Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Instagram - Analysis of User Behavior

Bachelor Thesis

2024
Matej Zelenák

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Instagram - Analysis of User Behavior

Bachelor Thesis

Study Programme:	Applied Informatics
Field of Study:	Applied Informatics
Department:	Department of Applied Informatics
Supervisor:	doc. RNDr. Damas Gruska, PhD.

Bratislava, 2024

Matej Zelenák

33149107

Comenius University Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname: Matej Zelenák
Study programme: Applied Computer Science (Single degree study, bachelor I. deg., full time form)
Field of Study: Computer Science
Type of Thesis: Bachelor´s thesis
Language of Thesis: English
Secondary language: Slovak

Title: Instagram - analysis of user behavior

Annotation: The work is devoted to the behavior of users of the social network Instagram with an emphasis on the presidential elections in the Slovak Republic in 2024. It examines the accounts of individual candidates and the posts or reactions to them, including the examination of sentiment.

Supervisor: doc. RNDr. Damas Gruska, PhD.
Department: FMFI.KAI - Department of Applied Informatics
Head of department: doc. RNDr. Tatiana Jajcayová, PhD.

Assigned: 29.05.2024

Approved: 29.05.2024 doc. RNDr. Damas Gruska, PhD.
Guarantor of Study Programme

....
Student Supervisor

33149107

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta:	Matej Zelenák
Študijný program:	aplikovaná informatika (Jednoodborové štúdium, bakalársky I. st., denná forma)
Študijný odbor:	informatika
Typ záverečnej práce:	bakalárska
Jazyk záverečnej práce:	anglický
Sekundárny jazyk:	slovenský

Názov: Instagram - analysis of user behavior
Instagram - analýza správania sa používateľov

Anotácia: Práca sa venuje správaniu sa používateľov sociálnej siete Instagram s dôrazom na prezidentské voľby v SR v roku 2024. Skúma účty jednotlivých kandidátov a príspevky či reakcie na ne vrátane skúmania sentimentu.

Vedúci:	doc. RNDr. Damas Gruska, PhD.
Katedra:	FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry:	doc. RNDr. Tatiana Jajcayová, PhD.
Dátum zadania:	29.05.2024

Dátum schválenia: 29.05.2024

doc. RNDr. Damas Gruska, PhD.
garant študijného programu

.... študent

.... vedúci práce

Abstract

In the era where digitalization pervades every aspect of society, social media platforms like Instagram have become essential to our communication and access to information. This thesis examines social media activity in relation to real election outcomes, focusing on user behavior on Instagram during the 2024 Slovak presidential election.

The study aims to analyse user interactions with political content on Instagram, employing data scraping techniques to gather relevant data points, and utilizing SlovakBERT for sentiment analysis. Our findings reveal patterns in user engagement, highlighting how individuals express their sentiments and interact with political content. Notable findings include the differential support for candidates, with Ivan Korčok receiving more positive engagement compared to Peter Pellegrini, as indicated by likes and sentiment analysis.

This thesis underlines the importance of sentiment analysis in understanding public opinion dynamics and provides a framework for monitoring political sentiment on Instagram. While the study focuses on Instagram, future research could expand to other social media platforms for a more in-depth analysis. Additionally, integrating advanced machine learning techniques could enhance sentiment classification and predictive capabilities, offering deeper insights into the complex landscape of social media and political communication.

Keywords: Instagram, sentiment analysis, elections

Abstrakt

V dobe, v ktorej digitalizácia preniká do všetkých aspektov spoločnosti, sa sociálne siete ako Instagram stali nevyhnutnými pre našu komunikáciu a prístup k informáciám. Táto práca skúma aktivitu sociálnych médií vo vzťahu k skutočným volebným výsledkom so zameraním na správanie používateľov na Instagrame počas prezidentských volieb v roku 2024.

Cieľom štúdie je analyzovať interakcie používateľov s politickým obsahom na Instagrame, využívať techniky zbierania údajov na zhromažďovanie relevantných dát a využívať SlovakBERT na analýzu sentimentu. Naše zistenia odhaľujú zapojenie používateľov a zdôrazňujú, ako jednotlivci vyjadrujú svoje pocity a interagujú s politickým obsahom. Medzi pozoruhodné zistenia patrí rozdielna podpora kandidátov, pričom Ivan Korčok získal pozitívnejšiu angažovanosť v porovnaní s Petrom Pellegrinim, ako naznačuje analýza lajkov a sentimentu.

Táto práca zdôrazňuje dôležitosť analýzy sentimentu pre pochopenie dynamiky verejnej mienky a poskytuje framework na monitorovanie politického sentimentu na Instagrame. Zatiaľ čo sa štúdia zameriava na Instagram, budúci výskum by sa mohol rozšíriť na ďalšie platformy sociálnych médií pre hlbšiu analýzu. Okrem toho by integrácia pokročilých techník strojového učenia mohla zlepšiť klasifikáciu sentimentu a prediktívnych schopností, ktoré ponúkajú hlbší pohľad do komplexného prostredia sociálnych médií a politickej komunikácie.

Kľúčové slová: Instagram, analýza sentimentu, voľby

Contents

List of Figures

List of Tables

Introduction

In the era where digitalization is taking part in every aspect of society, social media platforms like Instagram are slowly becoming inseparable part of our lives, shaping the way we communicate and stay updated on news. While the digital landscape offers both benefits and drawbacks, its impact on various aspects of life, including political campaigns, is undeniable. As social media continues to evolve and shape societal norms, understanding its influence becomes crucial for navigating the complexities of modern life.

Political campaigns, once primarily conducted through traditional mediums, have progressively moved to the social media. Here, candidates exploit the vast reach and immediacy offered by the platforms to connect with voters, spread their message, and influence public opinion. The potential of social media in shaping electoral outcomes suggests the importance of conducting a separate and impartial analysis of user behaviour on these platforms.

Observing how people behave on Instagram during political campaigns is essential for understanding the evolving dynamics of political communication. Inspecting user behaviour on Instagram can offer valuable insights into how individuals interact with political content, shape their opinions, and participate in the democratic process. By investigating these phenomena, this thesis aims to analyse how users interact on social media. asdex This thesis starts by exploring Instagram user behaviour within the context of political campaigns, aiming to provide an understanding of the dynamics. We begin the process by data acquisition, as we describe the criteria for selecting relevant data points and employ scraping techniques to extract information from Instagram. Given the constraints imposed by social media platforms and the time-intensive nature of scraping, storing the data continuously is a critical consideration. Thus, we establish a local storage system to store the gathered data, ensuring accessibility for subsequent analysis.

Central to our investigation is the analysis of user interactions on Instagram, which involves examining patterns of engagement such as likes and comments. By examining the networks of users who engage with political content, we seek to gain insights into their connections and preferences. Furthermore, we employ sentiment analysis using SlovakBERT, a state-of-the-art language model trained on Slovak texts, to discern the prevailing attitudes towards political candidates. This approach enables us to create a picture of user behaviour on Instagram during political campaigns.

Chapter 1

Theory and Tools

This chapter offers all theoretical topics in this thesis. It begins by examining social networks, focusing on Instagram and Facebook. However, Facebook is included only for comparison purposes, as attempts to collect data from it proved to be ineffective. Then explores how graph theory can help to visualise some types of relationships in social network. The chapter then delves into scraping methodologies, detailing the use of tools like the Facebook Scraper and Instaloader for data acquisition. Following this, it discusses sentiment analysis techniques, including the application of SlovakBERT in revealing sentiment patterns within social media content. Lastly, the chapter reviews related work, summarizing the current study and highlighting significant contributions in the field.

1.1 Social Networks

Social networks, as described by Wasserman [1] and Freeman [2], are complex patterns of social interactions that form among individuals, groups, organizations, or other entities. These connections are established through various interactions such as communication, friendship, collaboration, or resource exchange.

In essence, people use them to interact, share information, and collaborate. These networks include both strong ties, like close relationships with family and friends, and weak ties, like casual contacts or colleagues.

Social networks play a relevant role in shaping individual behaviour, enabling the flow of information and ideas, while creating social interactions within communities. These platforms create opportunities for social support, mobilizing resources, and collective efforts, simultaneously shaping individuals' access to resources, opportunities, and social connections.

Understanding the structure and dynamics of social networks is crucial to see how social systems function and evolve over time. By studying the patterns of connections between individuals and groups within a network, researchers can understand the mechanisms driving social interactions, the formation of social norms, and the emergence of collective behaviour.

Overall, social networks represent the complex web of relationships that underlies human social life, and studying them provides valuable insights into the dynamics of social behaviour, communication, and interaction.

1.1.1 Facebook

Established in 2004 by Mark Zuckerberg, Facebook [3, 4] is a social media platform connecting users worldwide. Users can actively participate by sharing content, including posts, photos, and videos. Groups are a key feature that helps users connect with others who share similar interests. Events can bring together individuals who share common passions and perspectives, thereby enriching their collective experiences.

One of the biggest feature of Facebook is its news feed, which delivers an infinite stream of content from friends, pages, and groups based on relevance and user preferences. Public figures, including politicians, use Facebook to connect with followers, share news, and engage in political discussion.

Despite its widespread popularity, Facebook has frequently faced challenges such as privacy concerns, data breaches, and distribution of misinformation. However, despite these issues, it continues to maintain its position as the most widely utilized platform by billions of users worldwide [5].

In summary, Facebook represents more than just a social networking site. It is a dynamic platform that shapes how people connect, communicate, and collaborate online, while also providing a rich field for academic study and research.

1.1.2 Instagram

With a user base exceeding one billion globally, Instagram [6, 7, 8] has established itself as a widely utilized social media platform. It enables users to share photos, videos, and messages. Through various features like Stories, Feed, Live, and Direct messaging, individuals, including teenagers, use Instagram for diverse purposes such as documenting significant moments, connecting with friends and family, building communities, and exploring shared interests. Figure 1.1 presents an example of an Instagram post, showcasing a typical representation of content found on the platform.

Users can follow each other and be followed, although mutual following, as seen on Facebook, is not necessary. On Instagram, users can customize their privacy settings to control who can view their posts. By default, content is visible to anyone unless the user chooses to set their account to private.

Figure 1.1: A post on Instagram, which includes likes and comments, is made by the account *zuzana_ caputova*. Source: `https://www.instagram.com/zuzana_caputova/`

Being active on Instagram involves uploading photos or videos, which can be accompanied by captions, locations, or user tags. Users also have the flexibility to choose whether to share content exclusively with their followers or extend its reach beyond the app.

Additionally, commenters and likers play a significant role in user engagement on Instagram. They can interact with posts by leaving comments, expressing their thoughts, opinions, or appreciation for the content. Likers, on the other hand, indicate their approval or enjoyment of a post by tapping the heart icon. These interactions contribute to the overall engagement and community-building on the platform.

