Universitetet
i Stavanger

FACULTY OF SCIENCE AND TECHNOLOGY

MASTER'S THESIS

Study programme/specialisation: Computer Science	Spring / Autumn semester, 2020 Open/~~Confidential~~
Author: Henrik Lessø Mjaaland	…………………………………………… (signature of author)
Programme coordinator: Vinay Jayarama Setty Supervisor(s): Vinay Jarama Setty	
Title of master's thesis: Detecting Fake News and Rumors in Twitter Using Deep Neural Networks	
Credits: 30 ECTS	
Keywords: Deep Learning Fake News Detection Natural Language Processing	Number of pages:55…………… + 5 Stavanger, 15 June 2020

Universitetet
i Stavanger

Faculty of Science and Technology

Department of Electrical Engineering and Computer Science

Detecting Fake News and Rumors in Twitter Using Deep Neural Networks

Master's Thesis in Computer Science

by

Henrik Lessø Mjaaland

Internal Supervisor

Vinay Jayarama Setty

June 15, 2020

“If you can’t fly, then run,

if you can’t run, then walk,

if you can’t walk, then crawl,

but whatever you do,

you have to keep moving forward. ”

Martin Luther King Jr.

Abstract

The scope of this paper is to detect fake news by classifying them as either real or fake based on article content, tweets and retweets of news articles from the Politifact dataset using graph neural networks.

I developed multiple graph neural network based models, and the most effective one was based on LSTM using pre-trained GloVe Vectors. Before applying LSTM using the Keras Sequential model, I combined the categorical and continuous attributes into one input vector after word embedding the textual categorical attributes, normalizing and discretizing the continuous attributes.

Fake news spreads rapidly because they tend to have a shocking and dramatic theme and language, and they also have more cascade hops than real news, meaning they are retweeted more than real news. Real news on the other hand, have a more constant and slow spread, and the reach is not as wide as for fake news.

Considering the rapid spread of fake news, it is imperative to identify fake news in a timely manner before the twitter reach of the fake news escalates when preventing misinformation. Fake news can be identified primarily by analyzing the language used in the news articles.

Preface

This thesis is the culmination of my master's degree in Computer Science and the result of my fascination for computers and social media. I took an interest in computers at a young age, and what first sparked my interest in computers is how small they make the world. At an age of eight, I was already chatting with people from all around the world. As social media emerged, the world was made even smaller. Information was made more accessible to the public as social media can be used to share information, but they can also be used to share fake news.

Recently, I have taken an interest in deep neural networks. What interests me about deep neural networks is that they can perform very well when the data size is large enough, they can combine features without being explicitly told to do so, thus, simplifying the process of feature engineering, and they can be used to solve complex problems such as speech recognition, image classification, and natural language processing which can be used for identifying fake news as is done in this thesis.

I would like to express my deepest gratitude to Vinay Jayarama Setty for introducing me to deep neural networks and giving me the necessary tools to write this thesis, and for guiding me throughout the semester by giving valuable feedback.

I would also like to thank my friends and family for motivating and supporting me.

Contents

Abbreviations

NN	**N**eural **N**etworks
DNN	**D**eep **N**eural **N**etworks
RNN	**R**ecurrent **N**eural **N**etworks
CNN	**C**onvolutional **N**eural **N**etworks
LSTM	**L**ong **S**hort **T**erm **F**requency
TF	**T**erm **F**requency
IDF	**I**nverse **D**ocument **F**requency
NLP	**N**atural **L**anguage **P**rocessing

Chapter 1

Introduction

https://www.youtube.com/watch?v=LJxqxtU_EDg

According to the Cambridge Dictionary, fake news are "stories that appear to be news, spread on the internet or using other media, usually created to influence political views or as a joke".

Fake news is a problem of increasing concern in today's world. Fake news today spread like wildfire. They are usually spread by bots, and they spread exponentially per cascade hop in social media. Creating your news by manipulating videos and images is only getting easier and easier. 'Anyone' can forge their own news today. Lately, fake news has become a problem of concern especially in Norway, where sharing fake news through memes on social media has become a trend.

Recently, there have been a lot of fake news regarding the Corona Virus from China. The most widespread ones are that that the Chinese were infected from eating bats, and that there is already a vaccine for the virus, which is not true, since there are seven kinds of Corona Viruses, and the one causing the outbreak now is a new kind of Corona Virus.

URL{https://www.kofiannanfoundation.org/supporting-democracy-and-elections-with-integrity/annan-commission/post-truth-politics-afflicts-the-global-south-too/}

When it comes to fake news during elections, attention has mostly been diverted to the West, leaving the global south unchecked. This has led to an increase of 40 percent spread of fake news in India during election times.

