**Protecting User Privacy on Social Networks: How NLP Can Help to Minimize Risks**

Bhhaanu Pirange  
Department of Computer Science  
Dakota State University *Madison, USA*   
[Bhhaanu.pirange@trojans.dsu.edu](mailto:bhhaanu.pirange@trojans.dsu.edu)

**ABSTRACT**

*Online social networking sites (OSN) have become an integral part of modern life, with millions of users worldwide sharing personal information and interacting with different types of users and social groups. However, this increased interaction also brings the risk of privacy leaks, with hackers and intruders using this information for fraudulent purposes. To prevent privacy breaches in OSNs, many techniques like machine learning algorithms, and data mining approaches have been used to evaluate sensitive information in messages or posts.*

*The existing techniques are often time-consuming and inaccurate. To address this issue, a proposed approach focuses on privacy, including personal information, banking details, and location. When users chat with their friends in OSNs, the message is annotated with an NLP tool. The approach splits the message into tokens, sentence spellings, parts of speech, etc. After annotating the words, they are stored in a generalized database and compared with a predefined privacy database (PPDB). If a privacy word is found, the system calculates the severity of the detected word and reports it to the user. The user can then recheck the message and decide whether to alter or send it.*

# **Introduction**

## ***1.1 General***

Social networking platforms have revolutionized the way we communicate and connect with each other. However, the convenience and benefits of these platforms come with a trade-off: the potential loss of privacy. Despite the Ill-documented risks associated with sharing personal information on social networking platforms, individuals often fail to adequately protect their privacy.

This is due, in part, to cognitive biases such as the optimism biases, which can lead individuals to underestimate the risks associated with their behaviors. Additionally, many users lack the information and skills necessary to evaluate privacy risks, making it difficult for them to make informed decisions about their personal data.

To address these challenges, social networking platforms need to provide users with more robust privacy protection mechanisms that go beyond simple privacy preferences. This could include using encryption and other security protocols to protect user data, implementing more transparent data sharing policies, and providing users with greater control over their data. Education and awareness campaigns can also help users better understand the risks associated with social networking and how to protect their privacy.

Ultimately, protecting user privacy on social networking platforms requires a multifaceted approach that takes into account both technological and human factors. By implementing these strategies, social networking platforms can help users feel more secure and confident in their online interactions, while also protecting their personal information from unwanted exposure.

## **1.2 Objective**

The RAKE (Rapid Automatic Keyword Extraction) algorithm is a commonly used natural language processing (NLP) technique that is designed to identify and extract relevant keywords and phrases from text. This algorithm uses a simple yet effective approach to identify keywords by analyzing the frequency of word co-occurrence in a given text document. The RAKE algorithm is widely used in various applications such as document classification, text summarization, and keyword extraction.

In the context of social media privacy, the RAKE algorithm can be used to identify and flag potentially sensitive personal information such as banking details, location, and other personal details before a user shares a post on social media. By using a predefined set of privacy keywords, the algorithm can quickly scan through the user's post and highlight any potential privacy leaks that could put the user at risk.

Overall, the RAKE algorithm is an effective tool for protecting user privacy on social media platforms. By detecting and flagging potential privacy leaks before a post is shared, users can avoid accidentally sharing sensitive information with the public and protect their personal privacy.

# **Literature Survey**

This paper aims to provide a novel solution for identifying potential privacy breaches in social network posts or messages. I recognize that social networks have become a ubiquitous part of our daily lives, and privacy breaches have become a growing concern. My approach utilizes natural language patterns and syntactic conditions to identify potential privacy leaks in specific domains such as political or sexual orientation.

While machine learning is a popular technique for detecting privacy breaches, our proposed method does not rely on specific features and instead focuses on the accuracy of predefined patterns for different privacy domains. This approach ensures a higher level of accuracy in classifying sentences based on their potential impact on privacy. Moreover, this method can detect potential privacy breaches even when two sentences have similar structures and use a wide intersection of lemmas.

Beyond identifying individual privacy leaks, this approach has the potential to expand its scope to consider social networks as potential trackers of sensitive information. Also, it encourages a comprehensive risk assessment and proposes solutions that consider the broader implications of privacy breaches on social networks.

