisor 2 / Industry Advisor(if any):

SECTION B: Project Proposal

Introduction:

Natural Language Processing (NLP) is the subject of interpreting language via machines. Various aspects of language are considered for NLP such as semantics and syntax. This proposal aims to utilize sentiment analysis, a part of NLP analysis of the semantics of language onto journalistic news articles. The domain of sentiment analysis usually looks at data with shorter subjective text like social media opinions. This creates a research gap for longer more objective sounding text such as journalism. With the development of deep learning models like BiLSTMs, allow for language to be analyzed within context of surrounding words and sentences which can uncover more complex patterns of sentiment.

Problem Background:

Recent studies of Sentiment Analysis have increased in their application of Deep Learning models such as BiLSTM. The use of BiLSTM models and BiRNN models in general have increased the performances of sentiment analysis within social media text. One of the main challenges to sentiment analysis is the application towards ambiguous language such as more subtle forms of propaganda. Much has been documented on how otherwise neutral journalism may be push certain viewpoints and opinions and much has been documented about media bias, the question that this study aims to resolve is the relationship between media bias and sentiment, if sentiment analysis can properly detect narrativization in journalism and its relationship between media outlet bias and popularity. If it is possible, this research aims to also visualize and provide knowledge on how frequently narrativization occurs in journalism and what correlations can be

made to it.

Problem Statement:

This thesis aims to perform an analysis of the relationship between sentiment within text and use of language within news using deep learning models. The research aims to address the lack of sentiment analysis on long-form text data that is published within various media outlets rather than focusing on short-form text data within social media. In particular, it aims to uncover and analyze in detail how subjective opinions are embedded within objective text. It analyses the efficacy of sentiment analysis with regards to identifying the degree of opinionated news. The analysis will also determine the relationship between media outlets and narrativization in journalism of certain topics. With the use of Deep Learning technology, this research aims to use models such as LSTMs to better analyze textual data. This research aims to help citizens understand the use of language better within media and helps understanding how journalism can manipulate language to advocate or disagree with certain figures and policies.

Aim of the Project:

The aim of the project is to investigate challenges to perform sentiment analysis on long-form text, what features of data and data preprocessing methods can benefit the analysis, the construction of BiLSTM and BIGRU models for analyzing the data, and investigating the sentiment behind certain topics and opinions within journalism and news.

Objectives of the Project:

1. Perform an exploratory sentiment analysis on large news datasets to identify patterns and trends in the language use of journalism.
2. Construct BiLSTM and BiGRU model to uncover complex language patterns and potentially reveal sentiment hidden under objective language
3. Conduct comprehensive evaluation on the model to ensure it's performance.
4. Deploy the model onto a large dataset of articles and visualize article sentiment by each media outlet and topic.

Scopes of the Project:

The study uses large sets of data from recent time and only considers the use of models, BiLSTM and BiGRU The research will conduct preliminary analysis of the data, extensive data cleaning and preprocessing before it is used to train our choice of model. The model will then be meticulously evaluated for it's robustness and performance before being used to predict emotion sentiment of articles. The scope of the analysis of articles will cover the purported media bias of it's original institution and it's popularity as a source of news along with the sentimental score of the article. All of it will be presented in the study as a visualization using graphs. Finally, the project will examine and address it's limitations for future work and enhancements in both the use of data and the construction of the model.

Expected Contribution of the Project:

The research can help people identify the biases of new sources and how sentiment and opinions are used as forms of propaganda. Which can help people critique media institutions and ensure that people can differentiate opinion and facts within a long article. Additionally, the research contributes to the use of deep learning models within sentiment analysis of long-form textual data and sentiment analysis of objective and neutral language.

Project Requirements:

Software: Python, Pytorch, MatPlotLib

Hardware: Computer

Technology/Technique/ Methodology/Algorithm: Deep Learning, BiLSTM, BiGRU

Type of Project (Focusing on Data Science):

[ ] Data Preparation and Modeling

[✓] Data Analysis and Visualization

[ ] Business Intelligence and Analytics

[✓] Machine Learning and Prediction

[ ] Data Science Application in Business Domain

Status of Project:

[✓] New

[ ] Continued

If continued, what is the previous title? .....

SECTION C: Declaration

I declare that this project is proposed by:

[✓] Myself

[ ] Supervisor/Industry Advisor ( )

Student Name: Alexander Tan Ka Jin

22/12/2024

Signature **Date**

SECTION D: Supervisor Acknowledgement

The Supervisor(s) shall complete this section.

I/We agree to become the supervisor(s) for this student under aforesaid proposed title.

Name of Supervisor 1: .....

Signature **Date**

Name of Supervisor 2 (if any): .....

Signature **Date**

SECTION E: Evaluation Panel Approval

The Evaluator(s) shall complete this section.

Result:

[ ] FULL APPROVAL [ ] CONDITIONAL APPROVAL (Major)*

[ ] CONDITIONAL APPROVAL (Minor) [ ] FAIL*

***** Student has to submit new proposal form considering the evaluators' comments.

Comments:

Name of Evaluator 1:

Signature **Date**

Name of Evaluator 2:

Signature **Date**