****

**Maximising privacy, anonymity and reduction of digital footprint over the internet**

**By**

**Put your name here**

**Project unit: Put Your Project Unit Code Here**

**Supervisor: Put Your Supervisor's Name Here**

**Month Year**

**Acknowledgements – on a new page**

on a new page and centred

50 to 100 words

should not appear in contents list

can be coloured if you wish

**Abstract – on a new page**

on a new page and centred

about 100 words

should not appear in contents list

can be coloured if you wish

**Table of Contents – on a new page**

**Table of Figures – on a new page**

# Chapter 1: Introduction – on a new page

# (remember – past tense)

## A heading based on your project e.g. Formula 1 Racing  - ALL headings should be the same size

**ALL** main text should be font size 11 or 12

This section is basically an introduction - heavily based on your ‘background’ section of your PID

## Project aim and objectives

## Project constraints

## Log of risks – useful to discuss in your conclusions

## Project deliverables

## Project Approach

### Project Management

A description of the project management methodology you have chosen to use to manage your project. You must state why you chose this particular methodology (or made up one of your own), how it will be applied to your specific project and if there are any parts of the methodology that you won’t be implementing

### Research

Give a BRIEF overview of what research you will be doing i.e. secondary research (your literature review). You might want to use some of the content from your PID, approach section

## Legal, ethical, professional, social issues (mandatory)

## The report

A summary of what the rest of the chapters are about

# Chapter 2: Literature Reviews

## Introduction

In this chapter begins with a description of the fundamental research strategy needed to comprehend digital footprints and their origins. To begin, many sorts of digital footprints are addressed. Following that, a number of popular digital footprint instances are discussed, also highlight the gap analysis of previous research. Moreover, the concerns concerning. The availability of private as well as public digital footprint records is examined.

The words digital footprints are used alternately in this text and in the rest of the literature. Digital footprints are another term for digital foot printing. A digital footprint is a type of footprint that occurs when digital footprint information is produced while a person is surfing the Internet. It comprises contact data, surfing the past, and information shared by the user with online services such as online stores. Digital footprints are classified into two types: either passive or active.

## Research Strategy

(Cheng, 2018)The confidentiality of digital footprint content, and has enormous business potential, is said to be a moderately human right. Additionally, the type of private information involved determines when to utilise a DNT opt-out or default technique. The practical implications point to a technical compromise amongst digital footprint information privacy and traditional advertising usage. Because this crucial issue is fairly young in application sectors, and no basic academic research has created a complete theoretical legal foundation, the authors are among the first to do so. Apart from its originality, this study contributes to the area by providing a theoretically viable approach for both digital footprint privacy protection and marketing advantages.

(Tu, 2022)By merging a digital footprint dataset and multiple geographical auxiliary datasets, the study employs an ensemble learning technique to predict population trends. To create early weekly and monthly gridded population estimates, they used a spatially scaled regression model that included two tree-based classifier model. These results were then corrected based on county-level estimations and their nonlinear connection with grid-level variables. The study created high-resolution demographic maps of China for 2017 which proved more realistic than earlier datasets after appropriate training and parameter adjustment. In addition, the analysis showed transport systems and population mobility trends in small and major cities.

(Rizi, 2022) Reviews existing safety and confidentiality measures and describes the research difficulties and concerns that remain unanswered. This research resulted in the creation of a comprehensive map of the scholarship on the issue, which includes key concepts, evidence, difficulties, solutions, & omissions. This results summary streamlines a large and varied set of research that was previously difficult to access.

(Micheli, 2018) The study's goal is to persuade current digital disparity frames to include this extra factor

(Cheng X. B., 2022) The ubiquitous sensing of both individuals as well as the surroundings is the most important part in creating a digital city. Crowdsensing entails a large number of people participating to the collecting of data, which could also contain critical information such as personal names and whereabouts. Individuals, companies, and even nations can suffer tremendous harm if this knowledge is intercepted or exposed. Users may be unwilling to disclose their data as a result of this danger to their privacy. To address this issue, a variety of communication techniques must be used to spread sensing jobs and gather data from regular users' mobile devices. It is critical to provide privacy security in crowdsensing in order to complete difficult tasks. This essay examines the present state of crowdsensing in digital cities, with a special emphasis on privacy.

The research review shows how diverse groups may gain or suffer as a result of their digital footprints. Those on the outskirts, such as users from deprived neighbourhoods, should be given special attention. We conducted further research in the following classifications.

### 2.1.1 Breach of Sensitive data stored on the internet

Information leakage may not seem harmful at first, but it could be used by malicious attackers to link data from different sources and connect it to a single user. Recent studies on pseudonyms used on social media platforms have shown that even in the best-case scenario, attackers can still locate more than 60% of a user's Facebook account. This means that leaked information can be used to create a more comprehensive picture of a user's online activities, potentially compromising their privacy and security. It is essential to be cautious about sharing personal information online, even seemingly harmless data, as it can be exploited by attackers to target individuals and gain access to their sensitive information. As technology advances, it is important to stay up-to-date with the latest security measures and take steps to protect personal data to avoid falling victim to cyberattacks.

