

Automated Text Classification

Group:

Name1 a1

Name1 a2

Name3 a3

**The University of Adelaide**

4533\_COMP\_SCI\_7417\_7717 Applied Natural Language Processing

Lecturer: Dr. Orvila Sarker

**Table of Contents**

[1. Abstract 2](#_Toc162337384)

[2. Introduction 2](#_Toc162337385)

[2.1. Data Collection Process 2](#_Toc162337386)

[2.2. Xxxxx 3](#_Toc162337387)

[3. Preprocessing 3](#_Toc162337388)

[3.1. Xxxx 3](#_Toc162337389)

[3.2. Mxxx 3](#_Toc162337390)

[3.2.1. Dxxxx 3](#_Toc162337391)

[4. Graphical Representation of the Data 3](#_Toc162337392)

[4.1. Xxxxxx 3](#_Toc162337393)

[4.2. System Architecture Components 4](#_Toc162337394)

[5. Data Categorisation 4](#_Toc162337395)

[5.1. Definition of the Categories 4](#_Toc162337396)

[5.2. Discussion: xxx 5](#_Toc162337397)

[7. Conclusion 6](#_Toc162337399)

[8. References 7](#_Toc162337400)

# Abstract

COVID-19 pandemic research provides a wide range of scientific studies to discover possible insights to fight against this infectious disease. However, manually searching the documents and writing the paper's abstract is time-consuming. Therefore, we devised an Information retrieval-based question-answering system and presented a summary of relevant information to the user.

…..

# Introduction

COVID-19 was urgent for all researchers worldwide when the number of cases frequently increased. All researchers worldwide were prompted on their research for prevention methods, new drugs and developing new vaccines. At such an alarming time, conducting research is very valuable to help the research and medical community by extracting useful information from thousands of articles (Afsharizadeh 2020, p. 237). However, timing consumption also matters, and automatically generated summarisation will allow users to find important information from different texts and articles and gain knowledge quickly (Cai et al. 2022, p. 1).

1. **Automated Text Classification**

In the automated text classification….

## Xxxxx

Xxxxx

**Extractive Summarisation:** The extractive approach involves picking up the documents' most essential phrases and lines. It then combines all the critical lines to create the summary. Therefore, the final summary consists of only sentences and phrases from the original text.

**Abstractive Summarisation:** Abstractive summarization algorithms use parts of the original text to get its essential information and create shortened versions, which might include new words rather than just the original words.

# Preprocessing

## Xxxx

CORD-19 is a free resource …. .

## Mxxx

### Dxxxx

In the dataset preprocessing, we ….

# Data Visualisation

## Xxxxxx

This study …. .

**Figure 1.** Workflow of retrieving NLP related posts and issues from Stack Overflow.

## Data Categorisation

Xxxxx

**Figure 2.** Pre-processing steps of Stack Overflow documents

**Figure 3.** Tags vs the number of posts per tag

**Table 1:**

|  |  |  |
| --- | --- | --- |
| Tags | Total Posts | … |
| .. | … | … |

# Implementation Details

## Xxxxx

**Table 1.** Performance metrics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

From Table 1. we can see that:

* Xxx .
* Yyy .

## Discussion: xxx

**Zzz :** In ….

**Ddd :** In … .

This … .

# Conclusion

This … .

# References

1. Afsharizadeh M., Ebrahimpour-Komleh H., Bagheri A 2020. Automatic text summarization of COVID-19 research articles using recurrent neural networks and coreference resolution. Frontiers in Biomedical Technologies 7, 236–248. doi:10.18502/fbt.v7i4.5321
2. Askdata. (2021). Train T5 for Text Summarization. Available at: <https://medium.com/askdata/train-t5-for-text-summarization-a1926f52d281>
3. Beltagy I, Lo K & Cohan A 2019, ‘SciBERT: A Pretrained Language Model for Scientific Text’, arXiv.org.
4. Cai X, Liu S, Yang L, Lu Y, Zhao J, She, D & Liu T 2022, ‘COVIDSum: A linguistically enriched SciBERT-based summarization model for COVID-19 scientific papers’, Journal of Biomedical Informatics, vol. 127, pp. 103999–103999.
5. Gasic, M. (2019). Text Summarization Part 2 - State of the Art. Besedo Engineering. Available at: <https://medium.com/besedo-engineering/text-summarization-part-2-state-of-the-art-ae900e2ac55f>
6. Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, … Poon H 2022, ‘Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing’, ACM Transactions on Computing for Healthcare, vol. 3, no. 1, pp. 1–23.
7. Gupta, A. (2021). Simple Abstractive Text Summarization with Pretrained T5 Text-to-Text Transfer Transformer. Towards Data Science. Available at:

[https://towardsdatascience.com/simple-abstractive-text-summarization-with-pretrained-t5-text-to-text-tran](https://towardsdatascience.com/simple-abstractive-text-summarization-with-pretrained-t5-text-to-text-transfer-transformer-10f6d602c426#%3A%7E%3Atext%3DT5%20is%20an%20abstractive%20summarization%2Cdirectly%20from%20the%20original%20text) [sfer-transformer-10f6d602c426#:%7E:text=T5%20is%20an%20abstractive%20summarization,directly%](https://towardsdatascience.com/simple-abstractive-text-summarization-with-pretrained-t5-text-to-text-transfer-transformer-10f6d602c426#%3A%7E%3Atext%3DT5%20is%20an%20abstractive%20summarization%2Cdirectly%20from%20the%20original%20text) [20from%20the%20original%20text](https://towardsdatascience.com/simple-abstractive-text-summarization-with-pretrained-t5-text-to-text-transfer-transformer-10f6d602c426#%3A%7E%3Atext%3DT5%20is%20an%20abstractive%20summarization%2Cdirectly%20from%20the%20original%20text)

1. Jurafsky D, Martin J.H, 2023, ‘Speech and Language Processing’. Available at: https://web.stanford.edu/~jurafsky/slp3/ed3book\_jan72023.pdf
2. T5 for Text Summarization in 7 Lines of Code. Artificialis. Available at: <https://medium.com/artificialis/t5-for-text-summarization-in-7-lines-of-code-b665c9e40771>