Analysis of the Fairness of Disinformation Detection models concerning the Global North and the Global South

Final Project Evaluation Report

by

Sujit Mandava
(112001043)

INDIAN INSTITUTE
OF TECHNOLOGY
PALAKKAD

COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

CERTIFICATE

*This is to certify that the work contained in the project entitled "**Analysis of the Fairness of Disinformation Detection models concerning the Global North and the Global South**" is a bonafide work of **Sujit Mandava (Roll No. 112001043)**, carried out in the Department of Computer Science and Engineering, Indian Institute of Technology Palakkad under my guidance and that it has not been submitted elsewhere for a degree.*

Dr.Sahely Bhadra

Assistant/Associate Professor

Department of Computer Science & Engineering

Indian Institute of Technology Palakkad

Contents

Chapter 1

Introduction

1.1 Progress so far

The problem statement of the project can be broadly divided into two parts: proving the hypothesis that popular disinformation detection systems show a bias towards data originating from the Global North, and creating a new disinformation detection system that overcomes the bias if any.

So far, we concluded that any bias present in the models would result from a lack of context related to the Global South compared to the Global North and that the algorithm itself did not introduce any such discrepancies. However, the obtained results were not conclusive enough to prove that our hypothesis is indeed true, rather the results lean towards the fact that our hypothesis was false and no real bias exists in these disinformation detection systems. To create a new disinformation detection system that overcomes the bias towards the Global North, it is necessary to prove the existence of said bias and pinpoint the reasons for its appearance.

1.2 Problem Statement

Our goal is to verify that current disinformation detection models show a bias towards Global North data and pinpoint the origin of the said bias. Suppose there is conclusive evidence supporting the existence of said bias. In that case, the next task will be to create a new disinformation detection system using the obtained evidence as the starting point. Suppose there is conclusive evidence pointing in the opposite direction. In that case, we can conclude that there is no bias shown by disinformation detection systems towards the Global North. Therefore current models can be adopted on a global scale to tackle the same, albeit with minor case-to-case adaptations.

1.3 Organization of The Report

Chapter 1 gives us a brief overview of the work done in the project in the previous term and the following report's overview. Chapter 2 gives us a brief overview of the literature review and the work done in the previous stages of the project. Chapter 3 gives us an insight into the work done this semester. It talks about the different tests done and the conclusions that can be drawn from them, in line with the problem statement defined in Chapter 1. Chapter 4 draws a conclusion for the report and the project as a whole.

Chapter 2

Previous Work: Summary

2.1 Literature Review: Content-Based Disinformation Detection

Content-based detection methods aim to extract features from the semantics involved in the text and classify the articles based on the same. Studies have shown that fake news tends to be more provocative and subjective while real news is more specific and objective. The length of the content, the total number of words, abbreviations, and other similar linguistic features play a major role in determining the nature of the content.[8]

Various methods have been utilized to identify and counteract the effects of misinformation. Machine learning techniques, such as supervised and unsupervised learning, have become prominent in distinguishing between genuine and deceptive content. Classification models like Support Vector Machines, Naive Bayes' classifiers, and deep learning models using RNNs, CNNs, and LSTM networks are commonly used. Deep learning models generally outperform traditional machine learning models by approximately 6%.[8][9]

Natural language processing (NLP) methods play a significant role in analyzing linguistic patterns, sentiment, and semantic structures to detect anomalies indicative of disinformation. Deep learning models often employ pre-trained word embeddings to enhance decision-making. Fine-tuning pre-trained models for specific tasks has gained traction due to its success in improving performance.[10] For instance, Jamal et al. (2020)[11] proposed

a hybrid CNN-RNN approach for fake news detection, outperforming existing methods. Other research papers explore combining word embeddings with deep learning, such as "FNDNet – A deep convolutional neural network for fake news detection."[7] "DISCO: Comprehensive and Explainable Disinformation Detection"[2] presents a novel approach by combining graph algorithms with word embeddings to extract new features for neural classifiers.