The paper (Jain, 2021) also presents statistics on OSN attacks and explores different defensive approaches to OSN security. Lastly, the survey identifies the remaining challenges and recommends guidelines to ensure trustworthiness in online social networks.

Not all personal information, however, is equally important or distinctive. (Teng, 2022) By analyzing the discoverability of several forms of personal information, the research offers a technique for detecting the content, size, and influence of an individual's digital footprint.

(Quan‐Haase, 2020) The study examines comprehensive conversations with forty older persons, to learn about their internet confidentiality concerns and the techniques they employ to address them. They discovered that seniors are most worried about protection and structural privacy, with only a minor worry about societal privacy. The most serious concerns were material abuse by third parties and unauthorized access to personal information. Authors discovered that some older persons' significant privacy concerns prevented them from fully using the prospective benefits of online communication. The usage of privacy protection tactics varied greatly; some older individuals utilized no measures, while others were keen to safeguard their privacy by employing all available methods.

### Online behavior leaves digital footprints

Humans are increasingly users in a huge array of diverse online places in an Internet-enabled society, and the behaviors granted to these areas are growing increasingly complicated. There is an assumption in the study of computer-mediated communication (CMC) that the conduct seen in CMC is identical to that reflected in the conveyed text (Kaye, 2022).

Many studies do not consider activities such as liking, favoriting, following, or commenting on social media platforms as online content production, but they do contribute to a person's digital footprint. Sensitive data such as browsing history, web searches, purchase habits, and geographic location also contribute to an individual's digital footprint, even if the user is not aware of it. The algorithms used by social media platforms play a significant role in generating this data, not only by prompting users to fill out profiles and forms but also by creating digital footprints from every online action (Valanarasu, 2021). Social media firms value this data, as it can be used for behavioral predictions, monitoring, and advertising. Platforms and third-party service providers analyze and interpret this data to identify the meaning behind it. Digital footprints can offer an accurate picture of an individual if they are available and analyzed using appropriate technologies. For example, Facebook Likes can be used to predict personal variables such as behavioral tendencies, ethnicity, sexual orientation, religious and political beliefs, personality traits, or drug use with high accuracy.

This (Surmelioglu, 2019)study aims to investigate the level of awareness and experiences of digital footprints among higher education students. To gather data, the researchers created a survey tool called "A Survey for Digital Media Use" and administered it to 508 students from 41 Turkish universities via an online form. Participation was voluntary. The results reveal that the students had a high level of digital footprint awareness but a low level of digital footprint experience.

The aim of this (Agarwal, 2020) literature review is to explore the recent developments in FinTech and its impact on both banks and consumers. Authors examine the literature in two key areas: credit supply and payment and clearing services. The emergence of FinTech has brought about a digital transformation of traditional banking models, leading to significant changes in the delivery of financial services. Additionally, they highlight several areas for future research that may be of interest to policy-makers and researchers.

### 2.3 Personal Information Types and Classes

"Appropriately categorizing information is the basis for establishing formation control plan and sharing information," according to research at Beijing University (Shi, 2007).

This type of personal information is according to its importance and sensitivity (Norberg, 2007). Value is established by factors such as reliability.

### 2.4 The Price of Personal Data

A recent study claims that online privacy practises are not incompatible, but rather the outcome of differing privacy views. (Gencoglu, 2015).

This (Dai, 2020) The study looks on the effect of borrowers' digital traces on debt collection. The researchers discovered that information acquired from clients' online activities can boost the likelihood of payback for late loans by evaluating a sampling of residential loans given by a Chinese fintech lender. This effect can be ascribed to two factors: linking the borrower's responsibility to their social network and pinpointing their physical location. Therefore, lenders with data sets are more likely to be approved for loans, despite their higher risk of delinquency. These findings imply that digital footprints can be used as a new type of enforcement instrument and help to increase financial inclusion.

(Orlova, 2021) Using cluster analysis and the k-means approach, we offer a novel model for grouping borrowers based on creditworthiness and credit risks. In addition, we create a classification model utilising the SGB approach to reliably predict the cluster frequency and potential severity for new borrowers. These models can help make educated judgements regarding lending price, interest rates, and loan conditions for each homogenous group of applicants with identical risk levels. Additionally, they provide an improved approach for assessing individual creditworthiness, with the goal of lowering credit risks and improving credit institutions stability as well as profitability.

## Gap Analysis

(Cinnamon, 2020) Developed the concept of a "digital footprint gap" in the context of digital disparities throughout the life course, referring to data shared by adults regarding (unaware) children. The phrase "digital footprint gap" refers to discrepancies in the amount of online traces between people or demographic groups in this method. We propose to broaden this term to include not only the amount but also the quality of online traces, as well as the consequences of digital traces for repeating inequality. As a result, (Micheli, Lutz, & Büchi, 2018) ask: How do various social groups differ in terms of their digital footprints, and in what way do these qualitatively as well as quantitatively differing digital footprints result in results that influence social disparities?