1.2 Graph Theory and Social Networks

Graph theory [9, 10, 11], a mathematical discipline focusing on the study of graphs provides a solid framework for unraveling the complex architectures of social networks. In these networks, nodes and edges are most commonly used as representations of individuals and their connections or interactions. By using graph theory, researchers can understand social relationships better. This can help to study how information spreads, how influence works, and how communities come together. In the end, graph theory becomes a powerful tool that helps people understand the complexities of social networks.

Figure 1.2 provides an example of a graph representation of connected actors in a social network, which can illustrate the relationships and interactions among users.

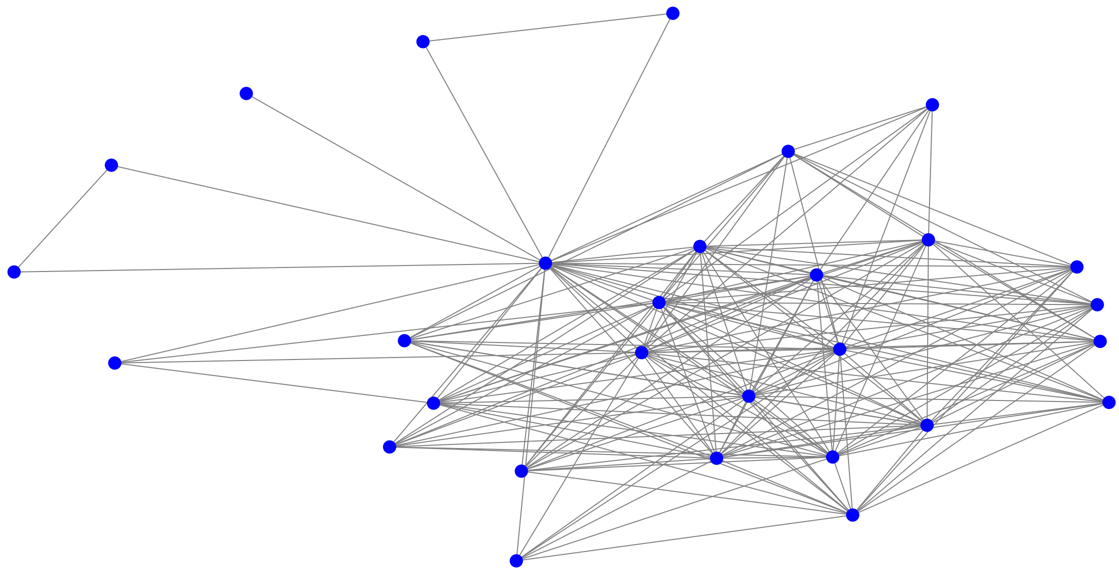

Figure 1.2: Example of a graph representation of connected actors in a social network, illustrating relationships and interactions among users.

It is essential to understand some fundamental concepts in graph theory, which are used to model and analyse social networks. These concepts can help us understand the structure and dynamics of these networks better.

- **Vertices**: Represent individual users or entities within the social network.
- **Edges**: Represent the relationships or interactions between the users. These can be directed (e.g., Instagram follows) or undirected (e.g., Facebook friendships).
- **Degree**: The number of edges connected to a vertex. In social networks, this could represent the number of followers or friends a user has.
- **Path**: A sequence of edges that connects two vertices. This can represent the shortest route of interactions or connections between two users.
- **Distance**: The number of edges in the shortest path between two vertices. This can measure the closeness or separation between users in a network.
- **Subgraph**: A subset of a graph's vertices and edges that forms a graph. Subgraphs can be used to study smaller communities within a larger network.
- **Clusters** (Communities): Groups of vertices that are more densely connected internally than with the rest of the graph. These can represent social circles or groups with common interests.

Additionally, the representation can be adjusted to almost anything to enhance visualization.

The application of graph theory to social networks helps in understanding various phenomena such as the spread of information, influence dynamics, and community detection.

1.3 Scraping

As shown in Figure 1.3, Web scraping [12, 13, 14] is a data extraction technique employed to retrieve information from websites across the internet. It involves a series of systematic steps designed to access, parse, extract, transform, and store data from web pages. At its core, web scraping involves accessing the HTML code of web pages, extracting relevant data elements, and transforming them into a structured format for processing or further analysis.

The process begins with accessing web pages using software tools or programming languages capable of fetching the HTML content. These tools acquire the web page data, providing a foundation for parsing and extraction tasks. Once the HTML content is obtained, the parsing stage comes into play, where the structure of the HTML document is analysed to identify specific elements containing the desired information.

Following parsing, the extracted data undergoes a transformation phase to convert it into a structured format suitable for analysis. This may involve cleaning and standardizing the data, removing HTML tags, formatting dates or numbers, and handling any inconsistencies or anomalies encountered during extraction. By organizing the data into a structured format, it becomes easier to handle.

Finally, the transformed data is stored in a database, spreadsheet, or other storage systems for future use. Storing the scraped data enables easy access, retrieval, and analysis, allowing researchers and analysts to use the extracted information for various applications. Common storage options include relational databases like MySQL or PostgreSQL, or simple file formats like CSV or JSON.

1.3.1 Facebook Scraper

Facebook Scraper [15] is a Python library developed by Kevin Zúñiga, available on GitHub, designed for scraping public data from Facebook. With this tool, users can extract various types of information from Facebook pages, including posts, comments, reactions, and other metadata.

The library leverages web scraping techniques to access and retrieve data from Facebook's web pages, mimicking the behaviour of a web browser. It does not rely on Facebook's official API, making it suitable for extracting data that may not be accessible through official channels.

Facebook Scraper provides a convenient interface for specifying the target Facebook

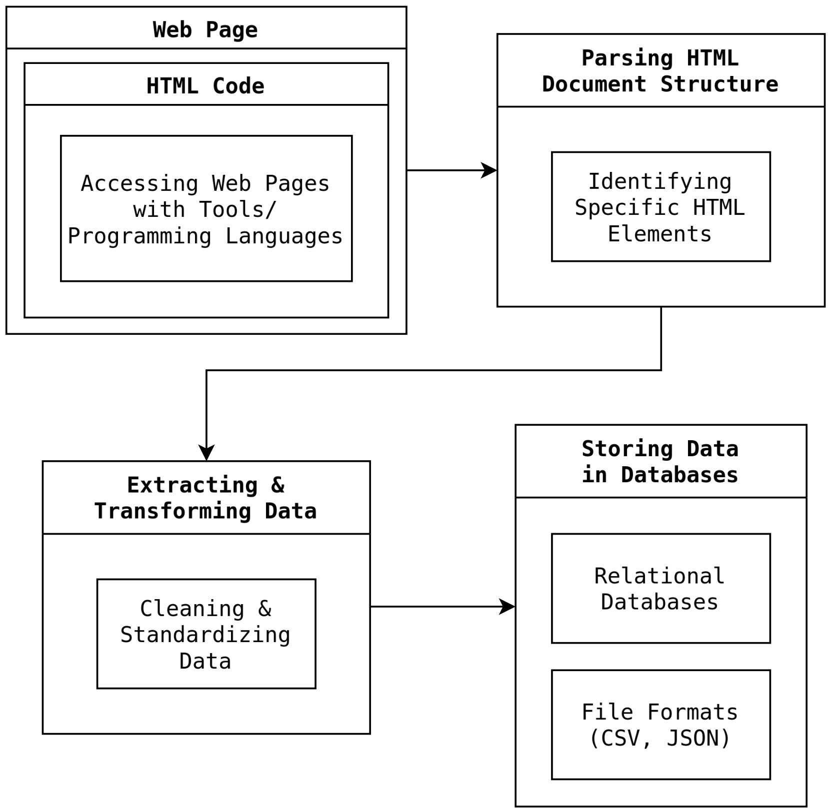

Figure 1.3: Process of scraping

page and the type of data to extract. It also includes features for filtering, and sorting the retrieved data, allowing users to customize their scraping operations according to their specific requirements.

However, it is essential to note that scraping data from Facebook may raise ethical and legal considerations, particularly concerning user privacy and data usage policies. Users of Facebook Scraper should be cautious and ensure compliance with Facebook's terms of service and relevant laws and regulations when using the library for data extraction purposes.

In Subsection 2.1.1, we will explore the practical application of this library while also examining potential setbacks that will be address in detail.

1.3.2 apify/instagram-scraper

Apify [16], a versatile web scraping and automation platform, offers Instagram Scraper [17], designed to extract valuable data from Instagram. With Instagram Scraper, users can effortlessly gather information such as user profiles, posts, comments, and hashtags, enabling them to gain valuable insights into trends, user behaviour, and market dynamics. Whether for market research, competitor analysis, or content curation, Apify's Instagram Scraper provides a reliable solution for businesses and researchers seeking to harness the wealth of data available on Instagram.

1.3.3 Instaloader

Instaloader, developed by Alexander Graf [18], is a Python tool designed for interacting with Instagram data programmatically. It provides a comprehensive set of functionalities for accessing and downloading Instagram content without the need for an official API.

One of the key features of Instaloader is its ability to fetch various types of data from Instagram, including profiles, posts, stories, likes, comments, and more. Users can specify the type of data they want to retrieve and define filters based on parameters such as hashtags, geotags, or user IDs.

In terms of functionality and features, the chosen library offers several key capabilities. Firstly, it provides a convenient means to retrieve all posts associated with a user, identified either by their username or user ID, along with their respective attributes. Additionally, the library can extract comments from each post. Notably, the library streamlines the data retrieval process by eliminating the need for interaction with the Instagram app, offering a self-contained solution within its framework. Furthermore, it is offering both command-line functionality with numerous customizable flags and a programming interface for precise control and integration within larger applications.

Moreover, Instaloader facilitates the downloading of media content from Instagram, allowing users to save images, videos, and profile pictures locally on their machines. This capability is particularly useful for researchers, data analysts, and enthusiasts who wish to archive or analyse Instagram content offline. Additionally, Instaloader supports advanced functionalities such as profile crawling, enabling users to go through followers, followings, and interactions of Instagram profiles. This feature is valuable for studying social network dynamics, user engagement patterns, and influencer marketing strategies on the platform.

Overall, Instaloader serves as a valuable tool for accessing, downloading, and analysing Instagram data in a flexible and efficient manner. Its wide range of functionalities makes it a popular choice among researchers, data scientists and journalists seeking to gain insights into Instagram's ecosystem of content and interactions.

In Subsection 2.1.2, we will look into the implementation of this library.

1.4 Sentiment Analysis

Sentiment analysis [19], also known as opinion mining, is a process in natural language processing (NLP) [20] that involves determining the sentiment expressed in a piece of text. The goal is to understand whether the text expresses positive, negative, or neutral sentiment towards a particular subject or entity.

Sentiment analysis algorithms typically work by analysing the words, phrases, and context within the text to determine the sentiment. This can involve various techniques, such as lexical analysis to identify positive or negative words, machine learning algorithms trained on labeled datasets, or more advanced methods like deep learning models.

1.4.1 BERT

Crafted by Google, this NLP model is designed to process and grasp human language in a manner more aligned with human cognition. BERT [21, 22] (Bidirectional Encoder Representations from Transformers) and other language models are trained on vast amounts of text from the internet, absorbing the patterns and meanings of words and sentences. What distinguishes it is its capability to simultaneously analyse words in a sentence from both directions, thereby considering their contextual significance comprehensively. This bidirectional approach helps to understand nuances and relationships between words better than previous models. Essentially, BERT acts as a sophisticated tool for computers to comprehend and generate human-like language.

1.4.2 SlovakBERT

A language model tailored specifically for Slovak, SlovakBERT [23] stands out in the field of transformers-based models. Despite the availability of multilingual models that support Slovak language, the development of a dedicated Slovak model holds promise for achieving superior results and optimizing language processing efficiency.

It adopts the RoBERTa [24] (Robustly Optimized BERT Approach) architecture and undergoes training using a web-crawled corpus. This selection of architecture and training data is customized to enhance performance specifically for the Slovak language.

The authors of SlovakBERT address the lack of evaluation standards for the Slovak language by creating their own tests. These tests encompass various tasks including part-of-speech tagging, semantic textual similarity, sentiment analysis, and document classification. By establishing these benchmarks, they provide a comprehensive framework for evaluating Slovak language models, including SlovakBERT, which could become a standard for future assessments.

Sentiment Analysis model based on SlovakBERT

It utilizes the specialized features and contextual understanding of the Slovak language provided by SlovakBERT to accurately determine the sentiment expressed in Slovak text. Leveraging the bidirectional contextual understanding of words and phrases, the model comprehensively analyses the text to discern whether the sentiment expressed is positive, negative, or neutral. By training SlovakBERT on sentiment-related tasks and datasets, the model learns to identify different expressions of sentiment, capturing subtle differences and contextual cues specific to Slovak language usage. This allows the sentiment analysis model to provide precise and contextually appropriate assessments of sentiment in Slovak text. This model has been trained on a large dataset of Twitter data, making it suitable for analysing sentiment in social media comments.

1.5 Related Work

The study of social media and its impact on elections and crises has attracted attention from researchers. This section reviews the existing literature on social media's role in political campaigns, digital campaign strategies, and social media analysis during crises. These studies provide a foundational understanding that will inform the analysis of Instagram user behaviour during the 2024 Slovak presidential election.

1.5.1 Social Media and Elections

Social media platforms have revolutionized the way political campaigns are conducted and how voters engage with political content. Numerous studies have explored the role of social media during election periods, highlighting its impact on political participation, campaign strategies, misinformation, and user engagement.

Bond et al. [25], in their study demonstrated that social media can significantly increase voter turnout and political engagement through informational and social mechanisms. By conducting a large-scale field experiment on Facebook, the researchers found that messages encouraging users to vote, combined with social pressure from seeing friends who had voted, led to higher voter turnout. This study underscores the potential of platforms like Instagram to mobilize voters and foster political participation during elections.

In "The Role of Social Media in Political Campaigns: Insights from the 2016 U.S. Presidential Election" Allcott and Gentzkow [26] analysed how political campaigns utilize social media to spread information and engage with voters. The study provides an examination of the strategies employed by political actors to influence voter opinions and behaviours through targeted advertisements and viral content. These insights are particularly relevant for understanding the digital campaign strategies that could be employed during the 2024 Slovak presidential election.

The issue of misinformation is critically examined by Shao et al. in their research "The spread of low-credibility content by social bots" [27] This study investigates how automated accounts, or bots, spread false information on social media, thereby influencing public opinion and potentially affecting election outcomes. The findings highlight the challenges of combating misinformation and the importance of monitoring automated activity to maintain the integrity of political discourse on platforms like Instagram.

These studies collectively enhance our understanding of social media's role in elections, particularly in terms of political engagement, campaign strategies, misinformation spread, and user behaviour.

1.5.2 Digital Campaign Strategies

The arrival of social media has transformed political campaign strategies, enabling candidates and parties to engage directly with voters, personalize messages, and rapidly distribute information. Various studies have explored these digital campaign strategies, highlighting their effectiveness and the unique characteristics of different social media platforms.

Kaplan and Haenlein, in their seminal work "Users of the World, Unite! The Challenges and Opportunities of Social Media" [28] discuss the strategic use of social media in political campaigns. They argue that social media platforms offer opportunities for targeted advertising, direct voter engagement, and real-time feedback. The study highlights how successful digital campaigns are those that effectively integrate social media into a broader communication strategy, leveraging the interactive nature of these platforms to build a strong and engaged supporter base.

A study by Enli and Moe, titled "Social Media and Election Campaigns: Key Tendencies and Ways Forward" [29] investigates the role of social media in modern election campaigns. The authors identify key tendencies, such as the personalization of political messages, the use of influencers and endorsements, and the strategic timing of posts to maximize reach and impact. They also discuss the challenges of managing a campaign's online presence, including the risks of negative feedback and the need for rapid response strategies.

Chadwick and Stromer-Galley [30], in their article explore how digital media is reshaping the internal dynamics of political parties and their campaign strategies. They argue that digital media has the potential to democratize political participation by enabling more direct interaction between candidates and voters. However, they also caution that this can lead to a centralization of power within parties, as those who control digital media strategies gain greater influence.

These works collectively illustrate the evolving nature of political campaigns in the digital age. They provide a understanding of the strategies employed by campaigns to leverage social media, highlighting the importance of integrating digital tools into traditional campaign efforts. For the 2024 Slovak presidential election, these insights are invaluable for analysing how candidates might use Instagram and other platforms to engage with voters, personalize their messages, and influence public opinion.

1.5.3 Social Media Analysis During Crises

This subsection explores recent studies on social media behaviour during crises, particularly focusing on the COVID-19 pandemic. Analyzing platforms like Instagram, researchers have analysed into user behaviour, information spread, and the impact on mental health. These findings offer valuable implications for crisis communication strategies and combating misinformation.

Zarei et al. in their study, "A First Instagram Dataset on COVID-19" [31] created a dataset of Instagram posts related to the COVID-19 pandemic. This dataset includes a wide range of data points such as post text, hashtags, user engagement metrics (likes and comments), and metadata like timestamps and user profiles. The study provides valuable insights into how users leveraged Instagram to share information, express emotions, and engage with content about COVID-19. By analysing this dataset, the researchers identified patterns in user behaviour, including the spread of information and misinformation, and the types of content that made the most engagement. This work is particularly relevant for understanding social media dynamics during global crises, offering a framework for analysing user behaviour and engagement on visual social media platforms like Instagram.

Another significant contribution is the work by Cinelli et al. [32], which analyses the spread of information across multiple social media platforms during the COVID-19 pandemic. The study highlights the differences in how information propagates on platforms such as Twitter, Instagram, and Facebook, providing a comparative analysis that underscores the unique characteristics of each platform in crisis communication.

Furthermore, Gao et al. [33] investigated the occurrence of mental health issues and the spread of related information on social media during the COVID-19 outbreak. Their findings suggest that social media can be a double-edged sword, offering support and community while also potentially worsening anxiety and misinformation.

These studies collectively enhance our understanding of social media's role during global crises, particularly in terms of user behaviour, information spread, and the impact on public perception and mental health. They provide a foundational basis for analysing how Instagram users behave in response to crises.

Chapter 2

Methods and Their Applications

In this chapter, we outline the methodology used in this thesis, detailing the processes of data collection, storage, and sentiment analysis. We explain the steps involved in gathering relevant data, setting up an organized storage system, and using NLP [20]. Our goal is to provide a straightforward overview of our research process, focusing on the methods we used. Detailed findings will be discussed in Chapter 3.

2.1 Data Collection

For data collection, we primarily relied on Instaloader [18], a tool that enabled us to extract posts and comments from Instagram [6] efficiently. While initially considering Facebook Scraper [15] for gathering Facebook [3] data, we encountered limitations that made it ineffective for our purposes.

2.1.1 Facebook Scraper

Exploring the utilization of Facebook Scraper [15] as part of our methodology aimed to gather valuable user data from Facebook. However, our experience with this tool introduced unexpected challenges.

We aimed to construct a Facebook Friends list visualization, where inserting one user would display a graph of all friendships, with the main node connected to everyone. However, to accomplish this task, we faced a significant challenge. We had to scrape not only the friends of the inserted user but also the friends of each of those friends, which considerably slowed down the whole process.

Upon initiating the scraping process, we faced an unforeseen obstacle — a three-day ban, as indicated by this error message:

"`facebook_scraper.exceptions. TemporarilyBanned: You're Temporarily Blocked`"

This interruption highlighted the challenges and limitations of scraping data from Facebook, particularly with the platform's restrictions in place.

Despite attempts to resume data collection after the ban was lifted, our efforts were short-lived. Shortly after resuming scraping activities, we encountered another ban, extending the interruption for an additional three days. These repeated bans made it clear to us that using the Facebook Scraper for data collection comes with its own set of challenges and unpredictability.

As a result, we had to rethink whether depending only on this method was practical and look into other ways to collect and analyse social media data for our thesis research.

2.1.2 Instaloader

Instaloader [18] played a key role in our project by enabling us to gather Instagram data efficiently. This powerful tool allows for the retrieval of various types of Instagram data, including posts, likes and comments, making it an ideal choice for social media analysis.

Our primary objective was to utilize Instaloader for gathering Instagram data and storing it locally. We chose this library due to its robust functionality and ease of use, which simplified the data collection process.

The process began by defining the scope of data needed, which included posts, comments, and user metadata from specific Instagram profiles. To manage the volume and complexity of the data, we employed a structured approach:

Initial Setup and Login

Our first action was to log into the Instagram account using a web browser. Following this, we stored the logged cookies into a separate file for future access and authentication. We utilized a script to export the cookie file using Firefox. This file was necessary for Instaloader to function properly with a logged-in user.

Data Retrieval

Our attention shifted to all candidates participating in the 2024 Slovak presidential election. We proceeded by scraping and collecting their posts, as well as the associated comments, within the specified timeframe. To accomplish this task, we developed a script capable of extracting all posts, comments, and replies to those comments within the given timestamp. We focused on retrieving older data as the newest posts were still dynamically evolving, with ongoing additions of comments and likes.

Checkpoints and Data Integrity

Given the prolonged duration of the scraping process, which involved retrieving data from specific dates, we implemented checkpoints to ensure efficient data retrieval. These checkpoints prevented redundant scraping of the same content and enabled us to retrieve posts within the specified time frame efficiently. This was particularly important because Instagram imposes rate limits on data retrieval, and redundant requests could result in temporary bans or slowed data access.

Local Storage

The scraped data were stored in a local database, providing easy access for further analysis. This approach ensured that the data remained readily available for future reference and utilization in various analytical endeavors. The structure and organization of this local storage will be discussed in detail in Section 2.2.

Final Remarks and Code Example

Outlined below are the advantages from our approach:

- **Efficiency**: By using checkpoints and local storage, we minimized redundant data retrieval, making the process more efficient and less time-consuming.
- **Data reliability**: Checkpoints ensured that data retrieval could resume from the last successful point in case of interruptions.
- **In-depth Analysis**: Storing data locally enabled us to perform detailed analyses without repeated access to Instagram's servers, reducing the risk of rate limits and improving reliability.
- **Scalability**: Our approach is highly scalable, allowing us to easily adapt and expand our data retrieval and analysis capabilities to accommodate larger datasets or additional sources. This scalability ensures that our methods remain robust and effective even as the scope of our project evolves.

The following Python code snippet illustrates the process of scraping posts and comments from a user profile using Instaloader:

Algorithm 2.1: Scraping posts and comments of Instagram profile "zuzana_caputova"

```
import instaloader

loader = instaloader.Instaloader()
loader.load_session_from_file(username, file_path)
```

```
profile = instaloader.Profile.from_username(
    loader.context, 'zuzana_caputova'
)

for post in profile.get_posts():
    print(post.caption)

    for comment in post.get_comments():
        print(comment.text)
```

These steps play a crucial role in the extraction of data, guaranteeing its analysis and interpretation for future purposes. With this systematic approach, we successfully collected data crucial for comprehending user behaviour and engagement patterns.

Disclaimer All data collected from Instagram and Facebook were used solely for research purposes in this study. The data will be securely stored during the research period and will be deleted upon the completion of the study.

2.2 Data Organization

In the context of our bachelor's thesis, effective data organization played a key role in enhancing our analysis of Instagram data collected during the 2024 Slovak presidential election. This section outlines our methodology for organizing and managing the data.

2.2.1 Storing Data in an SQLite Database

To ensure scalability and manageability of the data, we stored the retrieved Instagram posts and comments in a structured database. This database setup allowed us to perform complex queries and analyses efficiently, supporting the overall research objectives.

To handle the large amount of Instagram data we collected, we made a database system illustrated in Figure 2.1. This system allowed us to organize the data in a structured manner, making it easier to manage and analyse.

Within the database, we set up three main tables: "users", "posts", and "comments". The "users" table stored information about Instagram users, such as their usernames, the number of posts they have made, and their follower counts. The "posts" table was used to store details about each post, including unique identifiers, the number of likes, timestamps, and any captions (description) associated with the post. Similarly, the

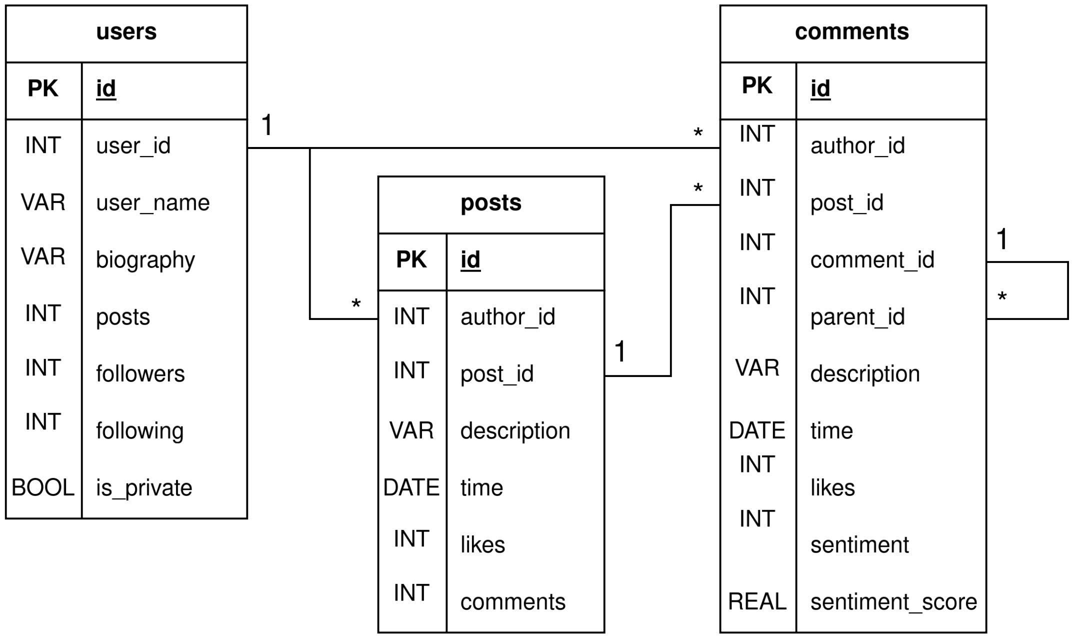

Figure 2.1: Diagram illustrating the structure of database we made to organize data from Instagram.

"comments" table was designed to hold comments (description) made on individual posts, along with information like the userids of the commenters, the timestamps of their comments, and a "parent_id" field.

The "parent_id" field in the "comments" table indicates whether a comment was made directly to the post or as a reply to another comment. If the "parent_id" is NULL, it means the comment was made directly to the post. On the other hand, if the "parent_id" contains an ID of another comment, it indicates that the comment is a reply to that specific comment.

This organization of tables ensured that we could efficiently store and retrieve Instagram data for our analysis.

2.3 Sentiment Analysis

Sentiment analysis is essential for understanding public perception and attitudes expressed in social media data. In this section, we describe our approach using a classifier based on SlovakBERT [23].

2.3.1 Sentiment Analysis Model

This SlovakBERT model, specifically designed for the Slovak language, can identify three levels of sentiment:

-1 → Negative sentiment

0 → Neutral sentiment

1 → Positive sentiment

By utilizing this model, we aimed to assign sentiment labels to each comment in our database, providing an understanding of the overall sentiment expressed within the Instagram data collected during the 2024 Slovak presidential election.

2.3.2 Implementation

To implement sentiment analysis on our Instagram data, we used a sentiment analysis classifier based on SlovakBERT. This involved processing each comment in our database through the classifier to generate sentiment label and score.

Scoring Comments

Here is an example of how this sentiment analysis classifier is used to label and score comments:

```
>>> from transformers import pipeline
>>> pipe = pipeline(
        "text-classification",
        model="kinit/slovakbert-sentiment-twitter"
    )

>>> pipe('Fakt môžu robit, vsetko, co chcu?!')
[{'label': '-1', 'score': 0.9985588192939758}]

>>> pipe('Bude sa dat volit prezident zo zahraničia?')
[{'label': '0', 'score': 0.9910099506378174}]

>>> pipe('Povedal len pravdu...a este dost slusne')
[{'label': '1', 'score': 0.9964924454689026}]
```

This example demonstrates the process of evaluating comments using the sentiment analysis model. Each comment is passed through the model, which assigns a sentiment label ('-1', '0', or '1') based on the sentiment expressed in the comment. Additionally, the sentiment score, indicating the intensity of sentiment, is computed and stored in the database for further analysis.

By employing sentiment analysis, we aimed to gain deeper insights into the sentiments expressed by Instagram users during the 2024 Slovak presidential election, allowing for a more precise analysis of public opinion and attitudes.

Chapter 3

Obtained Results

In this chapter, we present the findings of our analysis of Instagram [6] posts and comments related to our candidates. We provide insights into the relationship between likes and comments, commenter behaviour, and sentiment expressed in comments.

3.1 Scraped Overview

This section presents a detailed overview of the data gathered from Instagram. Table 3.1 presents an overview of key metrics extracted from the platform, including number of posts, likes, and comments associated with each candidate. Additionally, Table 3.2 provides insights into candidate engagement metrics, showcasing the mean and median number of likes and comments received per candidate.

Following the data extraction process from Instagram, the collected data underwent thorough analysis to uncover patterns and trends in candidate engagement. Next sections go deeper into the analysis, offering detailed examination of the dynamics of social media activity surrounding the candidates.

3.2 Likes and Comments

In this section, we present our thorough analysis of likes and comments on Instagram posts over the whole scraped time, as illustrated in Figure 3.1 and 3.2. We employ a variety of analytical methods to delve into engagement trends and patterns.

3.2.1 Trend Analysis

Our analysis begins with a detailed examination of engagement trends over time. By analysing the data, we identify significant patterns and fluctuations in likes and comments. This exploration enables us to gain a deeper understanding of how user engagement evolves across different phases of the 2024 Slovak presidential election.

Table 3.1: Overview of scraped data from Instagram, including the number of posts, likes, and comments for each candidate, covering the period from August 1, 2023 to May 19, 2024. The "Total Comments" column represents the total number of comments according to the posts attribute.

Name	Posts	Likes	Comments	Total Comments
Andrej Danko	284	155,683	9,297	11,647
Krisztián Forró	332	2,781	342	340
Štefan Harabin	193	41,438	2,175	2,330
Ivan Korčok	713	4,278,244	64,719	86,071
Marian Kotleba	129	45,867	2,787	3,128
Igor Matovič	19	63,251	3,186	5,018
Peter Pellegrini	431	674,776	26,584	62,415
Sum	2,101	5,262,040	109,090	170,949

Table 3.2: Summary of candidate engagement metrics, including mean and median likes and comments per candidate, over the period from August 1, 2023 to May 19, 2024.

	Likes per Post		Comments per Post	
Name	Mean	Median	Mean	Median
Andrej Danko	548	298	32	18
Krisztián Forró	8	6	1	0
Štefan Harabin	214	179	11	6
Ivan Korčok	6,000	3,560	90	39
Marian Kotleba	355	244	21	12
Igor Matovič	3,329	2,644	167	162
Peter Pellegrini	1,565	1,184	61	43

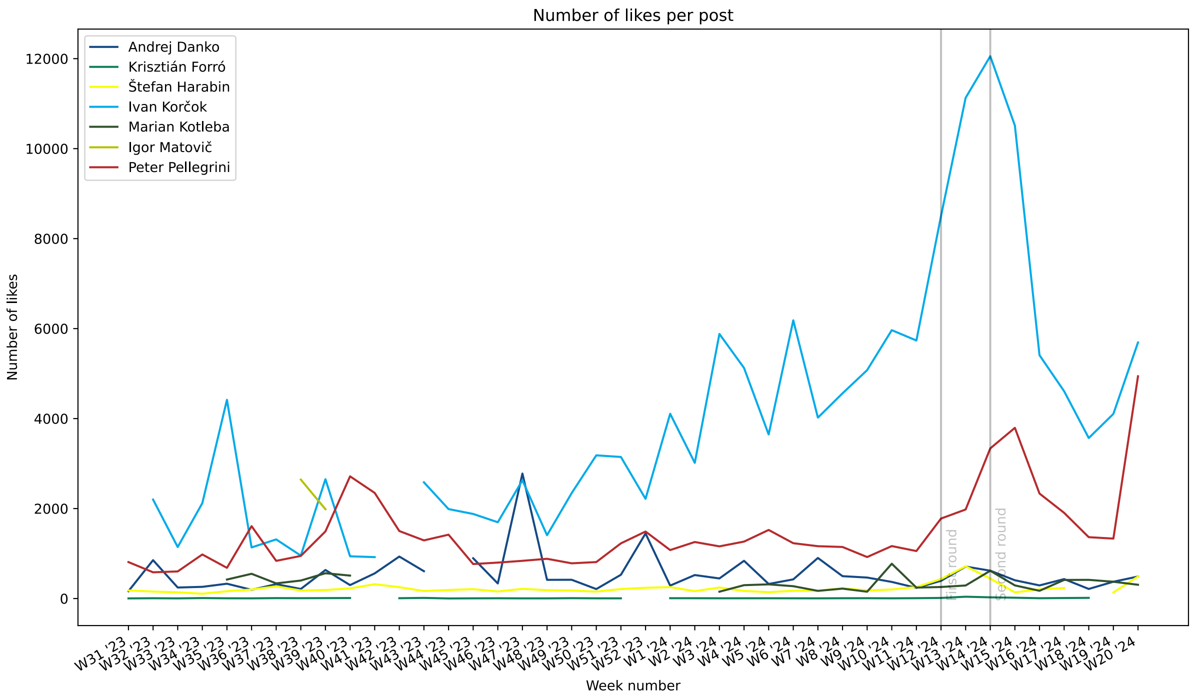

Figure 3.1: Illustrates the weekly likes received by each candidate on Instagram. Y-axis shows total likes, while X-axis denotes weeks. Offers a perspective on engagement trends among candidates over time.

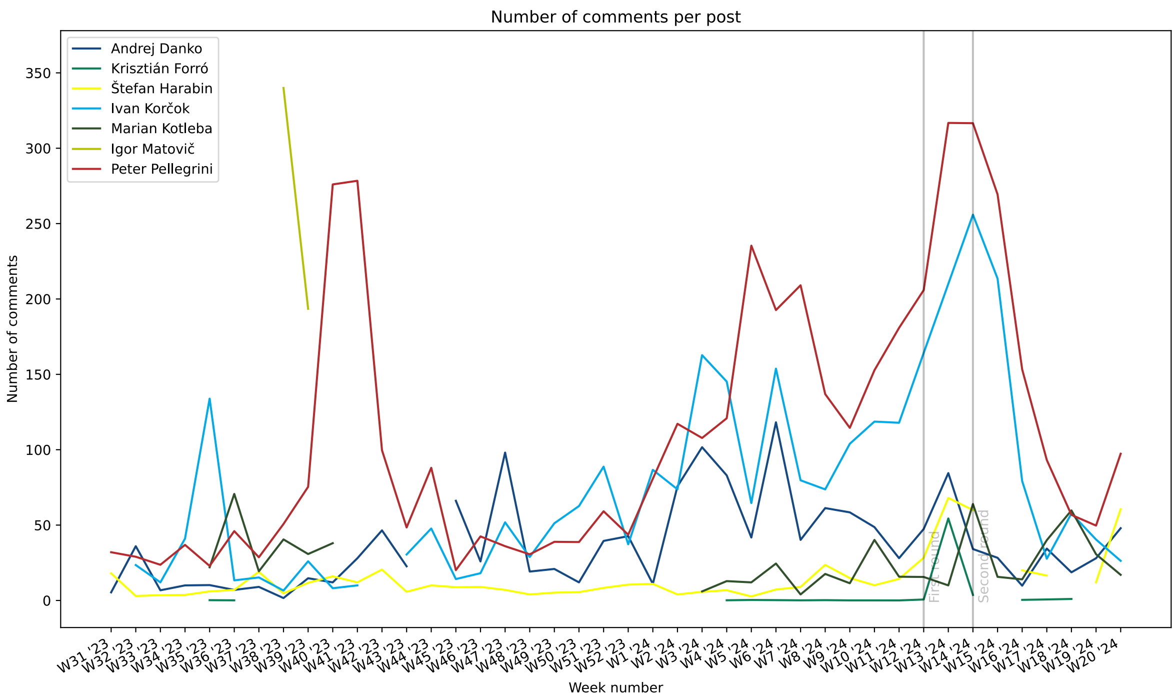

Figure 3.2: Displays weekly comments (total) received by each candidate on Instagram. Y-axis represents total comments, and X-axis indicates weeks. Highlights comment trends among candidates over time.

3.2.2 Peak Identification

Finding from our analysis is the presence of a prominent peak in engagement towards the end of both the first and second rounds. This spike indicates increase in user interactions during those specific periods. This pattern suggests that users are particularly active or interested at these times, which could be due to the fact that the election is taking place. Identifying these peaks helps us understand user behaviour better and can guide us in planning future analyses to accurately interpret engagement during these critical periods.

3.2.3 Interpretation Challenges

Our analysis faced challenges due to missing data from weeks when the candidates did not post anything. These gaps made it difficult to fully understand engagement trends during those times. Although we could not control this, it is important to note how it impacted our analysis.

In summary, our analysis of likes and comments on Instagram posts offers an overview of user engagement dynamics during the election period. By examining trends, identifying peaks, and acknowledging interpretation challenges, we have deepened our understanding of how users interacted with candidates' content over time.

3.3 Commenters

Next, we present the findings of our analysis of commenter behaviour on Instagram posts related to our candidates. We explore the total number of commenters and overlap between commenters on posts related to different candidates.

3.3.1 Total Number of Commenters

Our analysis uncovers the total number of unique commenters across all candidates' posts, providing insights into the size of the commenting audience and their commonalities (see Table 3.3). This information helps us understand the scope of engagement and the types of users who actively participate in discussions.

Additionally, our findings reveal patterns within the commenting community. For instance, the commenter with the most comments is Ivan Korčok, with an impressive total of 422 comments. Furthermore, it is intriguing to note that 53 commenters alone contributed over 100 comments each, indicating a subset of highly engaged users driving significant discussion within the community.

Table 3.3: Aggregate count of unique commenters across posts from all candidates, recorded from August 1, 2023 to May 19, 2024.

Name	Comments	Commenters
Andrej Danko	9,297	3,728
Krisztián Forró	342	174
Štefan Harabin	2,175	847
Ivan Korčok	64,719	20,400
Marian Kotleba	2,787	1,086
Igor Matovič	3,186	1,607
Peter Pellegrini	26,584	8,425
Sum	109,090	31,053

3.3.2 Overlap Analysis

Delving into the overlap analysis, as depicted in Figure 3.3 we explore the engagement overlap between commenters on posts made by various candidates. This analysis helps us identify the extent of cross-engagement between different candidates' audiences, providing valuable insights into audience interaction dynamics.

3.4 Comment Sentiment

We present the results of our sentiment analysis of comments using the SlovakBERT [23] pipeline. We analyse the sentiment expressed in comments and visualize the sentiment distribution across all comments.

3.4.1 Sentiment Analysis Results

Our analysis reveals the sentiment expressed in comments as positive, negative, or neutral. In Figure 3.4 we present the distribution of sentiment labels and scores across all comments.

The visualization of sentiment analysis results offers a graphical representation of sentiment distribution within our comments.

In summary, this analysis shows findings regarding user interaction patterns, commenter tendencies, and sentiment expression observed across Instagram posts made by candidates. These insights can be used to refine content strategies and improve the effectiveness of future campaigns.

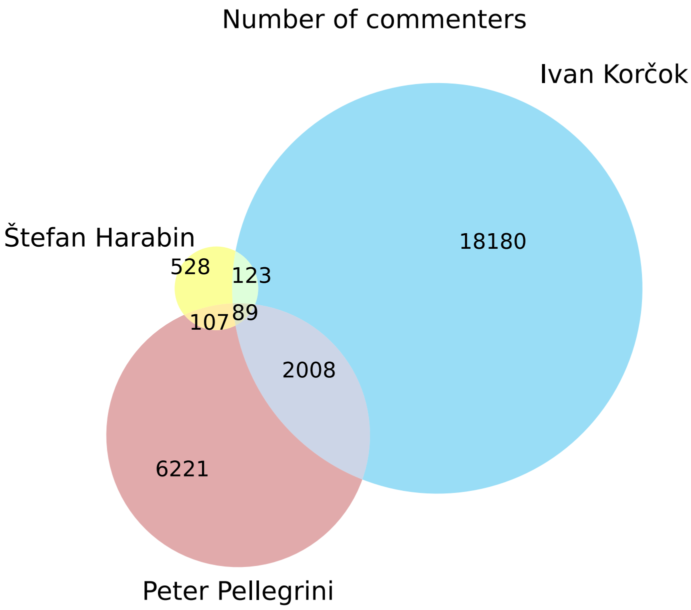

Figure 3.3: Venn diagram illustrating the overlap of commenters among the top three candidates in the 2024 Slovak presidential election, based on data collected during the entire scraping process. The numbers represent the count of unique commenters for each candidate.

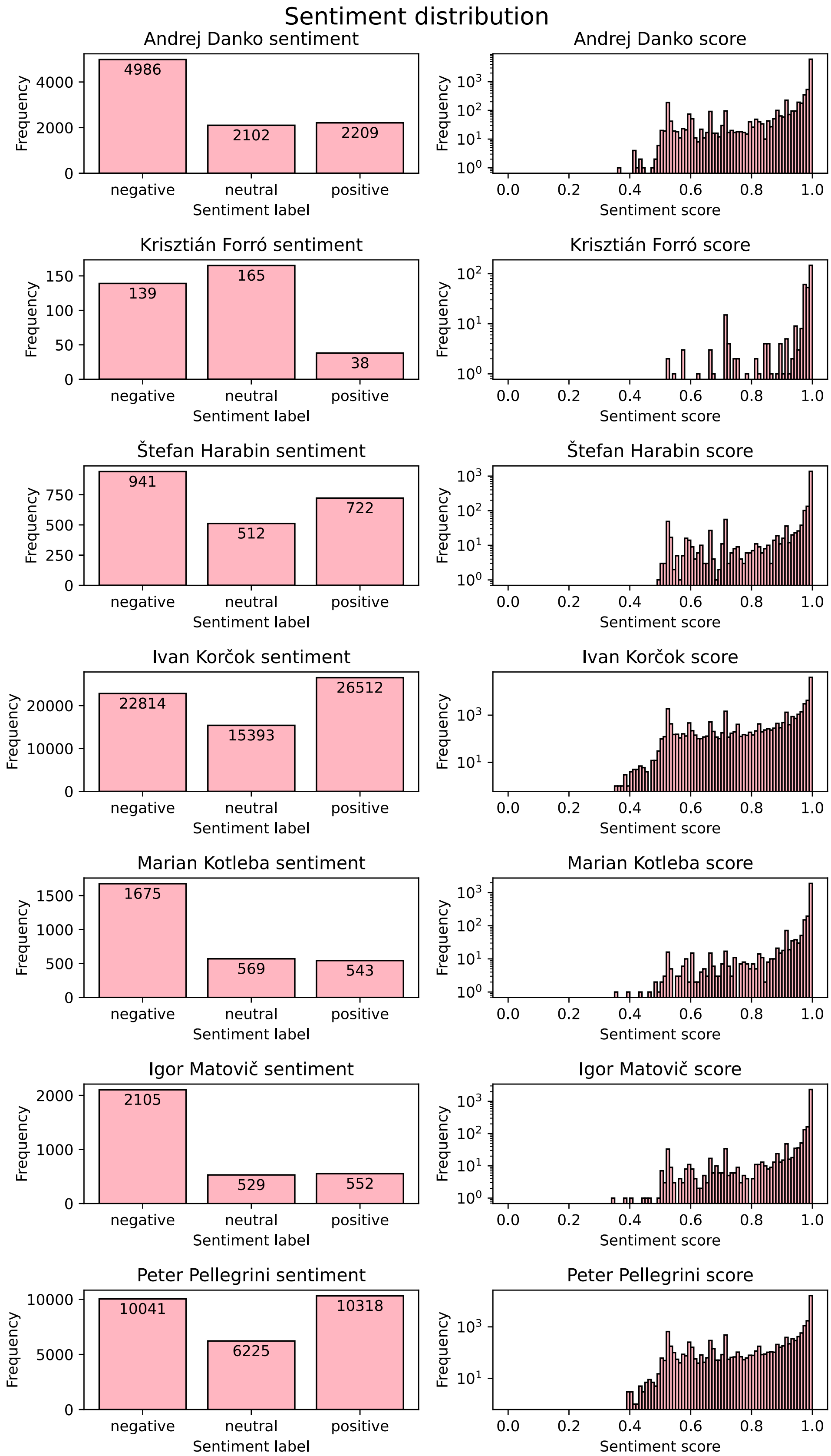

Figure 3.4: The figure presents two plots for each candidate: the count of positive, negative, and neutral comments (left); and the sentiment score assigned to the comments (right) using logarithmic scale from August 1, 2023 to May 19, 2024.

Table 3.4: Overview of scraped data from Instagram, including the number of posts, likes, and comments for each candidate, collected between the first and second rounds of the election. The "Total Comments" column represents the total number of comments according to the posts attribute.

Name	Posts	Likes	Comments	Total Comments
Ivan Korčok	133	1,522,102	21,241	28,255
Peter Pellegrini	66	157,496	7,003	20,204
Sum	199	1,679,598	28,244	48,459

3.5 Social Media Engagement During Runoff

In this section, we focus on analysing the social media engagement metrics of candidates who advanced to the second round of the 2024 Slovak presidential election. Specifically, we examine the data between the first round on March 23, 2024, and the second round on April 6, 2024, looking into the online activity and interaction dynamics during this crucial period. Table 3.4 provides an overview of scraped data only during this phase from Instagram, including the number of posts, likes, and comments for the candidates who advanced to the runoff phase.

3.5.1 Likes Analysis in Runoff Period

Examining the number of likes received by posts from the top two candidates provides a clear indicator of public engagement and support. Figure 3.5 illustrates the daily likes for these candidates on Instagram during the runoff period of the election.

This visualization brings attention to the trends in audience engagement for each candidate, showcasing peaks and fluctuations in public interest and support. By analysing these patterns, we can better understand the effectiveness of the candidates' social media strategies and the overall engagement levels during this critical final phase of the election.

3.5.2 Overlap of Commenters Among Top Two Candidates

The Venn diagram depicted in Figure 3.6 illustrates the overlap of commenters between the top two candidates who advanced to the runoff phase of the election. The data for this analysis was collected exclusively during the runoff phase of the election. Each section of the diagram represents the unique set of commenters for each candidate, while the overlapping region indicate commenters who engaged with both candidates' posts.

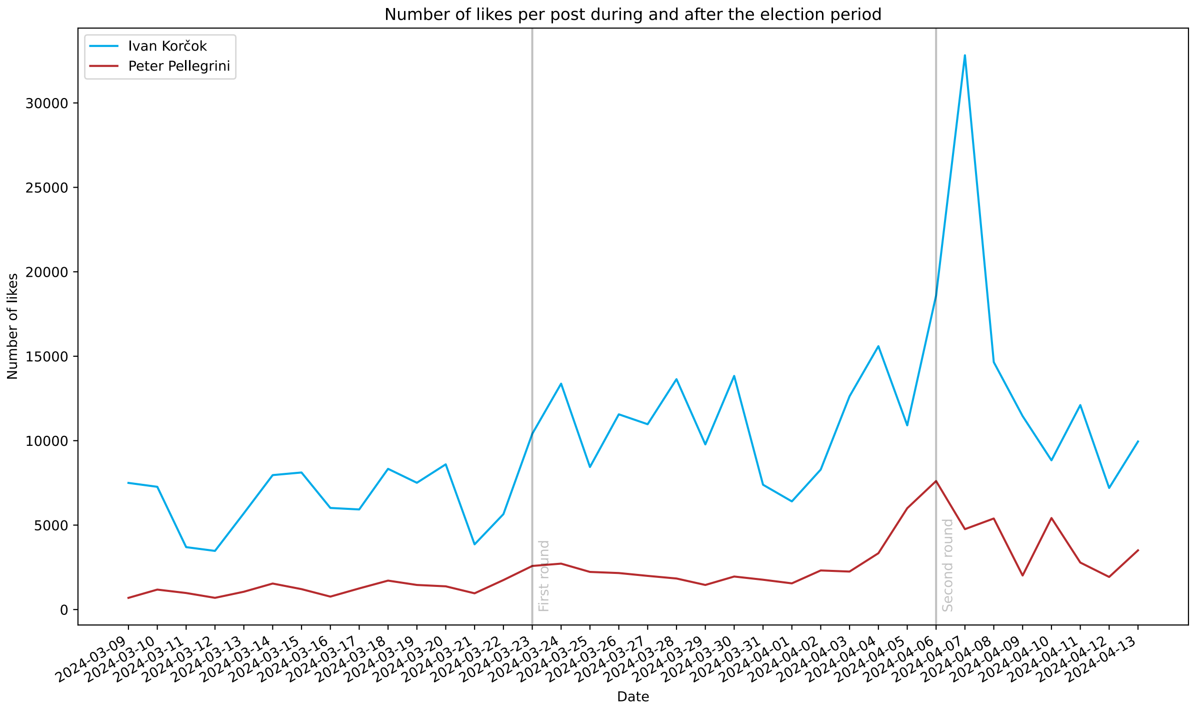

Figure 3.5: Illustrates the daily likes received by each candidate on Instagram during the end of the election period. Y-axis shows total likes, while X-axis denotes analysis days. Offers insights into engagement trends across candidates over time.

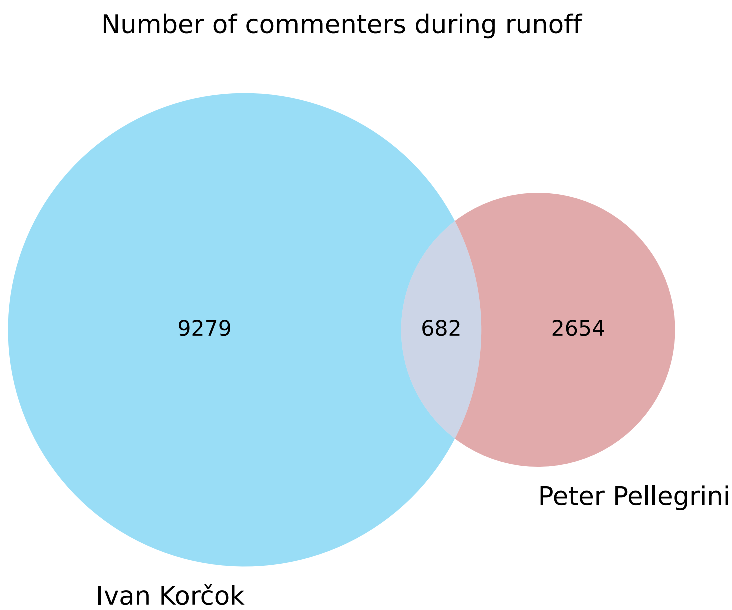

Figure 3.6: Venn diagram illustrating the overlap of commenters among the top two candidates who advanced to the runoff of the election, based on data collected during the runoff phase. The numbers represent the count of unique commenters for each candidate.

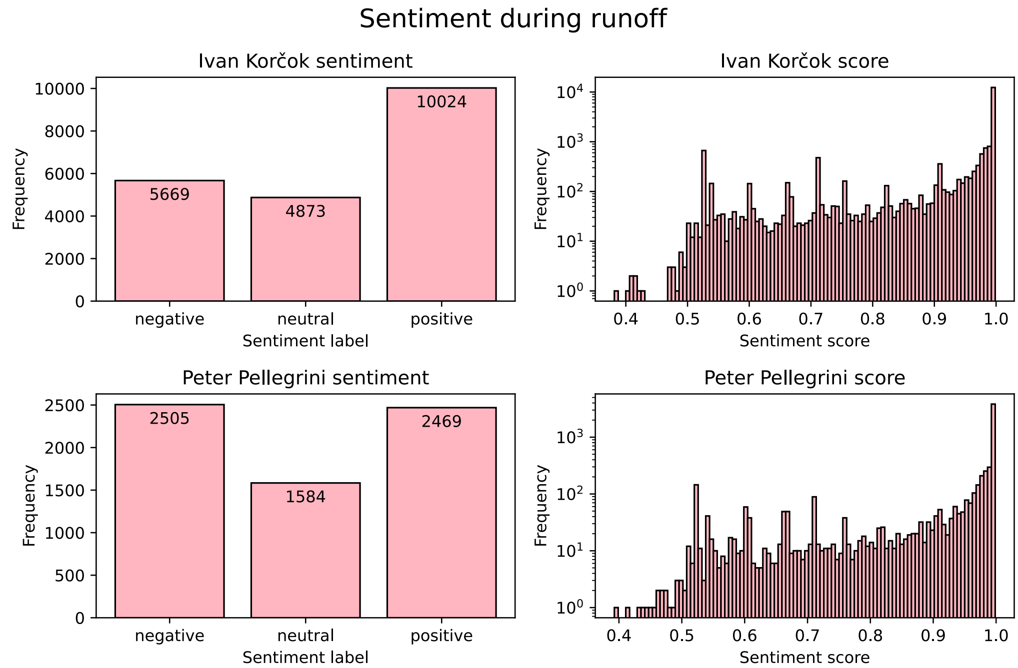

Figure 3.7: Histogram illustrating the sentiment of comments from posts made between the first round election and the second round election, specifically from March 23, 2024 to April 6, 2024.

3.5.3 Sentiment Analysis During Runoff Period

The histogram shown in Figure 3.7 depicts the sentiment of comments from posts made during the runoff phase of the election, between the first and second round of the election. This visualization offers a view into the sentiment expressed by users during this crucial period, helping to understand the dynamics of public sentiment and engagement patterns specific to the runoff phase.

3.5.4 Net Sentiment Analysis During Runoff Period

Understanding the sentiment expressed in comments during critical phases of an election can offer perspective on public opinion and the reception of candidates. Figure 3.8 illustrates the net sentiment trends during the runoff period of the 2024 Slovak presidential election. The net sentiment is calculated as the difference between the number of positive and negative comments each day. This visualization allows us to see fluctuations in public sentiment daily, highlighting key moments of positive or negative reception. By analysing these trends, we can identify periods of increased positive or negative engagement, offering a deeper understanding of voter sentiment dynamics during the crucial final stages and after the election.

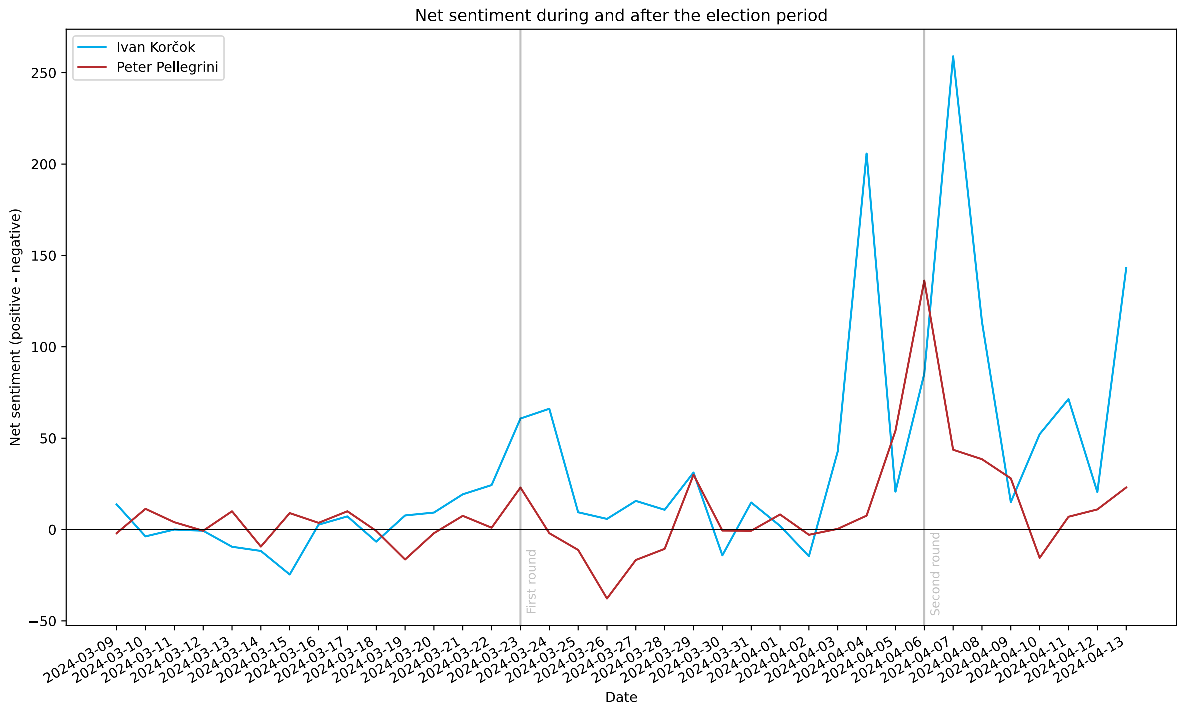

Figure 3.8: Net sentiment trends during the end of the 2024 Slovak presidential election period. The Y-axis represents the net sentiment, calculated as the difference between the number of positive and negative comments, while the X-axis represents the days during the election period.

3.6 Second Round Election Estimation

In this study, we utilized social media engagement data to estimate the outcomes of the second round of 2024 Slovak presidential election. We calculated the popularity of each candidate during two critical periods: the two weeks before the first round of election and the two weeks from the end of the first round to the end of the second round. Popularity metrics were derived by weighing comments more heavily than likes, using a predefined coefficient. These metrics were then used to compute a popularity coefficient, which adjusted the first-round results to estimate the second-round votes. The resulting analysis provided an estimate of vote distribution, reflecting the shifts in candidate popularity and engagement as observed through social media interactions. This method offers a novel approach to predicting electoral outcomes based on digital engagement metrics.

From the results, we observe that the engagement towards Peter Pellegrini was boosted more than towards Ivan Korčok in comparison to the first election round. When we applied the results from the first round to predict second and normalized it to include only the two candidates, the results were: Ivan Korčok at 50.63% and Peter Pellegrini at 49.37%. Although these estimates did not exactly match the actual election outcome, this variance could be attributed to the non-representative sample of voters, as Instagram is more popular among younger users. This imbalance in age groups could have affected the engagement data and how well it represented overall voter opinions.

One observation from this analysis was the significant difference in the ratio of likes to comments between the two candidates. For Ivan Korčok, the likes to comments ratio was approximately 50:1 in both rounds, indicating that for every comment, there were about 50 likes. In contrast, for Peter Pellegrini, the ratio was about 8:1, meaning he received roughly 8 likes for every comment. This difference highlights the varying engagement dynamics between the two candidates, with Korčok's posts attracting a higher proportion of likes relative to comments compared to Pellegrini's.

Chapter 4

Discussion

This chapter delves into the implications, challenges, and insights gathered throughout this thesis. It aims to reflect on the practical applications of the methodologies employed, particularly focusing on the use of Instaloader [18] for data collection from Instagram [6], and to discuss the broader implications of the findings.

4.1 Overview of the Data Collection Process

Using Instaloader, we were able to efficiently scrape valuable data from Instagram, which included posts, likes, and comments. The integration and implementation of this tool were crucial for the project, enabling us to construct a database for analysis.

4.1.1 Setbacks and Limitations

Throughout the project, several setbacks and limitations surfaced. One significant challenge was the requirement to authenticate with a genuine user account to access comments. This authentication step added complexity and occasionally resulted in slow retrieval speeds due to the reliance on scraping techniques.

Additionally, Instagram's strict policies often required additional verification measures, such as linking a phone number to the account (see Figure 4.1) or addressing automated behaviour suspicions (see Figure 4.2).

We encountered specific Instaloader-related challenges, such as temporary wait times between requests due to excessive queries. For example, we often received messages indicating the need to wait for certain periods, which disrupted the continuous data retrieval process:

> "Too many queries in the last time. Need to wait 22 minutes, until 19:44."

However, despite the inconvenience, these temporary bans served as a necessary precautionary measure. They acted as a buffer, preventing us from potentially triggering

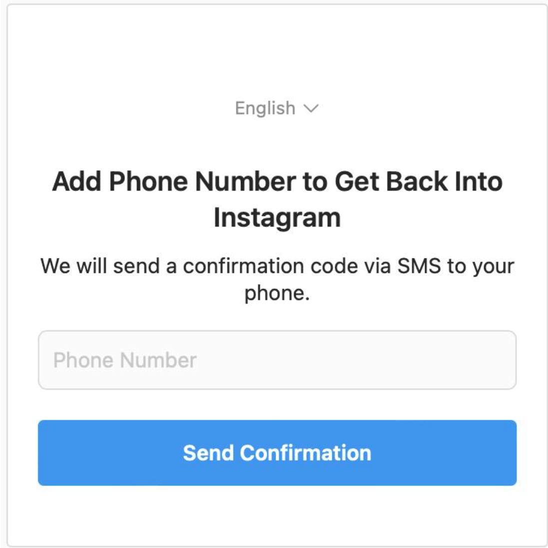

Figure 4.1: Prompt displayed on Instagram, encouraging users to add a phone number to their accounts.

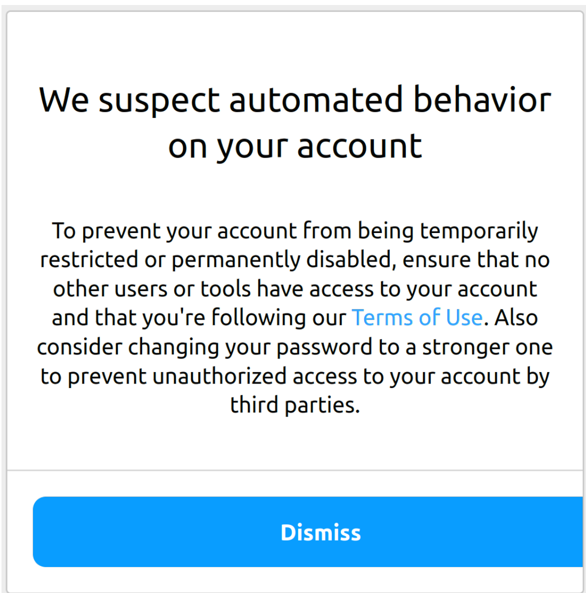

Figure 4.2: Notification from Instagram alerting users of automated behaviour detected on the platform.

more severe actions from Instagram, such as account suspension or more significant disruptions to our data collection efforts.

These functionalities significantly contributed to enhancing the efficiency and stability of our scraping process. By minimizing the risk of detection and potential penalties, they ensured smoother and more reliable data collection overall.

4.1.2 Workarounds and Solutions

In response to the encountered challenges, we implemented a range of workarounds and solutions to ensure the continuity of our scraping efforts.

For simpler issues, such as occasional login problems or dismissible prompts, we quickly resolved them by re-logging into accounts or dismissing the prompts as needed.

However, for more complex problems that persisted, we had to devise strategic adjustments. One such solution involved utilizing a larger pool of dedicated Instagram accounts specifically for scraping purposes. This approach helped distribute the scraping load across multiple accounts, reducing the risk of triggering detection mechanisms due to excessive queries.

However, expanding this pool of accounts posed its own set of challenges. Creating new accounts often required phone number verification, a process strictly monitored by Instagram to prevent automated behaviour. This verification process was particularly challenging as it also involved creating new email addresses, which are crucial for setting up new Instagram accounts. As a result, we had to carefully navigate this verification process to ensure compliance while scaling up our scraping operations.

4.2 Age Distribution on Instagram and Facebook

When examining the dynamics of the 2024 Slovak presidential election, using Facebook data could enhance our analytical accuracy. Nonetheless, it is crucial to acknowledge that relying only on Facebook data might present limitations in capturing a full representation of the voters. While Facebook demographics can offer information about the preferences and behaviours of specific segments of the electorate, they may not accurately represent individuals who are not active on the platform or who misrepresent themselves on social media. Thus, while integrating Facebook data into our analysis can be beneficial, it is imperative to supplement it with other data sources to ensure a more holistic understanding of voter demographics and preferences.

As illustrated in Figure 4.3, the more evenly distributed age groups on Facebook compared to Instagram suggest a broader reach. This implies that Facebook may offer a more general snapshot of voter characteristics across various age brackets.

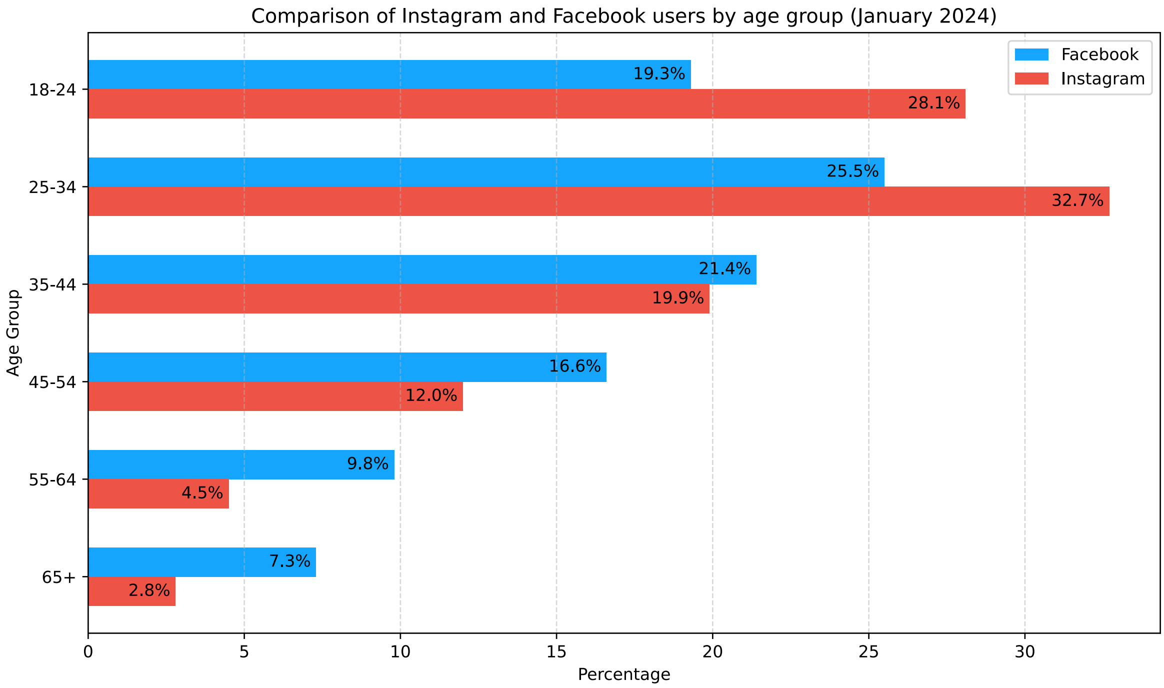

Figure 4.3: Comparison of Instagram and Facebook users by age group [34, 35].

Conclusion

In this thesis, we focused on the exploration of Instagram and its users, especially in Slovakia, during the 2024 presidential election. Our main focus was to analyse user engagement, behaviour, and sentiment expressed towards candidates. With a combination of data scraping, sentiment analysis and data visualization techniques, our aim was to discover insights of public opinion, views on candidates, and sentiment trends during the elections.

By examining Instagram data, we gained perspectives on how individuals engage with political content, express their sentiments, and interact with each other on these topics. These points provide a perspective of the social and political environment during the election period.

A key contribution of this thesis is the application of sentiment analysis to Instagram comments. Using SlovakBERT model, we were able to classify user sentiments accurately, looking into the sentiment trends expressed towards candidates. This approach not only improves our grasp of public sentiment dynamics but also provides a tool for monitoring and analysis of public opinion.

Furthermore, the implementation of data visualization played a crucial role in interpreting the entire dataset, allowing a clearer understanding. Visualizations such as histograms, Venn diagrams, and time-series plots helped us in presenting the findings of our analysis in an intuitive and accessible way.

We analysed user behaviour over time and estimated the second round results based on changes from the first round. Notably, Ivan Korčok received considerably more likes than Peter Pellegrini, indicating a shift in support. Furthermore, sentiment analysis revealed that during the second round, Pellegrini's comments were evenly split between positive and negative, while Korčok had twice as many positive comments as negative ones, signaling a positive sentiment trend towards him.

While this thesis presents valuable findings regarding the dynamics of social media engagement during the 2024 Slovak presidential election, it is important to acknowledge its limitations and identify possibilities for future research. The scope of this study was limited to Instagram, which is more commonly used by younger generations. Future studies could benefit from including other social media platforms for a more extensive analysis. Moreover, using different machine learning techniques and language models could enhance the accuracy and predictive capabilities, allowing for the integration of a broader spectrum of expressions beyond the typical positive, negative, and neutral sentiments.

Bibliography

[1] Stanley Wasserman and Katherine Faust. Social network analysis: Methods and applications. 1994.

[2] Linton Freeman et al. The development of social network analysis. *A Study in the Sociology of Science*, 1(687):159–167, 2004.

[3] Facebook. Facebook, 2004. Prominent social media platform for connecting users globally.

[4] Robert E Wilson, Samuel D Gosling, and Lindsay T Graham. A review of facebook research in the social sciences. *Perspectives on psychological science*, 7(3):203–220, 2012.

[5] DataReportal. Essential facebook stats. `https://datareportal.com/essential-facebook-stats`, 2024. Accessed: June 5, 2024.

[6] Instagram. Instagram, 2010. Social media platform for sharing photos and videos.

[7] POSTING ON Instagram. Instagram. *Online unter: https://www. wissenschaftskommunikation. de/format/instagram*, 2016.

[8] DataReportal. Essential instagram stats. `https://datareportal.com/essential-instagram-stats`, 2024. Accessed: June 5, 2024.

[9] John Adrian Bondy, Uppaluri Siva Ramachandra Murty, et al. *Graph theory with applications*, volume 290. Macmillan London, 1976.

[10] Chris Godsil and Gordon F Royle. *Algebraic graph theory*, volume 207. Springer Science & Business Media, 2001.

[11] Dmitri Goldenberg. Social network analysis: From graph theory to applications with python. *arXiv preprint arXiv:2102.10014*, 2021.

[12] Bo Zhao. Web scraping. *Encyclopedia of big data*, 1, 2017.

[13] Moaiad Ahmad Khder. Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, 13(3), 2021.

[14] Ismael Camargo-Henríquez and Yarisel Núñez-Bernal. A web scraping based approach for data research through social media: An instagram case. In *2022 V Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC)*, pages 1–4. IEEE, 2022.

[15] Kevin Zúñiga et al. Facebook scraper. `https://github.com/kevinzg/facebook-scraper`.

[16] Apify. Apify. `https://apify.com/`.

[17] Apify. Apify instagram scraper. `https://apify.com/apify/instagram-scraper`.

[18] Alexander Graf et al. Instaloader. `https://github.com/instaloader/instaloader`.

[19] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113, 2014.

[20] Bing Liu et al. Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2(2010):627–666, 2010.

[21] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[22] Ian Tenney, Dipanjan Das, and Ellie Pavlick. Bert rediscovers the classical nlp pipeline. *arXiv preprint arXiv:1905.05950*, 2019.

[23] Matúš Pikuliak, Štefan Grivalský, Martin Konôpka, Miroslav Blšták, Martin Tamajka, Viktor Bachratý, Marian Simko, Pavol Balážik, Michal Trnka, and Filip Uhlárik. SlovakBERT: Slovak masked language model. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 7156–7168, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.

[24] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[25] Robert M Bond, Christopher J Fariss, Jason J Jones, Adam DI Kramer, Cameron Marlow, Jaime E Settle, and James H Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298, 2012.

[26] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–236, 2017.

[27] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. The spread of low-credibility content by social bots. *Nature communications*, 9(1):1–9, 2018.

[28] Andreas M Kaplan and Michael Haenlein. Users of the world, unite! the challenges and opportunities of social media. *Business horizons*, 53(1):59–68, 2010.

[29] Gunn Enli and Hallvard Moe. Introduction to special issue: Social media and election campaigns–key tendencies and ways forward. *Information, communication & society*, 16(5):637–645, 2013.

[30] Andrew Chadwick and Jennifer Stromer-Galley. Digital media, power, and democracy in parties and election campaigns: Party decline or party renewal?, 2016.

[31] Koosha Zarei, Reza Farahbakhsh, Noel Crespi, and Gareth Tyson. A first instagram dataset on covid-19. *arXiv preprint arXiv:2004.12226*, 2020.

[32] Matteo Cinelli, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. The covid-19 social media infodemic. *Scientific reports*, 10(1):1–10, 2020.

[33] Junling Gao, Pinpin Zheng, Yingnan Jia, Hao Chen, Yimeng Mao, Suhong Chen, Yi Wang, Hua Fu, and Junming Dai. Mental health problems and social media exposure during covid-19 outbreak. *Plos one*, 15(4):e0231924, 2020.

[34] NapoleonCat. Distribution of instagram users in slovakia as of january 2024, by age group. https://www.statista.com/statistics/1024807/instagram-users-slovakia/, February 2 2024.

[35] NapoleonCat. Distribution of facebook users in slovakia as of january 2024, by age group. https://www.statista.com/statistics/1029958/facebook-users-slovakia/, February 2 2024.