2.2 Literature Review: Public Wisdom Matters! Discourse-Aware Hyperbolic Fourier Co-Attention for Social-Text Classification[1]

An interesting research paper tackling the problem of disinformation detection uses "public wisdom" on social media in the form of comments and replies to aid their decision-making process. It exhibits state-of-the-art performance on 10 benchmark datasets. However, its need to leverage public comments and replies limits its use to social media, and its true potential can only be revealed when the number of users interacting with the post is high. Combining this with a state-of-the-art content-based disinformation detection model can possibly yield great results.

2.3 DISCO: Comprehensive and Explainable Disinformation Detection[2]

This research paper employs graph and machine learning techniques for disinformation detection by transforming the problem into a graph classification task, where each article is represented as an embedding vector. The salient features of this paper are:

1. Building Word Graphs: Each word in the article forms a node in a word graph. An edge is created between two words if they co-occur within a defined window of text

units. Node features are assigned using pre-trained word embedding vectors.

2. Geometric Feature Extraction: A personalized PageRank vector is computed for each node in the word graph, capturing the importance of each word concerning others. This helps to extract geometric features that jointly model the heterogeneous semantic meanings of words.

3. Graph-Level Representation: Word-level hidden representations are aggregated to obtain a graph-level representation for the entire article. This is done using permutation-invariant functions like sum or average pooling.

4. Misleading Words: It also returns a list of words it deems most "misleading", i.e., words that help camouflage disinformation and hinder the performance of the model. By removing these words, the decision-making process becomes easier. This allows us to compare the writing styles of real and fake news.

This approach is successful because it substitutes the traditional method of exchanging messages between nodes with a more dependable way of amalgamating information. This enhances its effectiveness and speed. It can swiftly adapt to alterations in the graph's structure without requiring retraining, which is crucial for DISCO's explanatory capabilities.

Chapter 3

Progress

3.1 Dataset Generation

This study combines the use of three published datasets alongside data scraped from the internet. The datasets used are ISOT Fake News Dataset[3][4], FakeNewsIndia[6] and the Indian News Dataset[5].

3.1.1 ISOT Fake News Dataset[3][4]

The ISOT Fake News Dataset[3][4] was published by the ISOT Research Lab at the University of Victoria and consists of sociopolitical news gathered from around the world. It consists of both real news, gathered from reputed sources around the world, and fake news that have been flagged by various fact-checking organizations all over the world. A major drawback that can be observed is the lack of information related to the Global South. There is almost no fake news related to the Global South and only about 5% of the real news is regarding the Global South.

3.1.2 Indian News Dataset[5]

The Indian News Dataset is a collection of news originating from the Indian Subcontinent, published by the University of Calicut. This dataset contains around 200 news articles

related to sociopolitical spheres, and this data is added to increase the amount of Global South data.

3.1.3 FakeNewsIndia[6]

FakeNewsIndia is a dataset published by The International Institute of Information Technology Hyderabad consisting of fake news that has been flagged over 3 years from December 2016 to December 2019. This dataset provides our models much the needed context of fake news incidents in the Global South.