The District Court in the Hague halted the online welfare fraud detection device known as SyRI on human rights groun, and the UK Court of Hearings found in 2021 that the computerised process for assessing the credit transactions had interpreted the income-based regulation (R[oao Johnson] v SSWP, 2020). In other words, in the short period that e-government services have grown, this change has made digitisation essential, exposing disparities in accessibility to and implementation of the services in question.

Table 1Previous Research finding and weakness

|  |  |  |  |
| --- | --- | --- | --- |
| Paper | Year | Findings | Weaknesses |
| (Gao, 2019) | 2019 | surveys privacy-preserving techniques for social media data publishing. | does not provide a comprehensive analysis of the strengths and weaknesses of each technique, and some of the techniques may not be practical for large-scale social media datasets. |
| (M. Xiong, 2022) | 2022 | examines cloud computing's safe & private information data-sharing methodologies | does not provide a detailed analysis of the practicality or scalability of the techniques for real-world cloud computing environments |
| (H. Guo, 2019) | 2019 | proposes a privacy-preserving platform for personalized health analytics using techniques such as differential privacy. | does not give a thorough examination of the user privacy and hazards connected with individualized health care. |
| (N. Mohammed, 2020) | 2020 | an overview of privacy-preserving ML techniques, such as federated learning | does not provide a detailed analysis of their practicality or scalability for real-world applications. |
| (Directions", 2020) | 2020 | surveys privacy-preserving DL techniques | does not provide a comprehensive analysis of the strengths and weaknesses of each technique, and some of the techniques may not be practical for large-scale deep learning models. |
| (System, 2020) | 2020 | The study offers a cryptocurrency electronic health record system using privacy-preserving mechanisms also including encryption techniques and safe multi-party computing. | The research does not go into depth on the privacy issues and risks that are associated with electronic medical record systems, or about the suggested system's possible influence on healthcare results. |
| (Shokri, 2021) | 2021 | offers a thorough examination of privacy-preserving machine learning approaches such as different datasets, distributed learning, as secure multi-party computing. | The paper does not provide a detailed analysis of the practicality or scalability of the techniques for real-world applications, and some of the techniques may require significant computational resources. |
| (Z. Yang, 2020) | 2021 | surveys privacy-preserving biometric authentication techniques | The paper does not provide a comprehensive analysis of the strengths and weaknesses of each technique, and some of the techniques may not be practical for real-world biometric authentication systems. |
| (X. Zeng, 2021) | 2021 | The paper proposes a privacy-preserving deep learning framework for financial risk management using techniques such as federated learning and differential privacy. | The paper does not provide a detailed analysis of the privacy risks and threats |

## Conclusion

In chapter 2 underscores the significance of safeguarding privacy, and anonymity, and minimizing one's digital footprint over the internet. The study establishes that the internet is a public domain, and unauthorized individuals or organizations can easily access and misuse personal information. Therefore, it is imperative for individuals to take proactive measures to protect their privacy and anonymity while online.

The review outlines several techniques and tools that can be employed to achieve maximum privacy and anonymity on the internet, such as Virtual Private Networks (VPNs), Tor browser, encryption, and strong passwords. It further emphasizes the importance of being cautious about the information shared online and its potential impact on personal security.

In conclusion, the literature review highlights the need for individuals to be responsible for their online privacy and security. By implementing the recommended measures and tools, people can minimize their digital footprint, and enjoy a secure and private online experience. It is vital for individuals to stay informed and updated about the latest trends and developments in online security to ensure the safety and security of their personal information.

# Chapter 3: Methodology

In this chapter, the approach/method for doing research on maximising privacy, anonymity, and minimising digital footprint over the internet.

## Introduction

This section presents an overview of the research methodology followed in this study. The section discusses the rationale behind choosing the qualitative or quantitative research methodology and the reasons for its appropriateness in addressing the research questions. It outlines the research design and the data collection methods used for the study.

## Research Strategy in chapter 1

To answer the study's research questions, one must devise a strategy (or select a method) for developing the study design. The research technique specifies the strategy for gathering enough data to answer the research questions. There are several research methodologies available:

* qualitative research
* quantitative research
* Mix research methods.

This study's research strategy is a combination of research. The exploratory research strategy is used to get a better knowledge of the many approaches and tools available for maximizing privacy, and anonymity, and minimizing one's digital footprinting on the internet. This comprises a study of existing research and an examination of various internet resources in order to determine the best methods and tools for reducing digital footprints.

The qualitative research method is used to assess the efficacy of employing a false news detection dataset to minimise digital traces while maximizing online anonymity and security. This entails gathering data from a variety of sources like Kaggle & open-source repositories on Github. The data is pre-processed, and feature extraction algorithms are utilised to detect distinguishing features of fake news pieces and distinguish them from actual news articles. Methods from machine learning are then utilised to build a model able to perform detecting bogus news items, as well as to evaluate the model's performance.