URL{https://www.theguardian.com/world/2019/oct/30/whatsapp-fake-news-brazil-election-favoured-jair-bolsonaro-analysis-suggests}In Brazil, 2019, the vast majority of false information shared on WhatsApp in Brazil during the presidential election favoured the far-right winner, Jair Bolsonaro. Out of 11,957 viral messages shared across 296 group chats on the instant-messaging platform in the campaign period, approximately 42 percent of rightwing items contained information found to be false by factcheckers.

In the global South, the preferred messaging app is WhatsApp. WhatsApp offers encrypted peer-to-peer messaging and is very challenging to monitor, unlike apps like Facebook. Monitoring users' conversations violates the users' privacy, but is useful for detecting fake news and their source.

https://www.gouvernement.fr/en/against-information-manipulation Multiple initiatives has been taken to battle fake news such as the law against information manipulation. The law prevents influencing of election results, which was done e.g. during Brexit.

The law prevents political influencing by requiring “a ‘transparency obligation for digital platforms’, who must report any sponsored content by publishing the name of the author and the amount paid. Platforms exceeding a certain number of hits a day must have a legal representative in France and publish their algorithms.”[], and by requiring a judge to qualify fake news based on these three criterias:

“- the fake news must be manifest,
 - be disseminated deliberately on a massive scale,
 - and lead to a disturbance of the peace or compromise the outcome of an election.”[].

The law also demands a cooperation between the different digital platforms during election periods. The French Broadcasting Authority is responsible for enhancing law enforcement for this requirement.

Furthermore, Emmanuel Hoog, former president of the French Press Agency was entrusted with producing a press ethics body.

Measures against fake news have also been taken in Finland, Malaysia and Singapore. Finland launched an anti-fake news initiative that aims to teach how to identify fake news already in 2014. The Malaysian government passed the Anti-Fake News Act in 2018. The Act was however accused of seeking to stifle criticism of the administration. In 2019, another law against fake news was passed in Singapore called the Protection from Online Falsehoods and Manipulation Act. This law received the same allegations as the law against fake news from Malaysia.

1.1 Motivation

Fake news has had a massive impact on the news industry since the dawn of the Internet. They are under increasing pressure to produce more material faster, and today they tend to use social media as a source of information. When using social media as a source of information, verifying the authenticity of the news is a big challenge for reasons such as the pressure to produce more, and when journalists incorrectly verify news sources, their brand’s reputation is tarnished and people lose trust in the news media. A brand’s reputation takes years to build up, but only seconds to tear down. The pressure to produce material fast also leads to more unserious journalism.

Fake news is a many-faced and complex issue. They can be spread to fuel conflicts, for economic gains, defamation or for political purposes. Also, fake news lead to people in general being less well informed.

URL{https://www.bloomberg.com/news/articles/2019-11-11/how-fake-news-is-stoking-violence-and-anger-in-hong-kong}

In 2019, fake news regarding the death of 22-year-old Chinese student Alex Chow circulated, claiming that he had not committed suicide, but was in fact chased and/or pushed off a parking garage by the Hong Kong police force. The news also claimed that officers had blocked off an ambulance from reaching Chow. The purpose of these fake news was to incite anti-government protests.

URL{https://www.businesstimes.com.sg/opinion/fake-news-can-make-or-break-stock-prices}

In 2013, fake news regarding two explosions in the White House that had injured the earlier US president Barack Obama led to the Dow Jones Industrial Average dropping by 143.5 points, which corresponds to the sum total of 130 billion US dollars in stock value.

Donald Trump legislated CNN and Washington Post as "fake news" because they were not supportive of him (they referred to his followers as a cult and claimed that he had "bewitched" the republican party on multiple occasions).

\url{https://www.cigionline.org/articles/beware-fake-news?gclid=EAIaIQobChMIl_X4nZn-5wIVkRsYCh11MQueEAAYASAAEgI7LPD_BwE}

Fake news has always been used in political contexts, such as spreading false poll numbers to discourage people from voting during political elections, e.g. in Mexico in 1988 by PRI (Institutional Revolutionary Party), but never on a scale and magnitude as with today’s technology.

During the 2016 presidential election in USA, which ended in the election of Donald Trump as the US president, there was a surge of fake news on social media. Investigations after the presidential election points to Russian influence during the campaign.

Another case of using fake news as a means of political influence, is back in 2014 when ISIS started sharing propaganda on every social media imaginable.

Democracy is built trust in the public's capacity for reasoned communication. Digital technologies have made information more accessible to the public, however as deep fake news is getting more and more advanced, people are getting less and less well informed, thus threatening democracy. Political influencing, which was done by PRI, Russia and Bolsonaro as discussed above, also threatens democracy by centralizing the power and taking it away from the people.