3.1.4 Times Of India - Web Scrapping

To supplement the Global South data with data in real-time, we used a simple web scraping tool to gather articles from the Times of India website and selected those related to sociopolitical domains. The code for the same is as follows:

```
import requests
import newspaper
from bs4 import BeautifulSoup
import urllib.request,sys,time
import pandas as pd


def timesofindia():
    url = "https://timesofindia.indiatimes.com/home/headlines"
    page_request = requests.get(url)
    data = page_request.content
    soup = BeautifulSoup(data,"html.parser")
    listArticles = []
    counter = 0
    for divtag in soup.find_all('div', {'class': 'headlines-list'}):
        for ultag in divtag.find_all('ul', {'class': 'clearfix'}):
```

```
        if (counter <= 1000):
            for litag in ultag.find_all('li'):
                counter = counter + 1
                print(str(counter))
                print(str(counter) + " - https://timesofindia.indiatimes.com"
                    + litag.find('a')['href'])
                curr_URL = "https://timesofindia.indiatimes.com" +
                    litag.find('a')['href']

                article = newspaper.Article(url=curr_URL, language='en')
                try:
                  article.download()
                  article.parse()
                  article ={
                    "title": str(article.title),
                    "text": str(article.text),
                    "authors": article.authors,
                    "published_date": str(article.publish_date),
                    "top_image": str(article.top_image),
                    "videos": article.movies,
                    "keywords": article.keywords,
                    "summary": str(article.summary)
                  }
                  listArticles.append([article['title'], article['text']])
                except:
                  continue
    return pd.DataFrame(listArticles, columns=["title", "text"])
```

3.2 Current Dataset Division

The datasets used for confirming whether a bias exists or not remain the same from the previous iteration of tests. The split of the data can be seen in Table 2.1.

Dataset Names	Global North		Global South	
	Real	Fake	Real	Fake
ISOT Fake News[3][4]	16544	23481	4873	-
FakeNewsIndia[6]	N/A	N/A	0	4843
IND[5]	N/A	N/A	200	0
ToI	N/A	N/A	668	0
Total	16544	23841	5741	4843

Table 3.1 Dataset Breakdown

There is a lack of reliable fake news detection methods employed in the Global South, which can be observed from the lack of fake news datasets for data originating from the Global South. This split of data is an accurate representation of the progress made in the sphere of disinformation detection in the Global South.

3.3 Models Used

The models used in the tests are as follows:

3.3.1 Model B/ Simple Model

```
|---------------------------------------------------------|
| Layer                    | Output Shape     | Param #    |
|--------------------------|------------------|------------|
| embedding (Embedding)    | (None, 1000, 100)| 3,000,000  |
| flatten                  | (None, 100000)   | 0          |
| dense                    | (None, 1)        | 100001     |
|---------------------------------------------------------|
```

Model B gives us the control for most of these experiments and allows us to understand how simple NLP models would perform in the current classification task at hand.

3.3.2 FNDNet[7]

Layer	Output Shape	Param #
embedding (Embedding)	(None, 1000, 100)	3,000,000
conv1d_1 (Conv1D)	(None, 998, 128)	38,528
conv1d_2 (Conv1D)	(None, 997, 128)	51,328
conv1d_3 (Conv1D)	(None, 996, 128)	64,128
max_pooling1d_1	(None, 199, 128)	0
max_pooling1d_2	(None, 199, 128)	0
max_pooling1d_3	(None, 199, 128)	0
concatenate(Concatenate)	(None, 597, 128)	0
conv1d_4 (Conv1D)	(None, 593, 128)	82,048
max_pooling1d_4	(None, 118, 128)	0
conv1d_4 (Conv1D)	(None, 114, 128)	82,048
max_pooling1d_4	(None, 3, 128)	0
flatten_2 (Flatten)	(None, 384)	0
dense_3	(None, 128)	49,240
dropout_2	(None, 128)	0
dense_4	(None, 1)	129

FNDNet is a deep convolutional neural network proposed for disinformation detection tasks, and has shown outstanding performance on large-scale real-world fake news datasets.[7]