A quantitative study technique is also included in the research plan, with the goal of providing numerical statistics and information on the efficiency of deploying a false news detection dataset for minimising digital footprints. The application of machine learning algorithms allows for the analysis of massive amounts of data using statistical approaches to detect trends and patterns.

The research technique incorporates a case study approach in which the usefulness of applying a false news detection dataset on a sample group of persons is assessed. This entails gathering data on their digital footprints before and after using the strategies and tools indicated in the research. The case study results give real-world proof of the usefulness of utilising a false news detection dataset to reduce digital traces and improve online privacy and anonymity.

### Dataset:

The dataset for the fake news detection was obtained mainly from Kaggle and open-source repositories on Github. The dataset comprises both real and fake news articles. The sources of the data should be properly evaluated to avoid potential biases based on political or ideological affiliations. Additionally, it is important to ensure that the labeling of the data is accurate and consistent to ensure reliable results (Anon., 2019).

## Research Design

The answers to several of the research questions will have a significant impact on the approach. Machine learning algorithms, for example, are employed in development.

The flowchart below displays the proposed steps that will be taken over the course of this investigation.

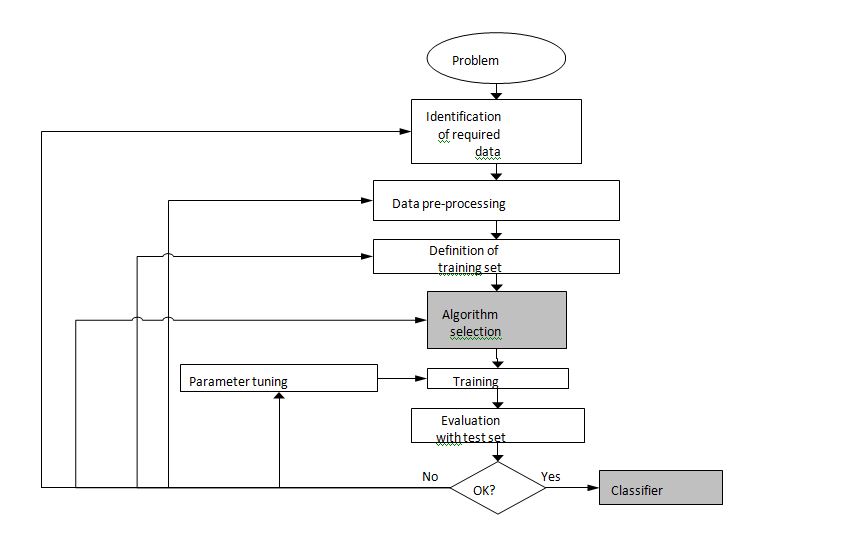


Figure 1 System design

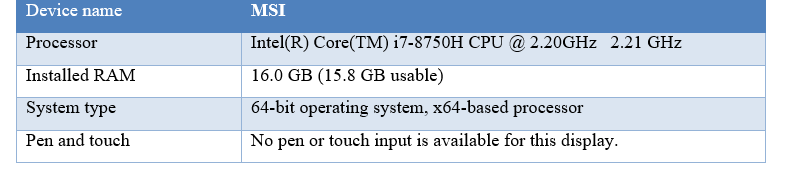
## Research Tool

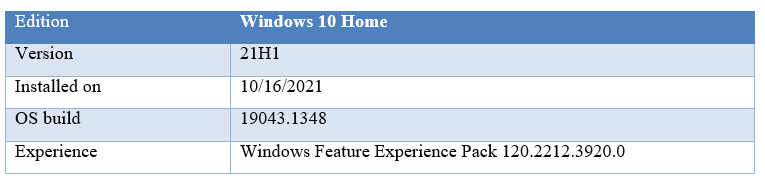
Jupyter Notebook with Python might be an excellent research tool for this system architecture. Python is a popular programming language for research technique and machine learning applications. Jupyter Notebook is a totally free and open-source web tool that allows users to create and share documents with live code, arithmetic, visualisation, and narrative prose.

We can use Python to develop software that acquire and evaluate data from fake news to put this system design into action. We may utilise the open source fake news dataset. Python used for coding.

Pandas and NumPy are data analysis and visualisation libraries. To chronicle the data collecting and analysis process, Jupyter Notebook may be used to produce a notebook containing code, data, and narrative prose. We may add narrative content to the notebook using Markdown cells to describe the goal of the study and offer context for the data. We can also utilise code cells to develop Python programmes that gather and analyse data from fake news and provide visualisations and reports depending on the results of the analysis.

### System Requirement:





## Research questions

Here are a few research questions on which we will concentrate our efforts in this study.