1.2 Problem Definition

This thesis investigates if articles' content and their respective rumour propagation paths can be utilized to improve fake news detection. The goal is to achieve this by modelling a graph neural network in order to classify the articles as either real or fake based on patterns in article content and sharing of tweets. The articles are also classified based on a series of descriptive parameters for both the articles and their respective tweets and retweets such as article id, metadata, tweet and retweet identification, tweet- and retweet content, and follower count.

1.3 Usecases/Examples

	created_at	text	contributors_enabled_user	...	followers	following	label
0	2015-10-09 21:18:53+00:00	Missed @marcorubio in NH this week? Come to a ...	False	...			real
1	2013-03-22 06:36:15+00:00	2013-HB-4469 Public employees and officers; et...	False	...			real
2	2008-03-03 06:10:37+00:00	Strong Words in Ohio as Obama and Clinton Pres...	False	...			real
...	...	...	...	...	...	...	...
790	2017-08-29 17:54:38+00:00	Floyd Mayweather Jr. donates a whopping sum of...	False	...			fake
791	2018-02-21 18:47:48+00:00	@TylerHuckabee @LagBeachAntifa9 @AntifaNantuck...	False	...	1140757886706102273, 289419759, 1450422866, 11...	1140757886706102273, 289419759, 1450422866, 11...	fake
792	2018-02-23 18:45:47+00:00	Obama Announces Bid To Become UN Secretary Gen...	False	...			fake

Figure 1: Sample Data

1.4 Challenges

There are many challenges when identifying fake news. Fake news platforms usually imitate real news platforms by design and content.

Fake news articles are not always one hundred percent fake, they might have some true statements mixed with false statements.

With today's technology, images and videos can be fabricated, making it impossible to instantly verify authenticity of news.

It is also easy to create fake entities to share fake news. There are free tools available online for this purpose, and existing photos can be used for face swap.

Fake news articles often contain more dramatic and intensive language than real news, hence **natural language processing (NLP)** can be used to identify fake news. NLP is however a complex and time-consuming process as words can have different meanings in different contexts, and each language and dialect requires a separate version of NLP.

The amount of data available is limited, and there are more real than fake news available. This affects training of classification models.

https://www.forbes.com/sites/forbestechcouncil/2020/01/02/detecting-fake-content-one-of-the-biggest-challenges-for-2020/

1.5 Contributions

In the past, classification and regression models have been used the most for fake news detection. This thesis, however, presents a deep neural network model based on Long Short Term Memory (LSTM). Continuous numeric attributes are normalized in the range [-0.5,0.5], and then discretized, before word embedding them along with categoric attributes and textual attributes. The word embedded attributes are fed into the model which uses them to classify news articles from the Politifact dataset as either real or fake.

Also implemented …..

High accuracy despite empirical studies showing to … being ineffective for fake news detection.

1.6 Outline

Chapter 2 introduces theories related to the work done in this thesis including neural networks, and related works.

Chapter 3 describes the method used in the proposed model in detail, and explains why the given method is used.

Chapter 4 presents the dataset, the experimental setup and the results of the experiments.

Chapter 5 discusses the results, models used as a basis and deviations from these models, implications of the results, and finally strengths and weaknesses of the proposed model.

Chapter 6 provides a summary of the thesis results and an analysis of the results, answers the research question, and points to future directions.

Chapter 2
Background

This chapter discusses relevant theories for this thesis. First off, neural networks are introduced. Next, neural networks used for NLP is discussed. Finally, different accuracy measures are presented.

2.1 Neural Networks

Neural networks are based on the human brain in the sense that they are algorithms consisting of multiple nodes connected by edges, mimicking neurons. Neural networks with minimum one hidden layer, are referred to as deep neural networks. Hidden layers are the layers between the input layer and the output layer (see Figure 2). For now, the required number of hidden layers for a neural network be considered ‘deep’ is more than one, but this number will likely increase soon. Deep neural networks enable computers to artificially learn, hence the term, Artificial Intelligence (AI). Deep neural networks can for example be used to learn features from data such as text, images, sound or they can be used for classification and regression.