3.3.3 FNDNet-derivative

Layer	Output Shape	Param #
input_7	(None, 1000)	0
embedding_1	(None, 1000, 100)	3,000,000
embedding_2	(None, 1000, 100)	3,000,000
conv1d_28	(None, 998, 128)	38,528
conv1d_29	(None, 997, 128)	51,328
conv1d_30	(None, 996, 128)	64,128
conv1d_32	(None, 998, 128)	38,528
conv1d_33	(None, 997, 128)	51,328
conv1d_34	(None, 996, 128)	64,128
max_pooling1d_28	(None, 199, 128)	0
max_pooling1d_29	(None, 199, 128)	0
max_pooling1d_30	(None, 199, 128)	0
max_pooling1d_32	(None, 199, 128)	0
max_pooling1d_33	(None, 199, 128)	0
max_pooling1d_34	(None, 199, 128)	0
concatenate_7	(None, 597, 128)	0
concatenate_8	(None, 597, 128)	0
conv1d_31	(None, 593, 128)	82,048
conv1d_35	(None, 593, 128)	82,048
max_pooling1d_31	(None, 19, 128)	0
max_pooling1d_35	(None, 19, 128)	0
concatenate_9	(None, 38, 128)	0
flatten_5	(None, 4864)	0

```
| dense_10          | (None, 128)       | 622,720   |
| dropout_5         | (None, 128)       | 0         |
| dense_11          | (None, 1)         | 129       |
|-------------------------------------------------|
```

The above model is derived from the structure of the FNDNet; the top half of FNDNet is duplicated, and the two halves are connected to a common bottom half.

3.4 Tests

The following tests were undertaken to verify the actuality of the hypothesis:[1]

1. Models are trained using both the Global North and Global South training data without any oversampling/undersampling

2. Individual classes in the Global North dataset and Global South dataset are oversampled so both classes have equal data, after which the Global South dataset as a whole is oversampled by a little bit to increase the amount of Global South data

3. The test setup is identical to test 2, with the exception that the Global North dataset is undersampled to have the same amount of data as the Global South dataset

4. Training and test data is split the same way as test 1; however:

   (a) Only Global North (or) Global South data is used to train the models

   (b) Training data from both datasets are used to train the models, however, the amount of Global North data varies from 0% to 90%

   (c) Training data from both datasets are used to train the models, however, the amount of Global South data varies from 0% to 90%

The results of the above tests allow us to analyze the performance of the models on both Global North and Global South data and also allow us to understand the impact each of the datasets has on the training of the model.

[1]For all tests, Global North and Global South datasets are individually split into training and test data in an 80-20 ratio.

3.5 Test Set Up

For all the tests, the following training and testing pipeline is used[2]:

1. The dataset is first preprocessed using the Keras library. It is tokenized and the sequenced tokens are used to generate a GloVe embedding.

2. The dataset is then split into training data and test data. The training data is further split into training and validation data.

3. The three models are trained using the GloVe embedding and the training data.

4. The performance is then tested on the unseen test data, both together and separately (Global North and Global South), and the performance metrics are collected.

The performance metrics are then tabulated and graphed to simplify the process of analyzing the results and deriving the necessary conclusions.

[2]Notebooks

3.6 Test Results and Observations

3.6.1 Test 1

	Precision		
	Training	Validation	Test
FNDNet	1	0.9969	0.9922
Model B	1	0.9870	0.9817
FNDNet(New)	1	0.9944	0.9931
	Recall		
	Training	Validation	Test
FNDNet	1	0.9958	0.9937
Model B	1	0.9736	0.9767
FNDNet(New)	1	0.9939	0.99304

Table 3.2 Performance Metrics for Different Models

Model	**Precision (GN)**	**Recall (GN)**	**Precision (GS)**	**Recall (GS)**
FNDNet	0.9910	0.9964	0.9955	0.9859
FNDNet(New)	0.9925	0.9946	0.9947	0.9885
Model B	0.9811	0.9853	0.9818	0.9514

Table 3.3 Performance Metrics for Different Models on GN and GS Test Sets

Observe that there is no noticeable difference in the performance of the models on the Global North and Global South test data. While the performance of FNDNet and its derivative is superior to that of Model B, the three models on an individual level perform equally well on both datasets.