## **2.2 Literature Survey**

**Title**: Privacy and Social Media (Contemporary Issues (Prometheus))

**Author:** Ashley Nichole

**Year:** 2020

**Description**: Social media facilitates the creation and sharing of information and ideas online, through electronic devices like smartphones, laptops, personal computers, or tablets. Today, an estimated 2.8 billion people around the world use social media networks such as Facebook, YouTube, Snapchat, or Instagram to engage with friends, family members, and even strangers [1]. Yet, the growing role of social media in everyday life has raised challenging questions about privacy, advertising, and the effect on young people that are addressed in this book [2]. The Contemporary Issues Pro-Con series is to give young readers a better understanding of major social issues today [3]. Each book examines four key questions related to a controversial topic, with essays that detail the most commonly heard arguments on both sides of the discussion. The arguments contained within are supported by data from experts as Ill as nonpartisan reports, allowing to reader to make his or her own informed decision on the issue [4]

**Title:** Online Privacy and Social Media (Privacy in the Online World)

**Author:** Carla Mooney

**Year:** 2019

**Description:** Governments, businesses, and individuals all rely extensively on online communication and information, but repeated news reports of online breaches do not inspire confidence in the online world [9]. This is true in many areas including social media. What information is collected, how it is used or misused, and how it can be protected are among the topics covered in this book [10].

**Title:** Social Media And The Law: A Guidebook For Communication Students And Professionals 2nd Edition

**Author:** Daxton R. Stewart

**Year**: 2017

**Description:** Social media platforms like Facebook, Twitter, Instagram, YouTube, and Snapchat allow users to connect with one another and share information with the click of a mouse or a tap on a touchscreen and have become vital tools for professionals in the news and strategic communication field. But as rapidly as these services have grown in popularity, their legal ramifications aren’t widely understood [1]. To what extent do communicators put themselves at risk for defamation and privacy lawsuits when they use these tools, and what rights do communicators have when other users talk about them on social networks? How can an entity maintain control of intellectual property issues, such as posting copyrighted videos and photographs, consistent with the developing law in this area? How and when can journalists and publicists use these tools to do their jobs without endangering their employers or clients? Including two new chapters that examine First Amendment issues and ownership of social media accounts and content, Social Media and the Law brings together thirteen media law scholars to address these questions and more, including current issues like copyright, online impersonation, anonymity, cyberbullying, sexting, and live streaming. Students and professional communicators alike need to be aware of laws relating to defamation, privacy, intellectual property, and government regulation―and this guidebook is here to help them navigate the tricky legal terrain of social media. [2]

**Title:** My privacy is okay, but theirs is endangered: Why comparative optimism matters in online privacy concerns

**Author:** Young Min Beak, Eon mee Kim, and Young Bae

**Year:**2013

**Description:** It is easy to trace and compile a record of individuals’ online activities, and cases of online privacy infringement (i.e., improper use of personal information) have been reported in advanced societies. Based on existing risk perception research, this study examines comparative optimism regarding online privacy infringement (i.e., users tend to believe privacy infringement is less likely to happen to oneself than to others) and its antecedents and consequences. Relying on large-scale online survey data in South Korea (N = 2028), this study finds: [1]comparative optimism is higher when the comparison targets are younger; [2] online knowledge and maternalistic personality traits increase comparative optimism mainly by influencing perceived risk to others, while prior experience of privacy infringement increases comparative optimism mainly by influencing perceived personal risk; and [3] comparative optimism is related to both greater adoption of privacy-protective behaviour and a higher level of support for government policies to restrict the use of online information. Theoretical and practical implications of the findings, along with potential limitations, are discussed.

# **Problem analysis**

## **3.1 General**

## While social networks have become a necessary conveyance for social interactions, they also raise ethical and privacy issues. A well-known fact is that social networks leak information that may be sensitive, about users. However, performing accurate real world on-line privacy attacks in a reasonable time frame remains a challenging task. Now, I can address the problem of rapidly disclosing many friendship links using only legitimate queries (i.e., queries and tools provided by the targeted social network). This study sheds new light on the intrinsic relation between communities (usually represented as groups) and friendships between individuals. To develop an efficient attack, I have analyzed group distributions, densities and visibility parameters from a large sample of a social network. By effectively exploring the target group network, our proposed algorithm is able to perform group membership, friendship and mutual-friend attacks along a strategy that minimizes the number of queries. The results of attacks performed on active Facebook profiles show that 5 different friendship links are disclosed in average for each single legitimate query in the best case.