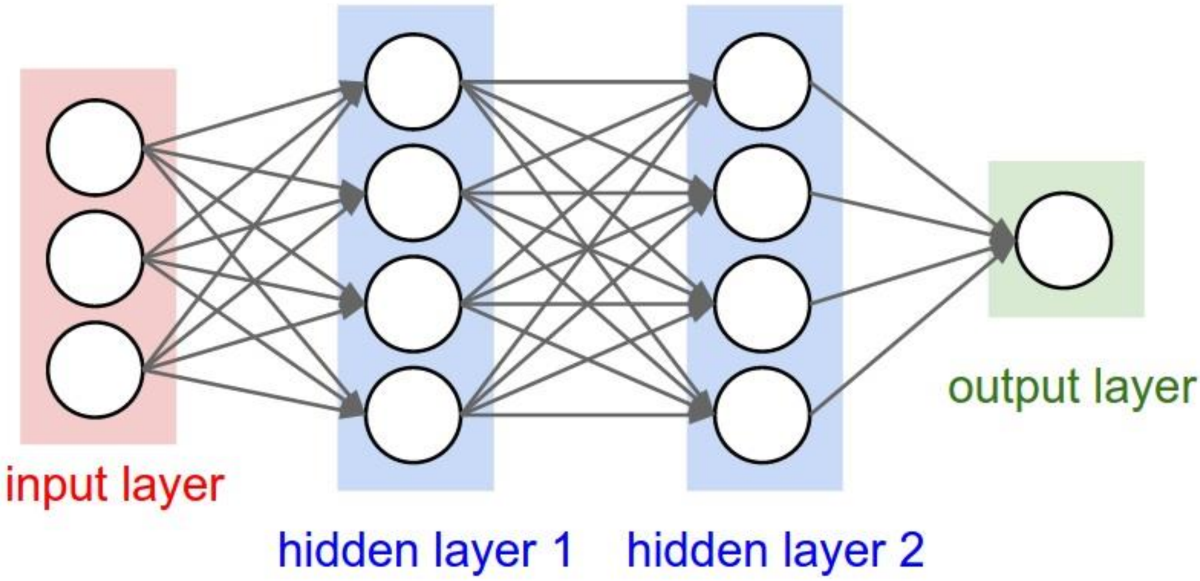

Figure 2: Neural Networks Layers [https://i.stack.imgur.com/Kc50L.jpg]

As mentioned above, neural networks have many uses, but in this case the target values are either real or fake, which means that this is a binary classification problem.

All the nodes in the hidden layers and the output layer, have each their classifier. The classifiers activate the respective nodes based on the inputs from the previous layer and an activation function. There are many activation functions, depending on numerous factors. This thesis uses **ReLU (Rectified Linear Unit)** followed by Sigmoid. ReLU is the most used activation function.

2.1.1 Feed-Forward (Forward-Propagation)

The first hidden layer receives input from the input layer, and then classification scores are passed on from layer to layer. Only activated nodes pass on their classification scores to the next layer. In the output layer, the target values (in this case real or fake) are classified based on the final classification scores, which means that the output is the predicted class. This process is referred to as feed-forward or forward-propagation.

2.2.2 Back-Propagation

Back-propagation is the process of computing the gradients with respect to the weights. While forward-propagation is used for computing outputs, the purpose of back-propagation is to minimize the cost function. This is done by adjusting the weights and biases backwards.

2.2.3 Gradient-Descent

For an unweighted neural network, a given set of inputs will always result in the same output. In a weighted neural network however, all the nodes have different weights, each resulting in a unique output. The gradient descent algorithm runs forward- and back-propagation multiple rounds or epochs, each time tweaking the weights in order to improve the classification accuracy as explained above.

Running forward propagation with a high number of epochs is time-consuming and one should find a balance between accuracy and number of epochs, where accuracy is maximized, and number of epochs is minimized.

Too many epochs can also result in overfitting. This can be prevented by plotting training- and testing loss. Overfitting is when the training loss is less than the testing loss, and underfitting is when the training loss is higher than the testing loss. The testing- and training loss should be about the same.

2.2 Neural Networks for NLP

2.2.1 Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF is the product of TF and IDF. TF is the number of terms across all the documents for each term. IDF is the inverse number of terms per document. TF-IDF says how rare a specific term is; the higher the TF-IDF is, the rarer the given term is. The most common words will thus have a low TF-IDF.

2.2.2 Word Embedding

Word embedding is the process of converting terms to vectors of float numbers. This is done to improve the accuracy of **Artificial Neural Networks (ANN)** because ANNs are built for continuous numeric inputs, and similarities between continuous vectors with more ease than similarities between terms. A common technique for word embedding is **GloVe (Global Vectors)**. GloVe is an algorithm for converting unlabelled data such as words to continuous vectors which reduces dimensionality. GloVe vectors are pre-trained on Wikipedia and Gigaword 5, and are good at capturing semantics

https://medium.com/@sthacruz/fake-news-classification-using-glove-and-long-short-term-memory-lstm-a48f1dd605ab

2.2.3 Recurrent Neural Networks

Recurrent neural networks consist of multiple **feedforward neural networks (FNNs)**, and each FNN is considered a time-step. Recurrent neural networks use the gradient descent algorithm to minimize error. However, since there are many time-steps that all share the same parameters now, **backpropagation-through-time (BPTT)** is used instead of normal backpropagation. The equation for BPTT contains an extra error gradient because of the time steps.

Traditional neural networks are not good for predicting based on sequential data such as speech, time series, or text, because they have no memory, and sequential data is ordered. Recurrent neural networks introduce memory by using loops between the hidden layers. The output from the loops are stored in the internal state (see figure 3).

