**RESEARCH PLAN**

**FAKE AND AUTHENTIC NEWS DETECTION USING LONG SHORT TERM MEMORY AND BI-DIRECTIONAL LONG SHORT TERM MEMORY**

**1.0 INTRODUCTION AND JUSTIFICATION**

Fake News is an intentional and verifiably false information published by a news outlet. Nowadays, fake news has become widely spread and has bad effects on many aspects of life such as political, economic and education. It is typically generated for commercial interests to attract viewers and collect advertising revenue. According to statistics, two million social media accounts are closed ever year to limit the spread of fake news or misinformation. During the COVID-19 pandemic, there were many instances of fake news on COVID-19. A Malaysian website spotted around more than 150 fake news in Malaysia only in March 2020. Producing fake news about the COVID-19 is still continuous. So, identifying fake news in social networks is very important because they have an impact that have been tremendous due to the massive user numbers globally, which is further boosted by the extensive information sharing and propagation among these users. Bangladesh also faces many measurable occurrences of fake news for the last few years. It can also hamper one’s life. Social platforms mislead different writers by spreading fake news. Some of the time, we can see that the rapidity for getting out fake report is quicker than the authentic report.

Majority of the research that have been done in fake news detection have been with a variety of machine learning algorithms. This project is going to include deep learning algorithms such as long short term memory algorithms (LSTMs)

**2.0 RESEARCH QUESTIONS, AIMS AND OBJECTIVES**

The aim of this research is to accurately predict fake and authentic news using Long Short Term Memory and Bi-Directional Long Short Term Memory algorithms.

2.1 RESEARCH QUESTION

The research question for this thesis is how will the state of the art deep learning algorithms compare to the traditional machine learning algorithms in terms of the performance metrics such as accuracy, precision, recall and the f1 score.

2.2 OBJECTIVES

* Conduct research on academic work that has been done on Long Short Term Memory and Bi-Directional Long Short Term Memory
* Conduct Research on optimization techniques used to improve the performance of Long Short Term Memory algorithms
* Identify appropriate dataset of text data from various data science platforms such as Kaggle and social media sites such as Twitter and Facebook
* Build machine learning algorithms using the two most conventional machine learning methods used in Natural Language Processing (Naïve Bayes and the Linear Support Vector Classifier)
* Build a Normal Long Short Term Memory and Bi-Directional Long Short Term Memory algorithms
* Compare the machine learning methods and the deep learning methods in terms of their performance metrics

2.3 DELIVERABLES

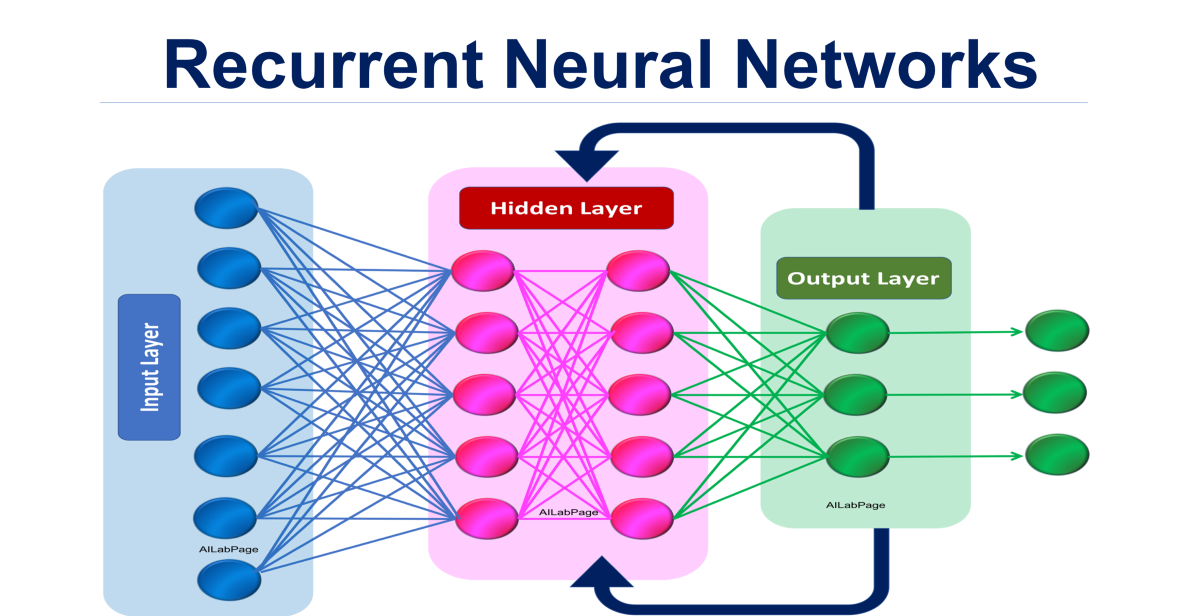
Two different machine algorithms (Naïve Bayes and Linear Support Vector Classifier) and deep learning algorithms (Long Short Term Memory and Bi-Directional Long Short Term Memory) would be built and then optimized.