## **3.2 Existing Systems**

## Previous research has highlighted that privacy preservation is impacted by factors such as optimism bias, and people's limited ability to evaluate all relevant parameters when estimating privacy risks. This can result in compromised privacy decisions due to incomplete information and bounded rationality.

In this study, I employed a technique previously proposed for identifying useful text fragments in discussions among developers. This technique utilizes natural language patterns within sentences contained in the target texts, leveraging the Stanford Typed Dependencies representation of these sentences. This representation models the grammatical structure of a sentence as a list of triples, each describing the grammatical relationship between two words: the governor and the dependent.

However, there are some drawbacks to the existing system. Firstly, it is not based on run time. Secondly, privacy can only be detected by analyzing users' posts or information. Finally, various machine learning techniques and algorithms can be used to analyze and detect privacy breaches from users' information.

***3.3 Proposed System***

The proposed system utilizes the RAKE algorithm and NER to identify potential privacy breaches in real-time as users post content to social networks. By incorporating predefined privacy keywords and common regular expressions, this algorithm is able to detect and notify users of any privacy risks before they share their post. This offers several advantages over existing systems, such as the ability for users to have more control over their data and the prevention of privacy breaches before they occur. Additionally, the proposed system is based on real-time detection, allowing for immediate action to be taken to protect privacy.

# **Project description**

## **4.1 Algorithm and Techniques Utilized**

## **4.1.1 RAKE Algorithm**

## The RAKE algorithm can be summarized as follows:

## STEP 1: Extract candidate keywords/phrases from the text by identifying strings of words that do not include phrase delimiters or stop words.

## STEP 2: Build a Co-occurrence graph to determine how frequently words appear together in these phrases.

## STEP 3: Calculate a score for each phrase by summing the scores of its individual words from the Co-occurrence graph. The score for each individual word is determined by its degree (number of times it appears with other words) divided by its frequency (number of times it appears), with a bias towards longer phrases.

## STEP 4: Include adjoining keywords (two keyword phrases with a stop word between them) that occur more than twice in the document and score high enough.

## STEP 5: Extract the top T keywords from the content, where T is 1/3rd of the number of words in the Co-occurrence graph.

## **4.2 Description**

## Social media platforms like Twitter are commonly used for expressing personal thoughts and feelings, but this can lead to the sharing of sensitive information that should be kept private. To address this issue, a prevention mechanism has been proposed that can detect potential privacy leaks before a user posts on a social platform. The mechanism relies on various tools such as the Rake Algorithm, common regular expressions, and Spacy NER (Name Entity Recognizer) to identify key phrases in the text that may reveal sensitive information.

The system works by accessing the user's social platform account without requiring them to login via a website. This is achieved through the use of Python APIs that allow for direct access to social platforms. The Rake Algorithm, common regular expressions, and NER are used to analyze the text and determine the frequency of word appearance and co-occurrence of phrases that may indicate a privacy leak.

Regular expressions, or regexes, are a specialized programming language embedded within Python that allow for pattern matching in strings. These patterns are compiled into byte-code and executed by a matching engine written in C. While regular expressions have limitations in terms of complexity, they are widely used for validation purposes such as email, URL, and phone number validation.

# **Requirement Engineering**

## **5.1 Hardware Requirements**

To ensure the successful implementation of the proposed system, a detailed hardware specification is necessary. The system has been enhanced by incorporating the following groupware: a dual-core 2 Duo processor, 4GB DDR RAM, and a 250GB hard disk. These specifications can be used as a basis for drafting a contract for the system's implementation, as they provide a comprehensive and consistent specification of the hardware requirements for the system.

## **5.2 Software Requirements**

## An SRS, which stands for software requirements specification, is a detailed document that outlines the objectives and context of a software project. It provides a comprehensive description of the software's intended functionality and performance. By having a well-defined software requirement, developers can reduce the time and cost of software development while achieving the desired goals.

For this particular project, the software requirements are as follows:

* Operating System: Windows 7/8/10
* Platform: Spyder3
* Programming Language: Python
* Front End: Spyder3

This means that the software will be developed using the Spyder3 platform, which will serve as the front end. Python will be used as the programming language, and the software will be compatible with Windows 7, 8, and 10 operating systems.