* In what way can a fake news detection dataset be used to reduce an individual's digital footprint while maximising online privacy and anonymity?
* Which strategies and technologies can be utilised to successfully eliminate a person's digital footprint and improve online privacy and anonymity?
* What is meant by the influence of someone's online behaviour on their digital footprint, and how may this be managed or minimised?
* In what ways can people be taught and informed about the dangers of digital footprints, as well as the value of online privacy and anonymity?
* What steps may be done to guarantee the privacy and security of sensitive information and data on the internet, and how successful are these steps in eliminating an individual's digital footprint?

## Conclusion

This chapter 3 presented a thorough summary of the research technique for investigating the efficacy of employing a false news detection dataset to reduce an individual's digital trace and maximise their online privacy and anonymity. The chapter described the many stages of the study design, such as data collection, data preprocessing, feature extraction, model selection, and assessment.

Overall, the research approach used in this work is comprehensive and appropriate for investigating the efficacy of employing a false news identification dataset to minimise an individual's digital trace and maximise online privacy and anonymity.

# Chapter 4: Data Analysis

In this chapter, we examine the specifications and put into effect our models for Data analysis.

## Introduction

We used a dataset for false news identification to meet our goal of maximising privacy, anonymity, and minimising our digital footprint on the internet. The dataset was gathered mostly from Kaggle and open-source sources on Github. The collection includes both true and false news stories.

In this chapter, they will analyse the dataset in order to clarify its properties and create models that can detect bogus news while minimizing the user's digital trace. To acquire insights into the dataset's qualities, we will employ a variety of data analysis approaches, including data visualization and statistical analysis. Furthermore, we will install different models and assess their performance in order to choose the optimal one for our needs.

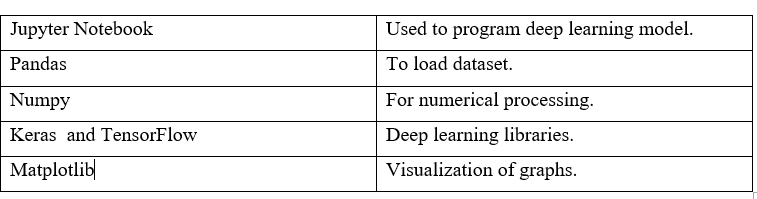
This chapter is crucial to attaining our study objectives. We can minimize the user's digital footprint and preserve their privacy and anonymity by analyzing the dataset and constructing effective models.

## Tools

This study extensively use the 64-bit Python 3.6 main programming languages.

The majority of the export extraction, pre-processing, and data preparation necessary to compile the imports to data was done in Python. Python is used for visualising and building the deep learning model. Matplotlib was used to display graphs, while Tensor Board was used to create the architectural graph. The deep learning model is built with Keras and has a TensorFlow backend.

Table 2 Tools used in this research



## Machine Learning Implementation

We use Jupyter Notebook and can generate a local machine by uploading the TensorFlow backend directly.

Jupyter Notebook makes it simple to create a local environment for utilising the TensorFlow backend. This entails directly downloading the required components, which may be achieved by running particular instructions within the laptop.

* ! Pip install tensorflow
* ! pip keras pip install

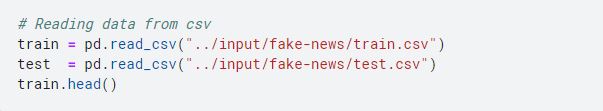
### Libraries:

I imported all the necessary libraries into Jupyter Notebook.



*Figure 2 libraries*

Next, I retrieved data from a CSV file.



*Figure 3 Call dataset*

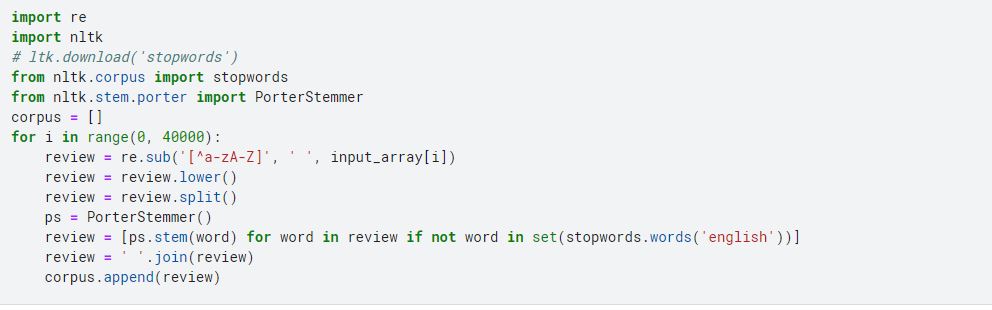
We combined these two datasets into one to make processing easier. To do this, I introduced a new column called "sentiment" that could distinguish between real and fake news. True news earned a sentiment value of one, while fake news received a value of zero.