**3.0 LITERATURE REVIEW**

This literature review will look at academic research that has been done on neural networks and those carried out on the use of Long Short Term Memory and Bi-Directional Long Short Term Memory for the purpose of fake and authentic news detection.

3.1 RECURRENT NEURAL NETWORKS

A recurrent neural network is a class of artificial neural networks where connections between nodes form a directed or undirected graph along a temporal sequence. This allows it to exhibit temporal dynamic behavior. Recurrent neural network (RNN) is a class of [artificial neural networks](https://en.wikipedia.org/wiki/Artificial_neural_network) where connections between nodes form a [directed](https://en.wikipedia.org/wiki/Directed_graph) or [undirected graph](https://en.wikipedia.org/wiki/Graph_(discrete_mathematics)) along a temporal sequence. This allows it to exhibit temporal dynamic behavior. Derived from [feedforward neural networks](https://en.wikipedia.org/wiki/Feedforward_neural_networks), RNNs can use their internal state (memory) to process variable length sequences of inputs.



*Fig 3.1 Diagram of a recurrent neural network*

3.2 LONG SHORT TERM MEMORY

Long Short Term Memory networks are the type of recurrent neural networks capable of learning order dependence in sequence prediction problems. This is a behaviour required in complex problem domains like machine translation, speed recognition and more. LSTMs are a complex area of deep learning. It can be hard to get your hand around what LSTMs are and how terms like bidirectional and sequence to sequence relate to the field.

According to (Felix A. Gers, et al., 1994), standard RNNs fail to learn in the presence of time lags greater than 5 – 10 discrete time steps between relevant input events and target signals. The vanishing error problem casts doubt on whether standard RNNs can indeed exhibit significant practical advantages over time window-based feedforward networks. A recent model, long short term memory is not affected by this problem. LSTMs can learn to bridge minimal time lags in excess of 1000 discrete time steps by enforcing constant error flow through constant error carousels within special units called cells.

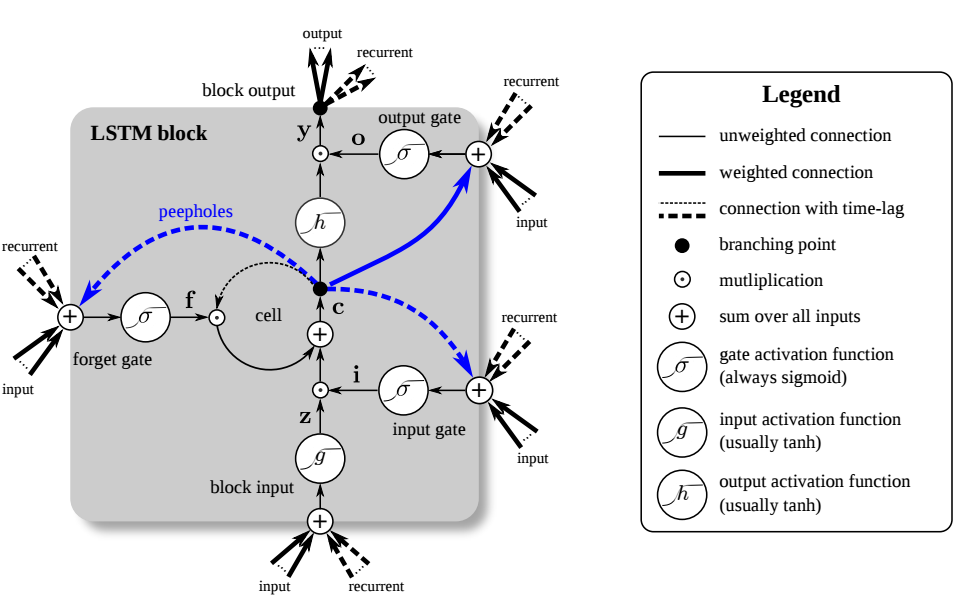
The key to the LSTM solution to the technical problems was the specific internal structure of the units used in the model. According to (Alex Graves, et al., 2009), the range of contextual information that standard RNNs can access is in practice quite limited. The problem is that the influence of a given input on the hidden layer, and therefore on the network output, either decays or blows up exponentially as it cycles around the network’s recurrent connections. This shortcoming is known as the gradient problem. Long Short Term memory is an RNN architecture specifically designed to address this problem. It is governed by its ability to deal with vanishing and exploding gradients, the most common challenge is designing and training RNNs. To address this challenge, the LSTM network was introduced and applied with great success to translation and sequence generation.

3.3 HOW LONG SHORT TERM MEMORY NETWORKS WORK

(Sepp Hochreiter and Jürgen Schmidhuber, 1997) LSTMs use networks with one input layer, one hidden layer, and one output layer. The (fully) self-connected hidden layer contains memory cells and corresponding gate units. Each memory cell’s internal architecture guarantees constant error flow within its constant error carrousel CEC. This represents the basis for bridging very long time lags. Two gate units learn to open and close access to error flow within each memory cell’s CEC. The multiplicative input gate affords protection of the CEC from perturbation by irrelevant inputs. Likewise, the multiplicative output gate protects other units from perturbation by currently irrelevant memory contents.

The Long Short Term Memory architecture was motivated by an analysis of error flow in existing RNNs, which found that long time lags were inaccessible to existing architectures, because back propagated error either blows up or decays exponentially. An LSTM layer consists of a set of recurrently connected blocks, known as memory blocks. These blocks can be thought of as a differentiable version of the memory chips in a digital computer. Each one contains one or more recurrently connected memory cells and three multiplicative units the input, output and forget gates that provide continuous analogues of write, read and reset operations for the cells. The net can only interact with the cells via the gates.

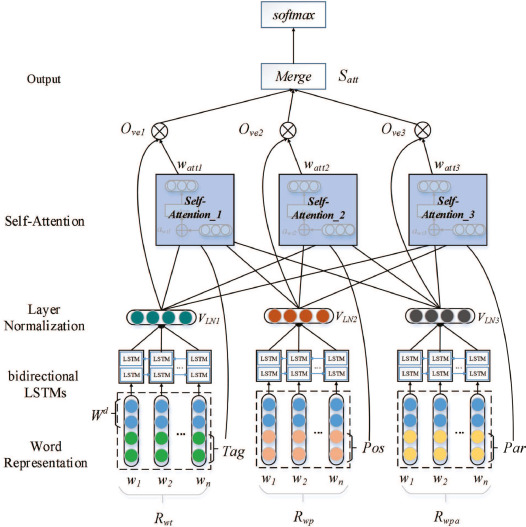
According to (Klaus Greff, et al., 2015), the most commonly used LSTM architecture (vanilla LSTM) performs reasonably well on various datasets. Learning rate and network size are the most crucial tunable LSTM hyper parameters. This implies that the hyper parameters can be tuned independently. In particular, the learning rate can be calibrated first using a small network, thus saving a lot of experimentation time.



*Fig 3.2 Diagram of a LSTM neural network*

3.4 BI DIRECTIONAL LONG SHORT TERM MEMORY

According to (Alex Graves, et., 2005) the basic idea of bidirectional recurrent neural nets is to present each training sequence forwards and backwards to two separate recurrent nets, both of which are connected to the same output layer. This means that for every point in a given sequence, the BRNN has complete, sequential information about all points before and after it. Also, because the net is free to use as much or as little of this context as necessary, there is no need to find a (task-dependent) time-window or target delay size for temporal problems like speech recognition, relying on knowledge of the future seems at first sight to violate causality … How can we base our understanding of what we’ve heard on something that hasn’t been said yet? However, human listeners do exactly that. Sounds, words, and even whole sentences that at first mean nothing are found to make sense in the light of future context. One shortcoming of conventional RNNs is that they are only able to make use of previous context. Bidirectional RNNs (BRNNs) do this by processing the data in both directions with two separate hidden layers, which are then fed forwards to the same output layer. Combining BRNNs with LSTM gives bidirectional LSTM, which can access long-range context in both input directions



*Fig 3.3 Diagram of a bidirectional neural network*

**4.0 RESEARCH DESIGN**

This section will highlight the research approach, methodologies and development tools and techniques used in the context of the research question to fulfil stated objectives. The philosophy of this research is **pragmatic** as it entails performance measurement and evaluation. A problem has been identified which is the proper classification of fake news and authentic news, and different techniques are being tested to measure impact on the accuracy of classification. The thought behind this research is to evaluate that some techniques, when applied to a long short term memory networks, will improve accuracy of image classification. This constitutes **deductive** reasoning. The result of the research is either true to back up initial theory or false. This involves theory verification or falsification, Saunders, Bristow, Lewis & Thornhill, (2015).

4.1 METHODOLOGY

The concept of fake and authentic news classification is qualitative in nature because it involves sorting images into categorical groups. However, the process of fine-tuning dataset to approach homogeneity and the outcome of results, which is a measure of accuracy via numerous parameters is quantitative. Qualitative data meets quantitative techniques thereby constituting a **mixed-method** methodological choice.

4.2 RESEARCH STRATEGY

The research is **experimental** as different adjustments and iterations of the long short term memory network algorithms are made on the texts from dataset to measure different outcomes

Time Horizon is **longitudinal**. Research data will be collected over an extended duration of the research on the same dataset.

4.3 DATA COLLECTION AND ANALYSIS

The fake news dataset from Kaggle is used for the purpose of this research. It is a collection of various texts. It solves the problem of lack of variety and difficulty in accessing such data in large volume. It is publicly available through Kaggle, which solves the privacy issue. It contains texts, which are labelled as 0 and 1. 0 meaning fake news and 1 meaning authentic news. Lots of other mediums are currently researched to add more to the amount of data points for better predictive power.

4.4 TOOLS AND TECHNIQUES

The research is centred around a machine learning framework and the choice of programming language that would be used is **Python**. This is chosen because of its popularity and longevity in the use of machine learning research, this implies python has a very large community, its relative simplicity compared to other languages and its massive collection of libraries for data transformation and visualization. These libraries include Scikit-learn, Keras, Pandas, Tensorflow and Matplotlib. An alternative programming language to python are **R,** which is good for visualization, also has a good collection of libraries but a steeper learning curve in comparison.

**Google drive** is the online storage option as it provides a free, stable and ample online storage from google.

**Google Colab** would be used as the development environment rather than a local pc because it provides a virtual space with CPU and GPU capabilities (more powerful than most personal computers) needed to perform machine learning functions. It is also easy to link to google drive to access files. A well-equipped local PC can be used as an alternative.

The Machine Learning model to be used is a Long Short Term Memory Networks (LSTMs) with different techniques employed to improve model accuracy, these include tuning the batch size and number of epochs.

Result from experiment will be in the form of a comparative analysis.

**5.0 ETHICS RISKS AND ISSUES**

5.1 RISKS

Project risks and mitigation measures outlined below

|  |  |  |
| --- | --- | --- |
| **Risk** | **Impact/Likelihood** | **Mitigation** |
| Loss of data via local PC | High/Low | Use google drive as an online storage solution |
| No access to development environment via local PC | High/Low | use Google Colab as an online development environment solution |
| No access to dataset | High/Low | Publicly available dataset considered that has mirror download sources |
| Underestimating research complexity | Medium/Medium | Conduct prior research on feasibility of project completion with intended methodology |
| Delay in project completion | Medium/Low | Design a proper scheduled plan |
| Lack of communication with supervisor | Medium/Low | Adhere to University guidelines and ensure meeting schedules are met with supervisor |

5.2 ETHICS

There are little to no ethical concerns on the research carried out. The content of the dataset is texts of different articles; there is no means of identification to who wrote the articles. It is also publicly available for use on the data science platform Kaggle for research purposes.

**6.0 TIME PLAN**

This project would be split into **5 phases**

* Project Initiation
* Data Preparation
* Algorithm/Model
* Results
* Project Closure

|  |  |  |
| --- | --- | --- |
| **PHASE** | **SUB PHASE** | **PERIOD** |
| **PHASE 1**  PROJECT INITIATION | Project Initiation | JULY 2022 |
| Dissertation topic selection |
| Research Question |
| Feasibility study |
| Dataset availability |
| Academic Research |
| Project Proposal |
| **PHASE 2**  DATA PREPARATION | Data retrieval | AUGUST 2022 |
| Prepare Dev Environment |
| Data Analysis/Pre-processing |
| **PHASE 3**  ALGORITHM/MODEL | Data Transformation/Splits | SEPTEMBER 2022 |
| Training |
| Testing and Validation |
| Optimization |
| **PHASE 4**  RESULTS | Comparative Analysis | OCTOBER 2022 |
| **PHASE 5**  PROJECT CLOSURE | Conclusion | NOVEMBER 2022 |
| Reporting/Dissertation |
| Dissertation Submission |

**REFERENCES**

Yoshua Bengio, et al., (1994). *Learning Long-Term dependencies with Gradient Descent*

Hassim Sak, et al., (2014). *Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling.*

Alex Graves, et al., (2015). *Frame wise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures.*

Di Wang & Eric Nyberg., (2015). *A Long Short-Term Memory Model for Answer Sentence Selection in question answering.*

Kyunghyun Cho, et al., (2014). *Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation.*