3.6.2 Test 2

Test 2 also shows that the models show no noticeable difference in performance against data originating from the Global North and Global South. Even though there appears to be a lack of data from the Global South compared to the Global North, all models seem capable of distinguishing real news from fake news for articles originating from both reasons. This leads to the conclusion that providing current models with adequate context

	Precision		
	Training	Validation	Test
FNDNet	0.9996	0.9958	0.9973
Model B	1	0.9880	0.9841
FNDNet(New)	1	0.9915	0.9841
	Recall		
	Training	Validation	Test
FNDNet	0.9997	0.9851	0.9953
Model B	1	0.9843	0.9859
FNDNet(New)	1	0.9945	0.9859

Table 3.4 Performance Metrics for Different Models

Model	**Precision (GN)**	**Recall (GN)**	**Precision (GS)**	**Recall (GS)**
FNDNet	0.9972	0.9979	0.9976	0.9857
FNDNet(New)	0.9898	0.9967	0.9960	0.9936
Model B	0.9837	0.9915	0.9854	0.9650

Table 3.5 Performance Metrics for Different Models on GN and GS Test Sets

from the Global North and the Global South should enable these models to identify fake news easily.

3.6.3 Test 3

	Precision		
	Training	Validation	Test
FNDNet	1	0.9896	0.9864
Model B	1	0.9804	0.9749
FNDNet(New)	1	0.9920	0.9825
	Recall		
	Training	Validation	Test
FNDNet	1	0.9935	0.9936
Model B	1	0.9712	0.9686
FNDNet(New)	1	0.9906	0.9936

Table 3.6 Performance Metrics for Different Models

The results of test 3 ratify the conclusion performance of the models on both Global North and Global South test sets after undersampling the Global North training dataset, which is similar to that observed in the previous two tests. Although there appears to be

Model	Precision (GN)	Recall (GN)	Precision (GS)	Recall (GS)
FNDNet	0.9783	0.9935	0.9934	0.9946
FNDNet(New)	0.9713	0.9943	0.9936	0.9948
Model B	0.9676	0.9747	0.9822	0.9626

Table 3.7 Performance Metrics for Different Models on GN and GS Test Sets

a minor boost in performance on the Global South test set, this can easily be attributed to the oversampling done on the Global South training and test sets to boost the number of data points and has no real effect on the observations.

3.6.4 Test 4

3.6.5 (a) GN/GS Training Data

All the models are trained using only Global North (or) Global South training data and tested on both Global North and Global South test sets. The results are as follows:

Model	Precision (GN)	Recall (GN)	Precision (GS)	Recall (GS)
FNDNet	0.9994	0.9983	0.9222	0.9428
FNDNet(New)	0.9806	0.9927	0.8462	0.9466
Model B	0.9894	0.9859	0.9379	0.8807

Table 3.8 Performance Metrics for Different Models on GN and GS Test Sets (GN Training Data)

Model	Precision (GN)	Recall (GN)	Precision (GS)	Recall (GS)
FNDNet	0.5623	0.9981	0.9968	0.9976
FNDNet(New)	0.5534	0.9976	0.9952	0.9968
Model B	0.5892	0.9810	0.9876	0.9841

Table 3.9 Performance Metrics for Different Models on GN and GS Test Sets (GS Training Data)

The performance of the models is in line with their expected behavior. However, a few key observations can be made from the above results. The models trained on Global North data exhibit significantly better performance on Global South test sets compared to that shown by models trained on Global South data on Global North test sets. This behavior

is in a sense contradictory to our hypothesis that models show bias to Global North data, as the performance on Global South data on average is better than that on Global North data based on these results. To understand the implications of these results and the effect of Global North and Global South data on model training, we employ tests 4(b) and 4(c) as mentioned in Section 2.3.

3.6.6 (b) Increasing GN data

The training data is generated for this test by combining the Global South training set and incrementing the amount of Global North test data to be added to the training set. The amount of data from the Global North training set ranges from 0-90%.[3] The results of the test can be seen in the table 2.10.