*Figure 4 merge both csvs*

### Dataset cleaning

After merging the datasets, the next step was to clean the data. In my case, selected the first 40,000 titles as input. The specific number of titles can vary depending on your system's capabilities, but it must be at least 10,000. To clean the data, I removed symbols and stopwords and applied Porter stemming.

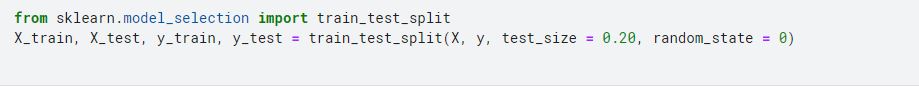


*Figure 5 cleaning of data*

To prepare the data for modeling, I initialized the independent variable, x, as the title and the dependent variable, y, as the sentiment (either 0 or 1). To limit the number of features, I chose to select the top 5,000 features, but this number can be adjusted based on personal preference. This step is crucial for reducing the complexity of the model and improving its performance.

### Data splitting

In this stage, we partition the dataset into two parts: training and testing. The goal of dividing the dataset is to test the model's performance on previously unknown data. The most frequent split ratio is 80/20 or 70/30, with the greater amount dedicated to training and the smaller portion dedicated to testing. In our situation, we utilised an 80/20 split ratio, which means that 80% of the data is used for training and 20% for testing.



*Figure 6 Data training*

### Models

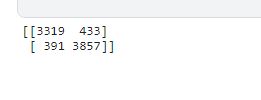
#### Naïve Bayes:

We used the NB class from the sklearn package to apply the Nave Bayes classifier. Because it implements the Nave Bayes method for multinomially distributed data, this class is especially built for text classification problems like the one we're working on.

The fit() function was used to fit the training data to the classifier, and the predict() method was used to forecast the false news (0 or 1) of the test data. We analysed the accuracies of numerous classification methods, including decision trees, support vector machines, and random forests, to find the best machine learning approach for our use case. The most accurate method was then picked as the final model for our investigation.



*Figure 7 NB implementation*



*Figure 8 NB output*

Based on the confusion matrix we obtained, we can calculate the accuracy of the Naïve Bayes classifier, which is 88.58%.

#### Logistic Regression:



*Figure 9 LR implementation*

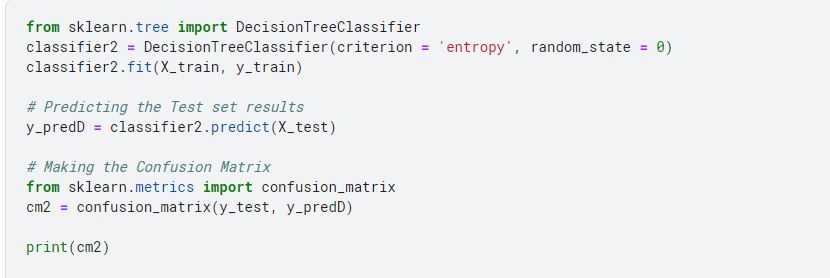
1.JPG

*Figure 10 LR confusion matrix*

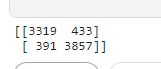
We applied the Logistic Regression algorithm to the training set and evaluated its performance by calculating the confusion matrix, which yielded an accuracy of 94.92%

#### Decision Tree:

The DT class from the sklearn package may be used to construct Decision Tree. Once the classifier has been trained, we can use it to predict the sentiment of the test data by fitting the training data to it.



*Figure 11 DT implementation*

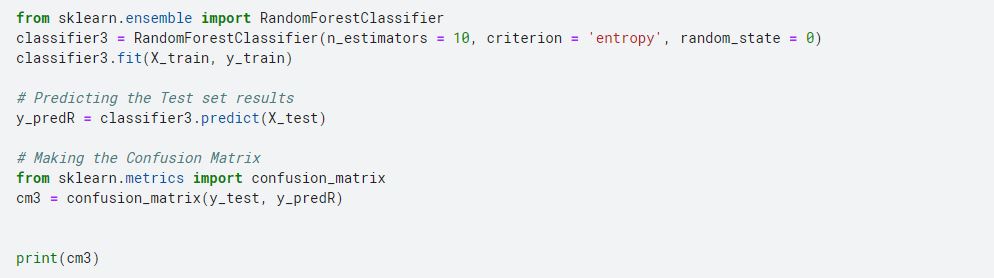


*Figure 12 DT output*

To implement Decision Tree Classification, we first fit the training set to the classifier using the fit() method from the sklearn library. Then we used the predict() method to predict the sentiment values of the test data then calculated the accuracy using the confusion matrix, which showed an accuracy of 89.70%.

#### Random Forest:

To implement the random forest algorithm, we use the RF class from the sklearn library. We fit the training data to the classifier and then use the predict() method to predict the fake (0 or 1) of the test data.