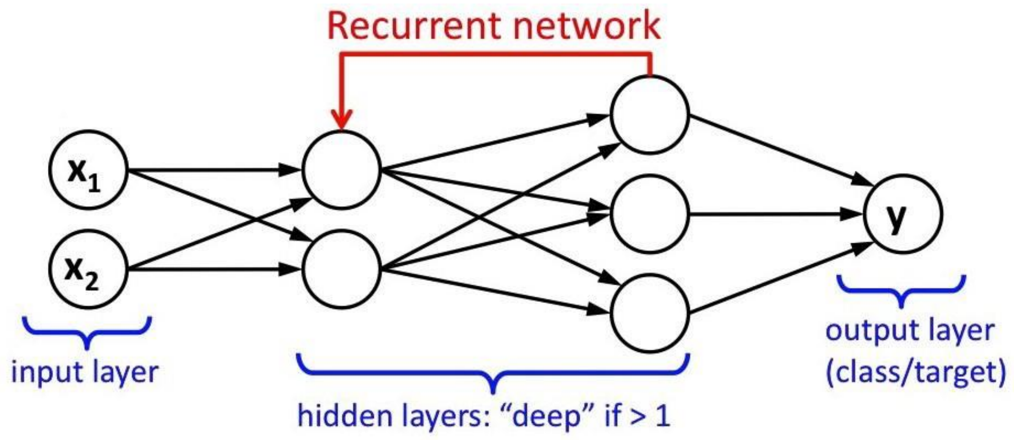

Figure 3: Recurrent Neural Network with Loops [https://towardsdatascience.com/implementation-of-rnn-lstm-and-gru-a4250bf6c090]

2.2.4 Vanishing Loss Gradient

When training RNNs with multiple layer a problem called the 'Vanishing Loss Gradient' can arise. The problem is as the name suggests, when the gradient 'vanishes', meaning that it's close to zero. The problem arises if many gradients are close to zero, because the gradients are products of previous gradients. The problem usually occurs in the earlier layers (gradients are computed using backpropagation). This is a problem because the weight updates are the subtraction of the gradients multiplied with the learning rate, from the weights themselves. This means that if the gradients are close to zero, the weight updates will have little to no effect at all. The vanishing loss gradient is not a problem for the ReLU activation function, hence we chose to use ReLU in the earlier layers and Sigmoid in later layers (see chapter 2.1). ReLU converts values above zero to the original value, and negative numbers to zero, while Sigmoid converts values above zero to one, and negative numbers to zero, and is thus more computationally taxing than ReLU.

2.2.5 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a type of recurrent neural network that solves the vanishing gradient problem by introducing long-term memory. As discussed in 2.2.3, recurrent neural networks have multiple time-steps and a cell for each time-step. LSTM networks on the other hand have a memory cell for each time-step (see figure 4).

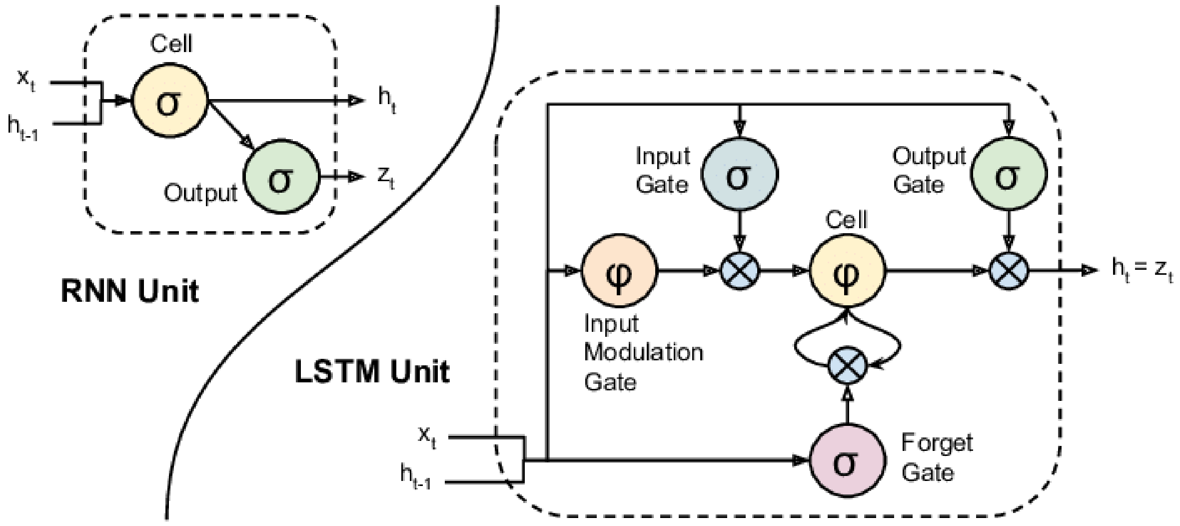

Figure 4: RNN Cell and LSTM Memory Cell [https://www.researchgate.net/figure/A-diagram-of-a-basic-RNN-cell-left-and-an-LSTM-memory-cell-right-used-in-this-paper_fig1_319770438]