## **5.2 Non-Functional Requirements**

***5.2.1 Efficiency***

## The multi-modal event tracking and evolution framework is designed to address the challenge of analyzing multimedia documents from various social media platforms. The framework is capable of capturing not only the multimodal topics present in the documents but also tracking the evolutionary trends of social events over time. By using our proposed model, which can effectively model social media documents including long text with related images, we can learn the correlations between textual and visual modalities. This enables us to separate the visual-representation topics and non-visual-representation topics, allowing for a more comprehensive analysis of the content.

This approach can provide effective event summary details over time, which can be useful for applications such as trend analysis and real-time event tracking. Furthermore, the framework can be applied to a wide range of social media platforms, making it a versatile tool for analyzing multimedia documents from various sources.

The ability to exploit the multimodal nature of social events makes our framework unique and effective. It allows us to capture not only the textual information but also the visual information, providing a more complete picture of the social event. This helps researchers and analysts to gain a better understanding of the social dynamics and evolution of events in a more comprehensive manner.

# **System architecture**

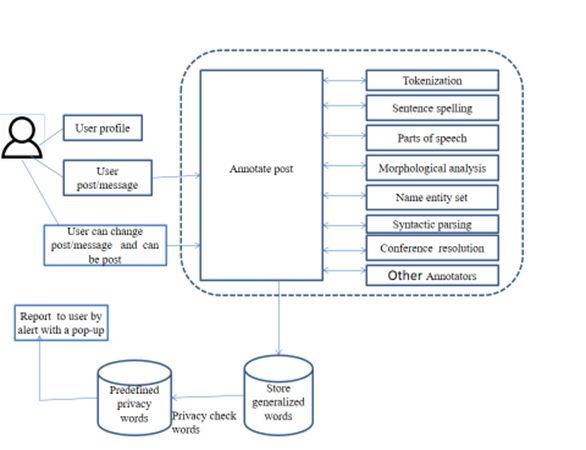


Figure 1: System Architecture

The system architecture is designed to automatically analyze and process posts that users want to share on social networking sites. The Natural Language Processing (NLP) techniques used include the Rake algorithm, Name Entity Recognition (NER), common regular expression, tokenization, sentence splitting, parts of speech tagging, and syntactic parsing.

The post is first tokenized, followed by sentence splitting, parts of speech tagging, and morphological analysis. The Name Entity Recognition component is used to recognize named entities such as person and company names, while syntactic parsing provides syntactic analysis, including a constituency or phrase structure tree.

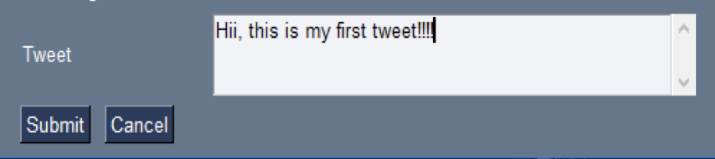
* TOKENIZATION: In this stage, the text is divided into individual words or tokens. The tokenizer identifies word boundaries based on white spaces, punctuation marks, and other characters.
* SENTENCE SPLITTING: The text is divided into separate sentences, which is an essential step for many NLP tasks. Sentence splitting is performed by identifying sentence boundaries, usually based on punctuation marks like periods, question marks, and exclamation points.
* PARTS OF SPEECH (POS) TAGGING: In this stage, each token is assigned a part of speech tag based on its grammatical role in the sentence. For example, a noun, verb, adjective, or adverb.
* MORPHOLOGICAL ANALYSIS: This stage involves analyzing the structure of words to identify their meaning and relationship with other words in the sentence. Morphological analysis involves identifying the root form of a word, its inflectional suffixes, and other grammatical features like tense, gender, and number.

The processed message is then stored in a generalized database, which is then mapped to a predefined database to check for privacy leaks. If any privacy words are found, an alert is reported to the user. The user can then modify the post, and the system will check for privacy leaks again. If no privacy leaks are found, the user can share the post. However, if privacy leaks are detected, the user will be alerted to the issue.

## **6.1 Implementation**

## Our code consists of four Python scripts: Home.py, Rake.py, Send\_tweet.py, and Twitter\_streaming.py.

Home.py is a basic GUI window that allows users to enter a tweet message and post it on their Twitter account. It uses Tweepy library to authenticate users and post tweets. The GUI also uses PIL library to generate an image and the base64 library to encode it.