We observe that as the amount of Global North data increases in the training data, the precision of the trained models on the unseen Global North data increases. The effect of the Global North data is so great that by the time about 10-15% of the data is introduced, the precision of the models on the unseen Global North data crosses 0.95 from less than 0.4 in the FNDNet-based models. From there, there appears to be minimal to no change in the models' performance metrics, as they all plateau over 0.95 in the case of Model B and around 0.98 for the FNDNet-based models.

3.6.7 (c) Increasing GS data

The training data is generated for this test by combining the Global North training set and incrementing the amount of Global South training data to be added to the training set. The amount of data from the Global South training set ranges from 0-90%. [4] The results of the test can be seen in the table 2.11.

Observe that while the performance of the models increases as the amount of Global South data increases, the initial performance of the models on Global South test data is

[3] 100% is omitted as it is a replica of test 1.
[4] 100% is omitted as it is a replica of test 1.

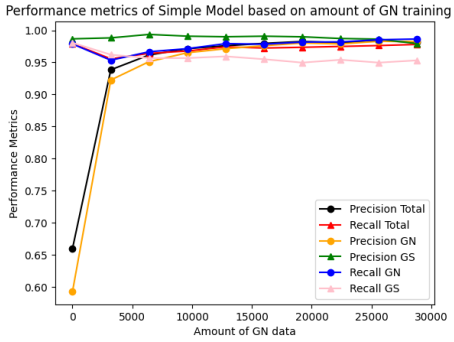
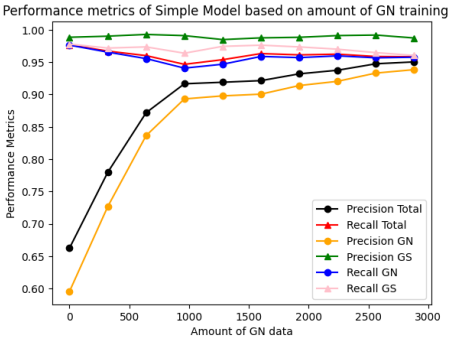
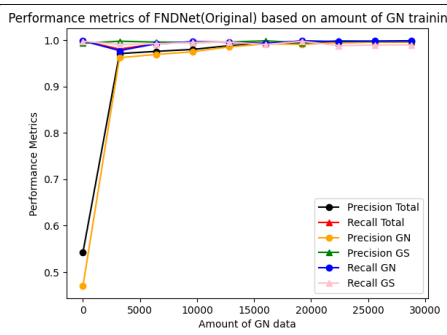
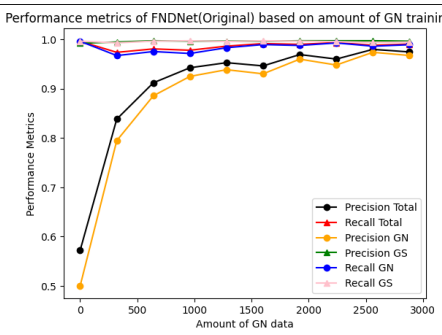
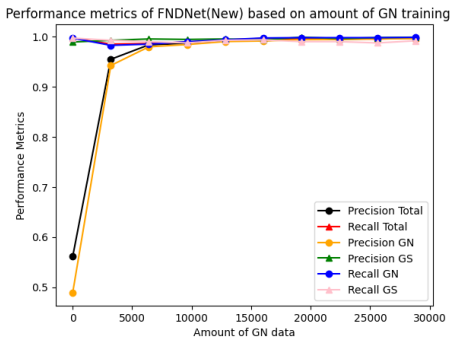
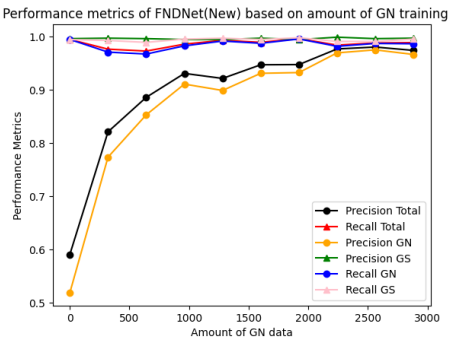
Model	0-90%	0-10%
Model B		
FNDNet		
FNDNet(new)		