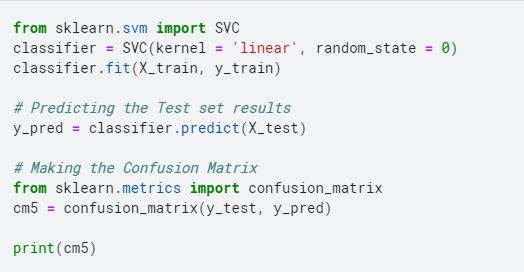
*Figure 13 RF implementation*

1.JPG

*Figure 14 Rf output*

After fitting the Random Forest Classification model to the training set, we evaluated its accuracy using the confusion matrix, which resulted in an accuracy of 92.37%

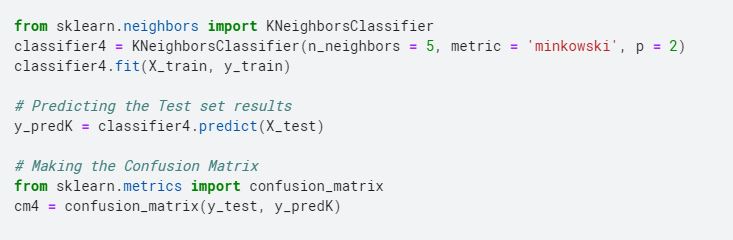
#### SVM:



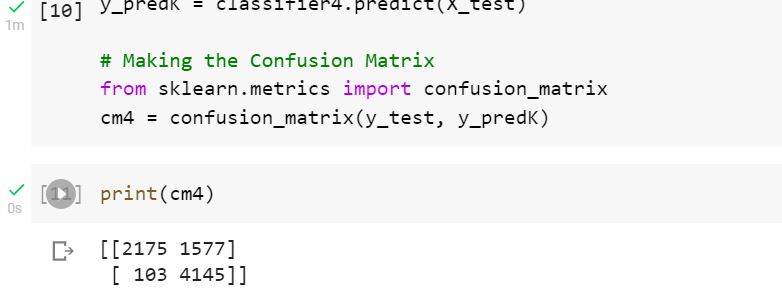
*Figure 15 SVM implementation*

Implementing SVM (Support Vector Machine) classifier and obtaining accuracy using confusion matrix, which is 94.92%

#### KNN:



*Figure 16 KNN*



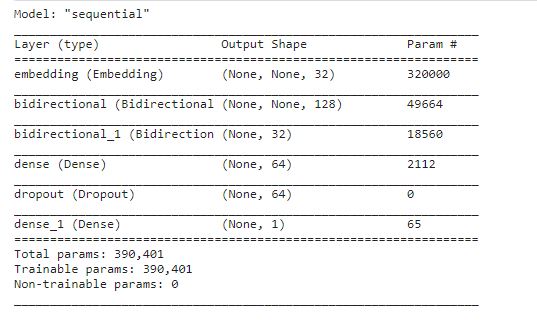
*Figure 17 KNN result*

To use K-NN, I fitted the algorithm to the training set with n\_neighbors set to 5 for quicker processing. The matrix of misinformation was used to calculate the accuracy, which was found to be 94.92%. It should be noted that altering the number of n\_neighbors might influence the model's accuracy.

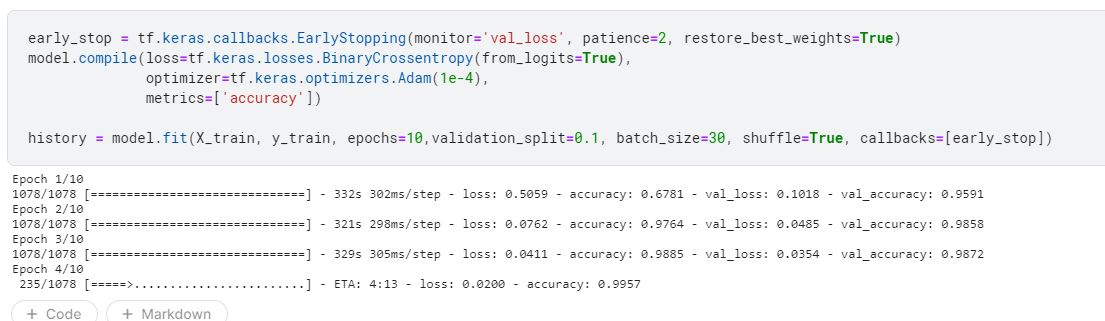
#### RNN model:



*Figure 18 RNN*



*Figure 19 model description*



*Figure 20 RNN result*

## Conclusion

In Chapter 4, we looked at multiple text categorization methods that use machine learning such as Nave Bayes, Logistic Regression, Decision Trees, Random Forest, SVM, and K-NN. We also used confusion matrices to test the accuracy of various algorithms on our dataset.

We obtained good accuracies with the majority of the algorithms, including Logistic Regression and K-NN. We also use Recurrent Neural Networks (RNNs) as a viable text categorization option. According to our case study and constraints, we may employ any of these algorithms, although RNN, Logistic Regression, and K-NN are the best possibilities for our present dataset. However, we may increase the accuracy of these algorithms even more by fine-tuning their parameters and employing more complex approaches such as ensemble learning and deep learning.