As depicted in figure 4, the memory cell contains an input gate, an output gate, and a forget gate. The input gate chooses which information to remember, and the output gate decides which information to output. The key to long-term memory is the forget gate. In traditional RNNs, all inputs/information are accepted. In this case, previously accepted relevant information might be substituted with new irrelevant information. The forget gate in LSTM networks prevents this problem by filtering irrelevant information, thus providing long-term memory. Memory cells also have an input modulation gate (this gate is often considered a sub-gate of the input gate). This gate normalizes the input information to increase convergence speed.

A con with LSTM is that it can cause redundant overhead because some predictions requires less context than others. An example is “the sun is a star”, when predicting the last word, ‘star’, no more information is needed. An example where more context, and thus LSTM, is needed, is “I like to watch soap operas… My favourite TV-show is ‘Friends’”. In this case the first statement is useful when predicting the last word of the second statement.

2.3 Accuracy Measures

A confusion matrix is an accuracy measure for labelled datasets. Confusion matrices can be computed for multiple classes, but for a two-class problem, it will have four cells containing **True Positives (TP)**, **False Positives (FP)**, **False Negatives (FN)**, and **True Negatives (TN)**, see figure 4.

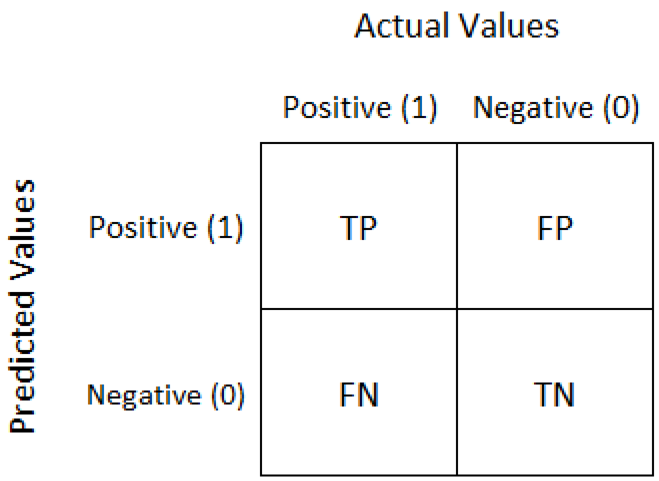

Figure 5: Confusion Matrix [https://www.google.no/search?q=confusion+matrix&tbm=isch&ved=2ahUKEwj9uISIk5LoAhWHapoKHcSIB04Q2-cCegQIABAA&oq=confusion+matrix&gs_l=img.3..35i39l2j0i30l8.96895.98917..99196...0.0..0.120.1171.17j1......0....1..gws-wiz-img.......0i19.q]

TP is the number of correctly predicted positive (real) records (articles).

FP is the number of wrongly predicted positive records.

FN is the number of wrongly predicted negative (fake) records.

TN is the number of correctly predicted negative records.

Precision is the number of correctly predicted positive records divided by all the records predicted as positive:

$$Precision = \frac{TP}{TP + FP}$$

Recall is the number of correctly predicted positive records divided by all the records that are actually positive:

$$Recall = \frac{TP}{TP + FN}$$

Accuracy is computed as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

and is the percentage of correctly predicted records for both classes.

F1-Score is a useful accuracy measure if you want a balance between precision and recall, and the class distribution is uneven. The formula for F1-Score is given below:

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

2.4 Related Works

https://matheo.uliege.be/bitstream/2268.2/8416/1/s134450_fake_news_detection_using_machine_learning.pdf

https://www.ijitee.org/wp-content/uploads/papers/v8i10/J94530881019.pdf SVM – 75.5, DT – 82.7%, Kaggle fake news challenge -1,

The last few years, there has been a lot of research on detecting fake news in social media. Many of these works have used machine learning techniques like decision trees and linear SVM as [https://www.ijitee.org/wp-content/uploads/papers/v8i10/J94530881019.pdf] and [https://arxiv.org/pdf/1905.04749.pdf]. The works achieved accuracies of 82.7% and 65% respectively using decision trees. When using SVM, accuracies of 75.5% and 66% was achieved.

Some of these works focus on article content like [https://arxiv.org/ftp/arxiv/papers/1901/1901.09657.pdf], which achieved an accuracy of 62.4% using Fakebox (a machine learning model) and McIntire's fake-real-news-dataset.