Table 3.10 Performance Metrics of the Various Models on both the combined test set (Total) and individual test sets (GN/GS)

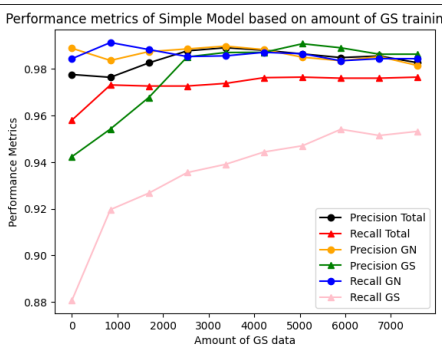
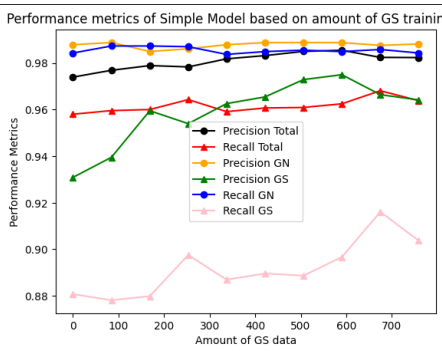
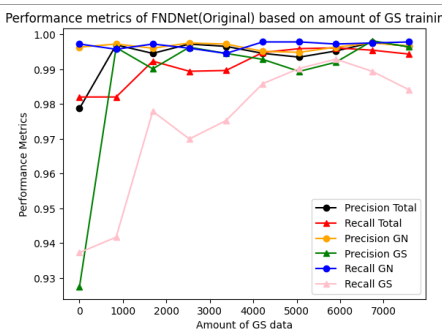
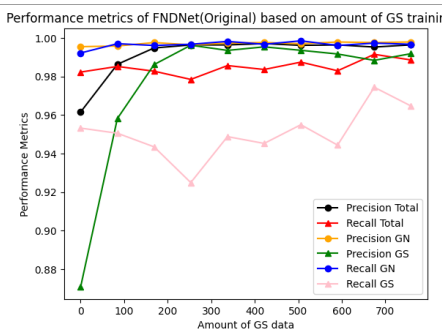
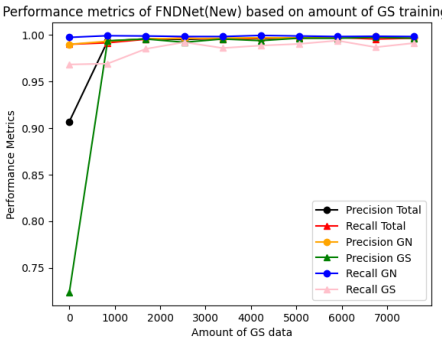
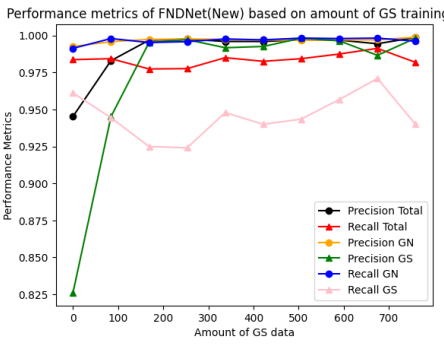
Model	0-90%	0-10%
Model B		
FNDNet		
FNDNet(new)		

Table 3.11 Performance Metrics of the Various Models on both the combined test set (Total) and individual test sets (GN/GS)

already more than satisfactory. As soon as a minimal amount of Global South data is added to the training set, the overall performance of the model comes close to the peak performance exhibited by the models.