# Chapter 5: Result & Discussion

# Chapter 6: Conclusion– on a new page

## Recommendations

## Aim & Objectives (were they met)

## Risks (did any happen, was your mitigation ok?)

## Project Management (did it work? Was your Gantt Chart ok etc)

## Etc

## Conclusion

# References

(n.d.). Retrieved from https://www.kaggle.com/datasets?search=fake+news

Agarwal, S. a. (2020). FinTech, lending and payment innovation: A review. *Asia‐Pacific Journal of Financial Studies 49, no. 3 (*.

Ahn, J. K. (2020). "A survey of privacy protection techniques for smart home systems. *Journal of Ambient Intelligence and Humanized Computing, 11(8), 3551-3566.q*.

Cheng, F.-C. a. (2018). The do not track mechanism for digital footprint privacy protection in marketing applications. *Journal of Business Economics and Management 19, no. 2*.

Cheng, X. B. (2022). A Survey of Crowdsensing and Privacy Protection in Digital City. *IEEE Transactions on Computational Social Systems* .

Dai, L. J. (2020). Digital footprints as collateral for debt collection. *Available at SSRN 4135159* .

Directions", ".-P. D. (2020). Privacy-Preserving Deep Learning: Current Trends and Future Directions. *IEEE Signal Processing Magazine, vol. 37, no. 6, pp. 72-86, Nov. 2020, doi: 10.1109/MSP.2020.3004063.*

*fake and real news*. (2019). Retrieved from https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset

Gao, T. a. (2019). Privacy-preserving sketching for online social network data publication. *In 2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), pp. 1-9. IEEE, 2019.*

Gencoglu, O. H. (2015). Collecting a citizen's digital footprint for health data mining. *In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 7626-7629*.

H. Guo, F. Z. (2019). owards a Privacy-Preserving Platform for Personalized Health Analytics. *in Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Virtual Event, August 14-18*.

Jain, A. K. (2021). Online social networks security and privacy: comprehensive review and analysis. *Complex & Intelligent Systems 7, no. 5*.

Kumar, P. R. (2019). A privacy-preserving and secure framework using blockchain-based machine-learning for IoT-driven smart cities. *IEEE Transactions on Network Science and Engineering 8, no. 3* .

M. Xiong, Y. L. (2022). Secure and Privacy-Preserving Data Sharing in Cloud Computing: A Survey,. *IEEE Access, vol. 10, pp. 14550-14572*.

Micheli, M. C. (2018). Digital footprints: an emerging dimension of digital inequality. *Journal of Information, Communication and Ethics in Society*.

N. Mohammed, Y. Z. (2020). Privacy-Preserving Machine Learning: Threats and Solutions. *IEEE Communications Magazine, vol. 58, no. 10, pp. 90-96, October* .

Norberg, P. A. (2007). The Privacy Paradox: Personal Information . *Journal of consumer affairs 41, no. 1*.

Orlova, E. V. (2021). Methodology and models for individuals’ creditworthiness management using digital footprint data and machine learning methods. *Mathematics 9, no. 15*.

Rizi, M. H. (2022). "A systematic review of technologies and solutions to improve security and privacy protection of citizens in the smart city. *Internet of Things* .

Solanas, A. P. (2021). A comprehensive review of privacy-enhancing technologies for online social networks.

Surmelioglu, Y. a. (2019). An examination of digital footprint awareness and digital experiences of higher education students.

System", P.-P. B.-B. (2020). Privacy-Preserving Blockchain-Based Electronic Health Record System. *IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA), Antalya, Turkey, Nov. 2020, pp. 1-8, doi: 10.1109/AICCSA50817.2020.9360656.*

Teng, C.-C. a. (2022). A size and impact analysis of digital footprints.". *International Journal of Business and Systems Research 15, no. 2* .

Tu, W. Z. (2022). An ensemble method to generate high-resolution gridded population data for China from digital footprint and ancillary geospatial data. *International Journal of Applied Earth Observation and Geoinformation 107*.

Valanarasu, M. R. (2021). Comparative analysis for personality prediction by digital footprints in social media. *Journal of Information Technology 3, no. 02*.

X. Zeng, Y. C. (2021). Privacy-Preserving Deep Learning with Application to Financial Risk Management. *IEEE Transactions on Information Forensics and Security, vol. 16, pp. 2184-2197, Sep.*

Z. Yang, Z. H. (2020). Privacy-Preserving Machine Learning: A Comprehensive Survey. *IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 1, pp. 1-20, Jan. 2021, doi: 10.1109/TKDE.2020.30304222*.

# Appendices – on a new page

# Appendix A: Project Initiation Document

Copy & paste in the document

# Appendix B : Gantt Chart – on a new page

Your original one as attached to your PID

# Appendix C : Ethics form – on a new page

Copy & paste in the document

You might also have as appendices your research tool, answers etc