Lately, research has also been conducted using deep learning. These works focus mostly on rumour propagation paths, and not so much on content as [file:///C:/Users/henri/Downloads/16826-76467-1-PB.pdf] and []. [file:///C:/Users/henri/Downloads/16826-76467-1-PB.pdf] claims to be able to detect fake news with accuracies of 85% and 92% on Twitter and Sina Weibo respectively in 5 minutes after they begin to spread.

A graph neural networks-based work focusing on image data, obtained an accuracy of 94%, using a Convolutional Neural Networks (CNN) based model [file:///C:/Users/henri/Downloads/Fake_news_detection_using_Deep_Learning%20(1).pdf]. [file:///C:/Users/henri/Downloads/Fake_news_detection_using_Deep_Learning%20(1).pdf] also obtained an accuracy of 91% using LSTM based on content and title of articles. The work used fake news articles from the dataset, 'Getting Real About Fake News', and real news articles from sources such as 'The New York Times' and 'The Washington Post'. Some works have also applied TF-IDF for fake news detection such as [https://arxiv.org/pdf/1905.04749.pdf], where an accuracy of 95% was achieved using unigram Naïve on a combination of the 'Liar Liar'- and 'Fake or Real News' datasets.

Chapter 3

Method

3.1 Solution Approach

Content, TF-IDF, and rumour path propagation have been successfully used in previous works, and this thesis combines these features (see figure 7). Article content and rumour path propagation related attributes are first read from the Politifact dataset and merged into a data frame ordered by articles as shown in figure 1.

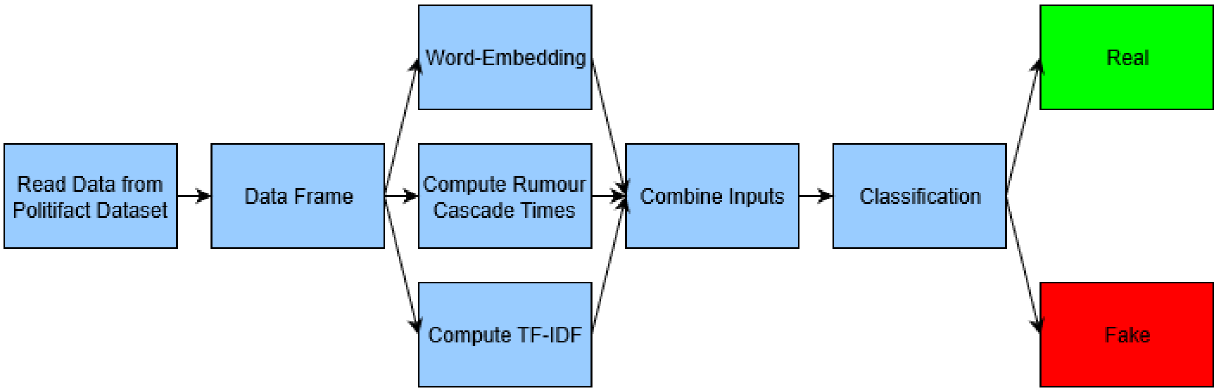

Figure 6: Proposed Solution

Article content is word-embedded (see 2.2.2). The content of each article corresponds to a vector with a length of one thousand (including padding). The weights for each of the vectors are generated using GloVe vectors. The vectors are fed into an embedding layer (see figure 8), and output as a matrix. Embedding layers are good at capturing sentiments.

TF-IDF is computed based on the articles' content (see 2.2.1). TF-IDF vectors are computed for all the terms in each article. Fake articles tend to have higher TF-IDF values, because they usually contain more unique words than real articles. This is most likely because fake article writers have a habit of using pretentious or even vulgar language in order to captivate readers and spread the article as fast as possible.

3.1.1 Cascade Propagation Path

Time series containing time in seconds passed per rumour cascade is computed for each article. A rumour cascade is simply put either a tweet or a retweet of an article. Studies have proven that fake articles are spread faster than real articles, meaning that fake articles should have less time passed per rumour cascade than real articles.

3.1.2 Layers

After combining article contents, time series, and TF-IDF vectors, these combined features are passed as input for the input layer. The input layer is followed by the embedding layer (see 3.1.1). The next layer is an LSTM layer with sixty neurons. There is no fixed rules for setting the number of neurons in the hidden layers in the LSTM cell, but in order to prevent overfitting, the number should be below

$$Nh = \frac{Ns}{\alpha * (Ni + No)},$$

where Ni is the number of inputs neurons, No is the number of output neurons, Ns is the number of samples in the training data set, alfa is an arbitrary scaling factor, usually between two and ten. Next up, is a 'GlobalMaxPool1D' layer, which downsamples by computing the most present features (maximum values) for each feature map. After the Max-Pooling layer is a dropout layer with a dropout of '0.1', meaning that ten percent of the neurons or inputs are dropped. Dropout is a regularization technique that prevents overfitting by dropping a random set of hidden units or nodes at each update during training. Nodes are dropped by setting them to zero. Finally there are two dense layers with another dropout layer between them for further regularization that classifies the articles as either fake or real. The first of these dense layers 50 neurons and apply the ReLU activation function. The last dense layer has 2 neurons and applies the Sigmoid activation function. The ReLU activation function is applied before Sigmoid to prevent the vanishing loss gradient (see chapter 2.2.4).