Chapter 4

Conclusion

4.1 Conclusion

The results of the tests in the previous sections provide no conclusive proof that our hypothesis, i.e., current disinformation detection models show a bias towards (or) perform better on data originating from the Global North compared to the Global South, holds. The results point in the opposite direction, i.e., there is little to no bias shown by current disinformation detection models towards data originating from the Global North or the Global South.

The obtained results point in the following direction: disinformation detection systems when trained with an adequate amount of data originating from both the Global North and Global South, perform equally well on both datasets. When trained using data from only one of the two regions, the model fails to perform equally well on data originating from the other region. This difference is much more pronounced when the model is trained on only Global South data due to a lack of reliable data from these regions. Models trained on Global North data exhibit much better performance on Global South data, although still inferior to the performance of these models on Global North data.

The new wave of disinformation online is virtually indistinguishable from real news to humans, making it difficult to identify and combat disinformation. It is necessary to deploy

these systems in all major sources of information dissemination to combat the spread of lies, propaganda, and misleading agendas. While current disinformation detection methods appear to combat this problem satisfactorily, these models must be supplemented with other tools and techniques to boost their performance.

Knowledge graphs containing information about the author and the source of information, although difficult to implement, can be useful in determining the authenticity of the origin of the news. Public sentiment and discourse in the form of comments etc. on social media platforms can be used to gather information regarding public opinion on the news.[1] Another method that can be explored is the deployment of an LLM-based disinformation detection system. Fine-tuned LLMs have a vast amount of knowledge to supplement their decision, making them more future-proof than their counterparts.

Studies also talk about the linguistic and stylistic differences between real news and fake news. Fake news tends to be more polarizing and subjective, whereas real news is much more objective and typically is accompanied by verified facts and statistics. This information can be extracted by LLMs from texts owing to the extraordinary capability of processing natural languages.[10]

References

[1] K. Grover, S. M. P. Angara, M. S. Akhtar, and T. Chakraborty, "Public wisdom matters! discourse-aware hyperbolic fourier co-attention for social-text classification," 2022.

[2] D. Fu, Y. Ban, H. Tong, R. Maciejewski, and J. He, "Disco: Comprehensive and explainable disinformation detection," 2022.

[3] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using n-gram analysis and machine learning techniques," in *Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments: First International Conference, ISDDC 2017, Vancouver, BC, Canada, October 26-28, 2017, Proceedings 1*. Springer, 2017, pp. 127–138.

[4] ——, "Detecting opinion spams and fake news using text classification," *Security and Privacy*, vol. 1, no. 1, p. e9, 2018.

[5] R. Suharshala, A. Kadan, M. P.Gangan, and L. V L, "Online news popularity prediction before publication: effect of readability, emotion, psycholinguistics features," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, pp. 539–545, 06 2022.

[6] A. Dhawan, M. Bhalla, D. Arora, R. Kaushal, and P. Kumaraguru, "Fakenewsindia: A benchmark dataset of fake news incidents in india, collection methodology and impact assessment in social media," *Computer Communications*, vol. 185, pp. 130–141, 2022.

[7] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "Fndnet – a deep convolutional neural network for fake news detection," *Cognitive Systems Research*, vol. 61, pp. 32–44, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389041720300085

[8] L. Yuan, H. Jiang, H. Shen, L. Shi, and N. Cheng, "Sustainable development of information dissemination: A review of current fake news detection research and practice," *Systems*, vol. 11, no. 9, p. 458, 2023.

[9] Z. Khanam, B. Alwasel, H. Sirafi, and M. Rashid, "Fake news detection using machine learning approaches," in *IOP conference series: materials science and engineering*, vol. 1099, no. 1. IOP Publishing, 2021, p. 012040.

[10] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever *et al.*, "Improving language understanding by generative pre-training," 2018.

[11] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: A hybrid cnn-rnn based deep learning approach," *International Journal of Information Management Data Insights*, vol. 1, no. 1, p. 100007, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667096820300070