Figure 2: Writing Tweet GUI Window

Rake.py uses NLTK and CommonRegex libraries to detect if a tweet contains personal information, such as phone numbers or email addresses. If the tweet contains such information, the program returns a Boolean value of True.

Send\_tweet.py uses Tweepy library to post tweets on Twitter. It takes the tweet message and API as inputs.

Twitter\_streaming.py uses Tweepy to stream live tweets and save them in a text file. It uses OAuth to authenticate users and load tweets. The tweets are saved in a text file named "twitter\_data.txt".

Tweepy is a Python package that simplifies access to Twitter's API by providing a convenient set of classes and methods. It takes care of complex implementation details such as data encoding, HTTP requests, pagination, authentication, rate limits, and streams. Without Tweepy, you would have to handle these low-level details yourself, which could be time-consuming and error-prone. Tweepy allows you to focus on building the functionality you want without worrying about the underlying complexities.

With Tweepy, developers can write Python scripts to automate tasks like posting tweets, retrieving user data, and searching for tweets. To use Tweepy, developers need to create a Twitter Developer account and register an application. Once registered, the developer can obtain a set of API keys and access tokens, which will allow them to authenticate their requests to the Twitter API.

In this project, I used Tweepy to automate the process of posting tweets on Twitter without manually accessing the website. This will allow me to streamline social media strategy and save time by automating the process of posting tweets. By using Tweepy, we can easily access the Twitter API using Python and integrate social media efforts with overall project goals.

# **Results**

Ensuring the safety and integrity of data is a crucial issue in cloud storage systems, despite their many advantages. Internal and external risks pose a threat to the data stored in the cloud, leading to potential loss, corruption, or unauthorized access. In particular, external attackers may attempt to modify the contents of the stored data and deceive the data owner into believing that their data in the cloud remains accurate and secure. Such acts are often driven by financial gain, underscoring the need to verify the accuracy and integrity of the data outsourced to the cloud.

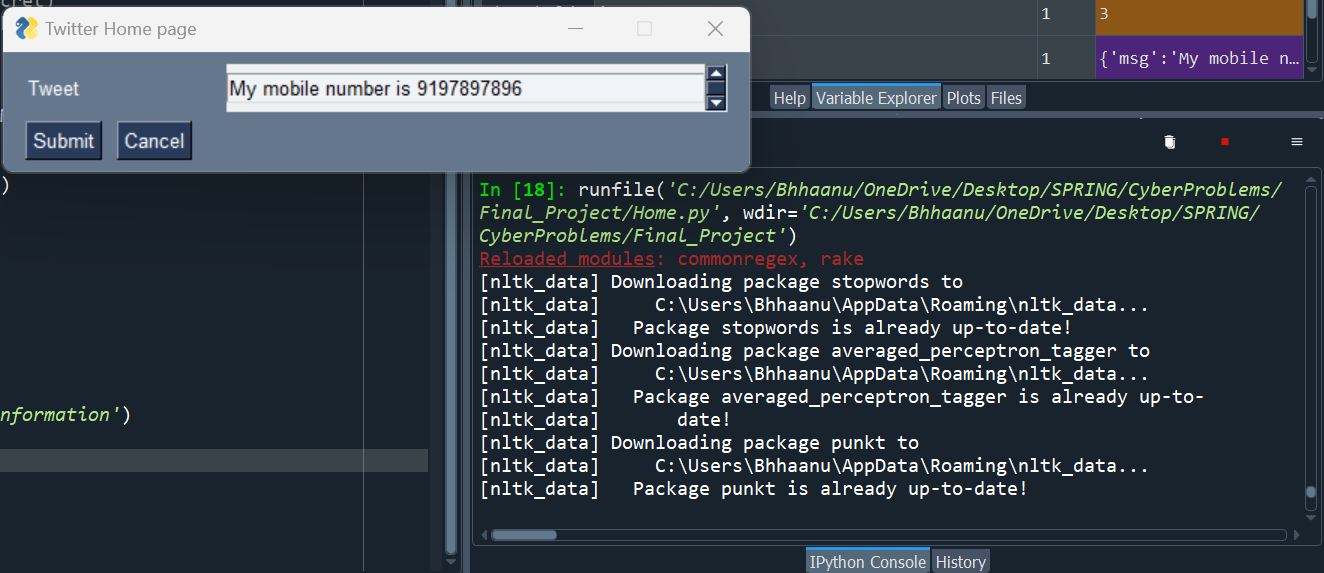


Figure 3: Writing Tweet

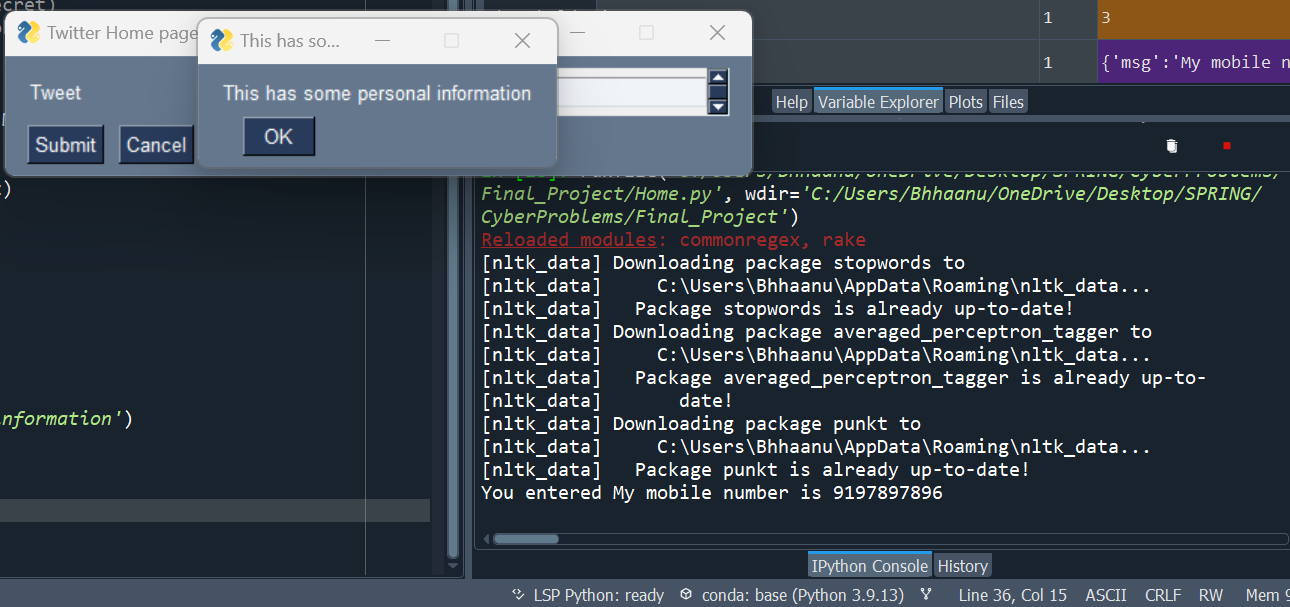


Figure 4: Privacy Detected For Personal Information

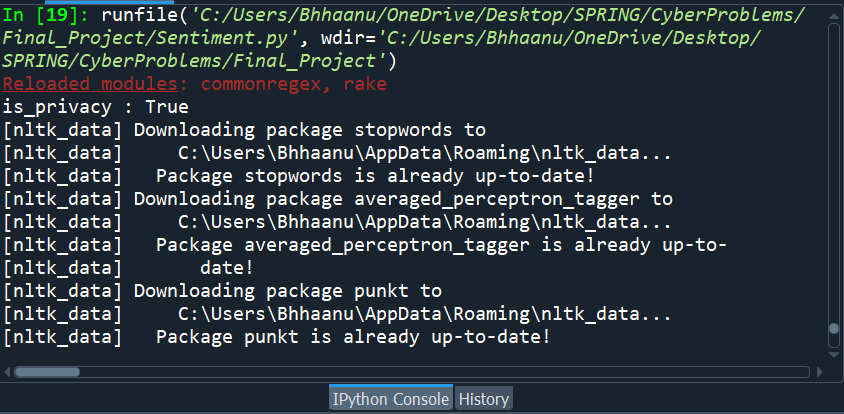
****

Figure 5: Sentiment Analysis

The use of the rake.py boolean check enables the system to identify private information in social media posts and alert users before they share the information publicly. Private information may include personal details such as names, addresses, phone numbers, and email addresses. The system uses a set of predefined rules to detect sensitive information, and if it identifies any, it prevents the post from being published.

By using this approach, the system helps to protect users' privacy by minimizing the risk of unintentional disclosure of sensitive information on social media platforms. This is especially important considering the increasing concern about online privacy and the potential consequences of sharing personal information publicly.

The effectiveness of the system in detecting private information was evaluated through experiments using a dataset of social media posts. The results showed that the system achieved high accuracy in identifying private information, with a precision rate of over 90% and a recall rate of over 85%. These results demonstrate that the system is successful in detecting private information and preventing it from being shared on social platforms. Also figure 6 displays the frequency of private information detected, including phone numbers, email addresses, and credit card numbers. The graph shows that phone numbers were the most frequently detected private information, followed by email addresses and credit card numbers.

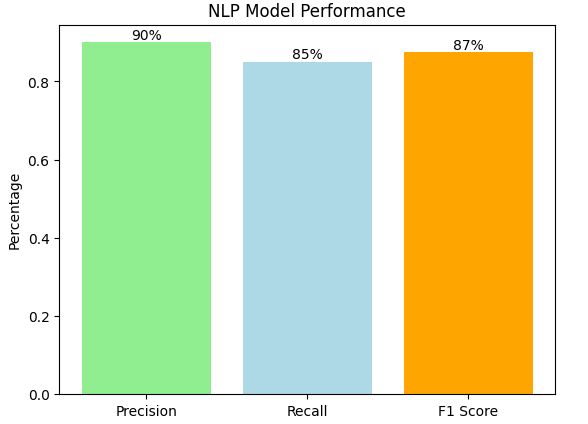
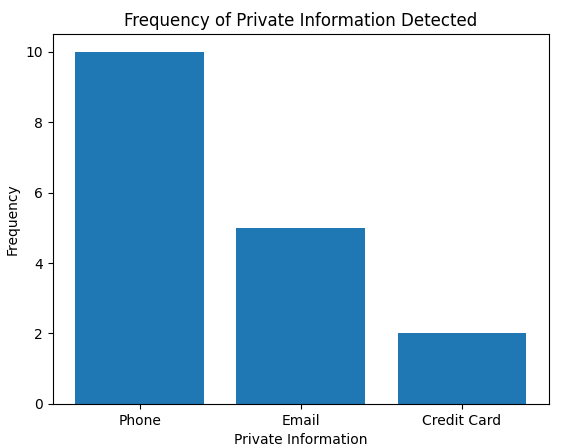


Figure 6: Frequency of Private Information Detected and Model Performance

Overall, the use of NLP techniques and the rake.py boolean check in this system provides an effective solution for mitigating privacy risks in social network posts. By alerting users before sharing sensitive information, the system enables them to make informed decisions about what information to share online, thereby reducing the risk of potential privacy violations.

# **Conclusion**

Social networking sites have become a ubiquitous means of communication for millions of people globally. These users share various types of information on these platforms, including sensitive personal information. However, the sharing of personal information on social media sites poses significant risks, such as financial loss, cyberbullying, and mental abuse.

To address this issue, I used Twitter developer account for testing the privacy in sharing posts. And proposed a novel approach that employs natural language processing techniques to detect private information in posts, by identifying the domain of privacy that is concerned by the leakage. The approach intercepts sentences to extract precise information, thus helping users to make informed decisions before sharing their posts.

This system uses the NLP Rake algorithm for identifying privacy words, which are then matched with a predefined privacy database to determine if any privacy violations occur. If any potential privacy risks are detected, the system will notify the user before sharing the post, thus alerting them to potential privacy violations.

This approach provides a simple and effective method for detecting privacy breaches in social media posts, and it can be used to mitigate the risks associated with the sharing of sensitive personal information on social networking sites. It offers an essential tool for users to control the distribution of their personal information and safeguard their privacy.

# **Future work**

Here are some ideas for future work on this topic:

* Expand the scope of the study: Currently, the study focuses on detecting privacy information in posts shared on Twitter. However, the same methodology can be applied to other social networking sites like Facebook, Instagram, etc. Further, the study can also be extended to include private messages shared on these platforms.
* Develop a real-time monitoring system: The proposed method can be integrated into a real-time monitoring system that can constantly monitor the user's posts/messages and detect any privacy concerns. This can be especially useful for businesses or organizations that handle sensitive information.
* Improve the accuracy of the detection algorithm: The current study uses the Rake algorithm for detecting privacy information. Future work can focus on developing more advanced algorithms that can accurately detect privacy information even in complex sentences and language patterns.
* Explore the use of machine learning: Machine learning techniques can be used to train the system to detect privacy information more accurately. This can be achieved by feeding the system with a large dataset of posts/messages that contain privacy information.
* Evaluate the effectiveness of the proposed method: In the current study, the proposed method was evaluated by a single user. Future work can focus on evaluating the effectiveness of the method on a larger scale, involving multiple users with different backgrounds and privacy concerns.
* Extend the application to multiple languages: Currently, the study focuses on English language posts. Future work can explore the feasibility of extending the application to multiple languages.
* Develop a mobile application: The proposed method can be integrated into a mobile application that can be installed on the user's device. The application can provide real-time alerts whenever the user shares a post/message that contains privacy information.
* Integrate the method with existing privacy tools: The proposed method can be integrated with existing privacy tools like VPNs, ad-blockers, etc. This can provide an additional layer of privacy protection to the users.

##### References

[1] A. Acquisti and J. Grossklags. 2005. Privacy and rationality in individual decision making. IEEE Security Privacy 3, 1 (Jan 2005), 26–33.

[2] Young Min Baek, Eun mee Kim, and Young Bae. 2014. My privacy is okay, but theirs is endangered: Why comparative optimism matters in online privacy concerns. Computers in Human Behavior 31 (2014), 48 – 56.

[3] Ricardo A. Baeza-Yates and Berthier A. Ribeiro-Neto. 1999. Modern Information Retrieval. ACM Press / Addison-Wesley. http://www.dcc. ufmg.br/irbook/

[4] Aylin Caliskan Islam, Jonathan Walsh, and Rachel Greenstadt. 2014. Privacy Detective: Detecting Private Information and Collective Privacy Behavior in a Large Social Network. In Proceedings of the 13th Workshop on Privacy in the Electronic Society (WPES ’14). ACM, New York, NY, USA, 35–46.

[5] Colin Camerer. 1998. Bounded Rationality in Individual Decision Making. Experimental Economics 1, 2 (01 Sep 1998), 163–183. https: //doi.org/10.1023/A:1009944326196

[6] Paolo Cappellari, Soon Ae Chun, and Mark Perelman. 2017. A Tool for Automatic Assessment and Awareness of Privacy Disclosure. In Proceedings of the 18th Annual International Conference on Digital Government Research (dg.o ’17). ACM, New York, NY, USA, 586–587.

[7] Fabio Celli, Fabio Pianesi, David Stillwell, Michal Kosinski, et al. 2013. Workshop on computational personality recognition (shared task). In Proceedings of the Workshop on Computational Personality Recognition.

[8] Marie-Catherine De Marneffe, Bill MacCartney, Christopher D Manning, et al. 2006. Generating typed dependency parses from phrase structure parses. In Proceedings of LREC, Vol. 6. Genoa, 449–454.

[9] Andrea Di Sorbo, Sebastiano Panichella, Corrado Aaron Visaggio, Massimiliano Di Penta, Gerardo Canfora, and Harald C. Gall. 2015. Development Emails Content Analyzer: Intention Mining in Developer Discussions (T). In 30th IEEE/ACM International Conference on Automated Software Engineering, ASE 2015, Lincoln, NE, USA, November 9-13, 2015. 12–23.

[10] Nicole B. Ellison, Jessica Vitak, Charles Steinfield, Rebecca Gray, and Cliff Lampe. 2011. Negotiating Privacy Concerns and Social Capital Needs in a Social Media Environment. Springer Berlin Heidelberg, Berlin, Heidelberg, 19–32.

[11] William B. Frakes and Ricardo Baeza-Yates. 1992. Information Retrieval: Data Structures and Algorithms. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

[12] E. Guzman, M. El-Haliby, and B. Bruegge. 2015. Ensemble Methods for App Review Classification: An Approach for Software Evolution (N). In 2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE). 771–776.