https://ai.stackexchange.com/questions/3156/how-to-select-number-of-hidden-layers-and-number-of-memory-cells-in-an-lstm

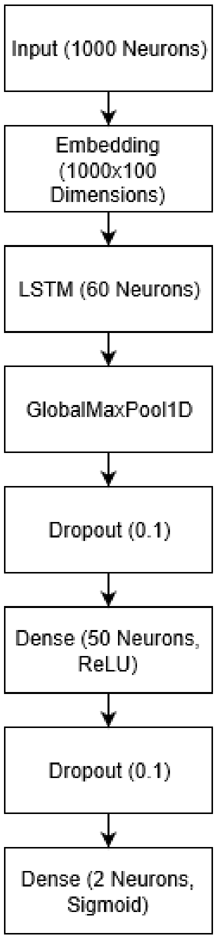

Figure 7: LSTM Graph

https://www.diagrameditor.com/

Chapter 4
Experimental Setup and Data Set

This chapter introduces the data set and the development environment used, and presents the experimental results.

4.1 Technologies

The code for this thesis was written in Python, because it has simple syntax rules which makes code easy to read, and it has many libraries that are handy for data science in general such as Pandas, which is good for data representation and can handle large amounts of data efficiently, Pickle, which can be used to save data structures as backups when running time demanding code (TQDM can offer progress bars for such cases), Numpy for scalable and efficient computations, and Pyplot, which offers a large variety of plots for visualization. Python also has good libraries for deep learning specifically, such as Keras.

4.1.2 Natural Language ToolKit (NLTK)

NLTK is a package for natural language processing. It contains for example stopwords from many languages. Stopwords are useless words (they do not give any context or meaning to a given sentence), which are commonly used (such as "the", "a", "and"), and should be removed.

4.1.3 wordcloud

The Wordcloud package is used to visualize text, where the text size corresponds to significance or frequency (in this thesis TF-IDF, thus, significance). Matplotlib is needed to visualize word clouds.

4.1.1 Keras

Keras has many perks. The library is efficient on both CPU and GPU, it has preprocessing features such as tokenizing and padding, and it also has many models, such as max pooling, dropout, embedding, and LSTM. These models can be combined, as they were in this thesis. Keras even have regularizers that can be used to avoid overfitting.

4.1.6 RE (Regular Expressions)

RE allows one to check if a given string matches a regular expression and can be used to remove numerals from strings.

4.1.7 Datetime
The Datetime package contains classes for manipulating dates and time by for example subtracting dates in order to compute time passed.

4.2 Dataset

All the data, including articles, tweets and retweets used for this thesis are retrieved from the Politifact dataset. Politifact was first started by the Tampa Bay Times in St. Petersburg, Florida in 2007. It is run by the Poynter Institute. Politifact is a fact-checking website that rates accuracy of claims by elected officials and others on its Truth-O-Meter [https://www.politifact.com/]. The Truth-O-Meter is a scale ranging from "True" (a hundred percent real) to "Pants on Fire" (a hundred percent fake). The labels in this thesis however, have a binary distribution, either "real" or "fake". The dataset has a total of X articles, of which A is labelled fake and B is labelled real.

4.2.1 Preprocessing
The data undergoes multiple preprocessing steps in this thesis. First off, columns and rows with mostly NAN (Not A Number) values are dropped. Next, the remaining NAN values are estimated. NAN values for continuous features are estimated to the mean of the column, while Nan values for categorical features are estimated to the most common value. Then, symbols, punctuation and stopwords are removed. Finally, pre-trained 6B 100d GloVe vectors are used for word embedding. Before training the LSTM model and classifying, the dataset is split into a training set consisting of eighty percent of the data and a validation set consisting of the remaining twenty percents.

4.2.2 Data Visualization and Analysis

This subsection discusses and analyzes the data visualizations generated from the articles and statistics in the Politifact dataset.

Followers scatterplot, CDF

tfidf CDF, wordcloud

rumour cascade propagation - scatterplot hops and time.

Chapter 5

Accuracies

Thesis guide:

https://student.uis.no/getfile.php/13242348/02.%20Nye%20student.uis.no/Studiehverdag/Vedlegg/M%C3%98AHOV%20Veiledningshefte%20engelsk.pdf

APA 6 - standard

example: