Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

[Analysis of Social Networks]

Bachelor Thesis

2024
Matej Zelenák

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

[Analysis of Social Networks]

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

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta:
Študijný program:

Študijný odbor:
Typ záverečnej práce:
Jazyk záverečnej práce:
Sekundárny jazyk:

Názov:

Anotácia:

Vedúci:
Katedra:
Vedúci katedry:

Dátum zadania:

Dátum schválenia:

garant študijného programu

....

študent vedúci práce

Comenius University Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:
Study programme:

Field of Study:
Type of Thesis:
Language of Thesis:
Secondary language:

Title:

Annotation:

Supervisor:
Department:
Head of department:

Assigned:

Approved:

Guarantor of Study Programme

.... Student

.... Supervisor

Acknowledgments: Tu môžete poďakovať školiteľovi, prípadne ďalším osobám, ktoré vám s prácou nejako pomohli, poradili, poskytli dáta a podobne.

Abstrakt

Slovenský abstrakt v rozsahu 100-500 slov, jeden odstavec. Abstrakt stručne sumarizuje výsledky práce. Mal by byť pochopiteľný pre bežného informatika. Nemal by teda využívať skratky, termíny alebo označenie zavedené v práci, okrem tých, ktoré sú všeobecne známe.

Kľúčové slová: jedno, druhé, tretie (prípadne štvrté, piate)

Abstract

Abstract in the English language (translation of the abstract in the Slovak language).

Keywords:

Contents

List of Figures

List of Tables

Chapter 1

Literature Review

1.1 Social Networks

Social networks [1, 2] are the intricate patterns of social connections that form among individuals, groups, organizations, or other entities within a social system. These connections are established through various interactions such as communication, friendship, collaboration, or resource exchange.

In essence, social networks serve as the underlying framework through which people interact, share information, and collaborate with one another. They encompass both strong ties, such as close relationships with family and friends, and weak ties, such as casual acquaintances or connections through professional networks.

Social networks play a fundamental role in shaping individual behavior, facilitating the spread of information and ideas, and influencing social dynamics within communities and societies. They provide opportunities for social support, resource mobilization, and collective action, while also influencing individuals' access to resources, opportunities, and social capital.

Understanding the structure and dynamics of social networks is crucial for comprehending how social systems function and evolve over time. By studying the patterns of connections between individuals and groups within a network, researchers can gain insights into the mechanisms driving social interactions, the formation of social norms, and the emergence of collective behavior.

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 behavior, communication, and interaction.

[TODO]make ref 1.1

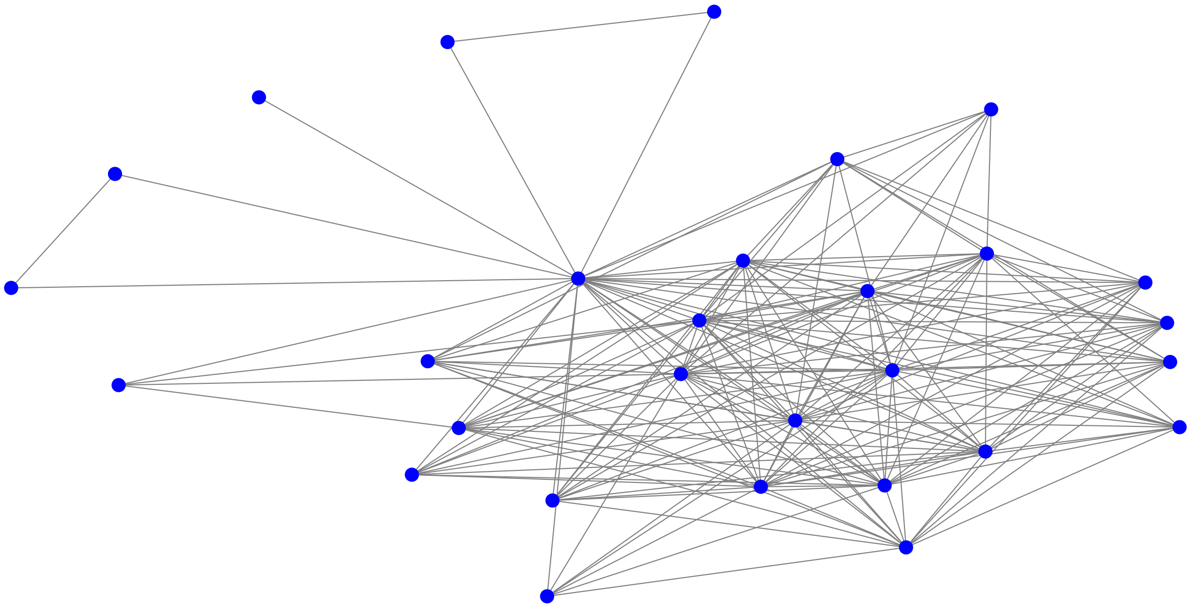

Figure 1.1: Example of connected actors in social network

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. Key features like groups facilitate connections among users with shared interests; events coordinate gatherings and activities, significantly enhancing interaction among individuals with similar interests and opinions.

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 discourse.

Despite its widespread popularity, Facebook has frequently faced challenges such as privacy concerns, data breaches, and distibution misinformation. However, despite these issues, it continues to maintain its position as the most widely utilized platform by billions of users worldwide. These challenges highlight the importance of ongoing efforts to address and mitigate the spread of misinformation and enhance user privacy and security on the platform.

In summary, Facebook represents more than just a social networking site; it's a dynamic platform that shapes how people connect, communicate, and collaborate online, while also serving as a fertile ground for academic inquiry and research.

Figure 1.2: Example of post on Instagram with likes, comments
`https://www.instagram.com/zuzana_caputova/`

1.1.2 Instagram

With a user base exceeding one billion globally, Instagram [5, 6] 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, IGTV (for longer videos), 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. It is compatible with Apple iOS devices and Android smartphones and tablets. See example of post shown in Figure 1.2

Users can follow others and be followed, although reciprocity isn't required like on Facebook. Users can follow accounts privately, controlling who can view their posts. However, by default, content is visible to anyone unless the account is set to private. Users maintain control over their privacy settings, with the option to make their account private and approve followers.

Posting on Instagram involves uploading photos or videos. Users can add captions, locations, tag people, and choose whether to share content exclusively with their followers or extend it beyond the app.

1.2 Scraping

As shown in Figure 1.3 Web scraping [7, 8, 9] is a data extraction technique employed to retrieve information from websites across the internet. It encompasses 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 underlying HTML code of web pages, extracting relevant data elements, and transforming them into a structured format for analysis or further processing.

The process begins with accessing web pages using software tools or programming languages capable of fetching the HTML content. These tools facilitate the retrieval of web page data, providing a foundation for subsequent parsing and extraction tasks. Once the HTML content is obtained, the parsing stage comes into play, where the structure of the HTML document is analyzed 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 and analyze, making it simpler to extract valuable insights.

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

1.2.1 Facebook Scraper

Facebook Scraper [10] 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 behavior 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 page and the type of data to extract. It also includes features for pagination, filtering, and sorting the retrieved data, allowing users to customize their scraping operations according to their specific requirements.

However, it's 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 exercise caution and ensure compliance with Facebook's terms of service and relevant laws and regulations when using the library for

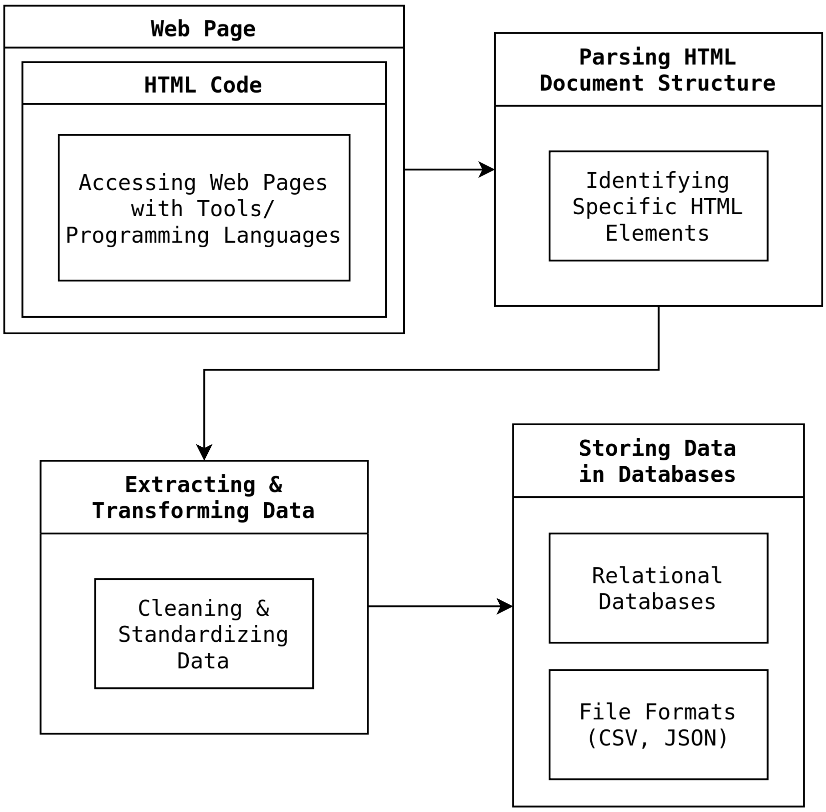

Figure 1.3: Process of Scraping

data extraction purposes.

1.2.2 Instaloader

Instaloader, developed by Alexander Graf [11], 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 primary features of Instaloader is its ability to fetch various types of data from Instagram profiles, including posts, stories, IGTV videos, 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 boasts versatility, 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 analyze Instagram content offline. Additionally, Instaloader supports advanced functionalities such as profile crawling, enabling users to traverse 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 analyzing Instagram data in a flexible and efficient manner. Its wide range of functionalities makes it a popular choice among researchers, data scientists, journalists, and social media enthusiasts seeking to gain insights into Instagram's vast ecosystem of content and interactions.

1.3 Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a process in natural language processing (NLP) [12] 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 analyzing the words, phrases, and context within the text to infer 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 such as BERT 1.3.1.

1.3.1 BERT

Crafted by Google, this NLP model is tailored to process and grasp human language in a manner more aligned with human cognition. BERT [13, 14] (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 analyze 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

SlovakBERT

A language model tailored specifically for Slovak, SlovakBERT [15] 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 [16] 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 tackle the absence of established evaluation standards for the Slovak language by devising their own set of evaluation 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 analyzes the text to discern whether the sentiment expressed is positive, negative, or neutral. By fine-tuning SlovakBERT on sentiment-related tasks and datasets, the model learns to identify nuanced expressions of sentiment, capturing subtle nuances 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 analyzing sentiment in social media comments.

1.4 Related Work

[TODO](some thesis/articles about Instagram scraping)

Chapter 2

Methods

In this chapter, we outline the methodology used in the bachelor's thesis, detailing the processes of data collection, storage, and sentiment analysis utilizing SlovakBERT [15]. We explain the steps involved in gathering relevant data, setting up an organized storage system, and using NLP to understand sentiment. Our goal is to provide a straightforward overview of our research process, focusing on the methods we used. Detailed findings will be discussed in the next chapter.

2.1 Data Collection

For data collection, we primarily relied on Instaloader, a tool that enabled us to extract comments and posts from Instagram efficiently. While initially considering Facebook Scraper for gathering Facebook data, we encountered limitations that rendered it ineffective for our purposes. As a result, we decided to focus solely on Instaloader for our data collection needs.

2.1.1 Facebook Scraper

Exploring the utilization of a Facebook Scraper [10] as part of our methodology aimed to gather valuable user data from the platform. However, our experience with this tool introduced unexpected challenges.

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

Moreover, the collection of user's friends proved to be significantly slow, further complicating the process. We aimed to construct a Facebook Friends list visualization, where inserting one user would display a graph of all connected nodes (users), 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.

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

2.1.2 Instaloader

Instaloader [11] played a key role in our project by enabling us to gather Instagram data efficiently.

Functionality and Features

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, though it may skip some irrelevant ones without user intervention. 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 boasts versatility, offering both command-line functionality with numerous customizable flags and a programming interface for precise control and integration within larger applications.

Integration and Implementation

During the integration and implementation phase, our primary objective was to utilize Instaloader for gathering Instagram data, storing it locally, and subsequently analyzing it. Given the prolonged duration of the scraping process, which involved retrieving data from specific dates, we implemented multiple checkpoints to ensure efficient data retrieval, preventing redundant scraping of the same content and enabling us to retrieve posts within the specified time frame efficiently. This process involved storing certain Instagram posts and comments in a database, crucial for laying the groundwork for our analysis. Here's a snippet of code illustrating how to scrape posts and comments of posts from one user

Algorithm 2.1: Scraping posts and comments of Instagram user 'zuzana_caputova'

```
import instaloader

loader = instaloader.Instaloader()
loader.login(username, password)

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 facilitated the extraction of valuable insights from the data, ensuring comprehensive analysis and interpretation in the future.

2.2 Data Organization

In the context of our bachelor's thesis, effective data organization played a pivotal role in facilitating our analysis of Instagram data collected during the 2024 Slovakia elections. This section outlines our methodology for organizing and managing the data.

2.2.1 Storing Data in an SQLite Database

To handle the large amount of Instagram data we collected, we used an SQLite database system (as shown in Figure 2.1). This system allowed us to organize the data in a structured manner, making it easier to manage and analyze.

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've 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 posts. Similarly, the

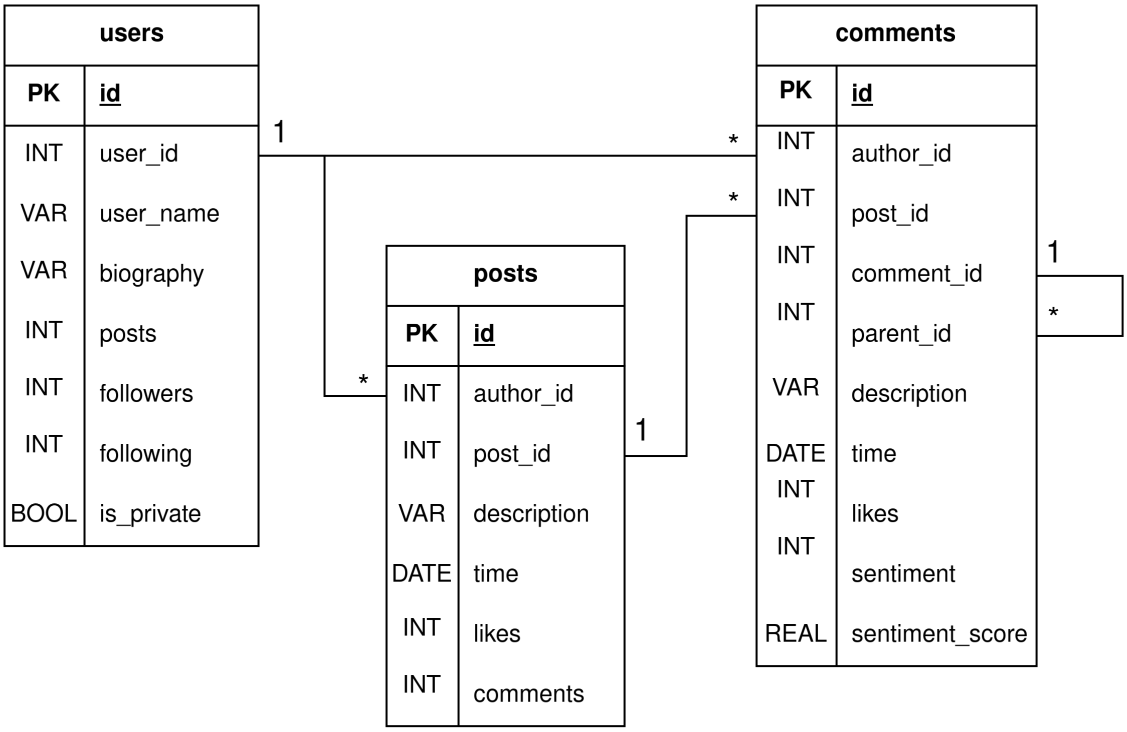

Figure 2.1: Diagram illustrating the structure of the SQLite database used to organize Instagram data.

'comments' table was designed to hold comments (description) made on individual posts, along with information like the usernames 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. If the 'parent_id' contains the ID of another comment, it indicates that the comment is a reply to that specific comment.

This careful organization of tables ensured that we could efficiently store and retrieve Instagram data for our analysis. Additionally, we implemented measures in the database to prevent duplicate entries. This meant that if a post or comment was scraped multiple times, it would only be stored once in the database, avoiding unnecessary redundancy and ensuring data accuracy.

2.3 Sentiment Analysis

Sentiment analysis plays a crucial role in understanding the public perception and attitudes expressed in social media data. In this section, we detail our approach to sentiment analysis using a classifier based on SlovakBERT.

2.3.1 Sentiment Analysis Model

The sentiment analysis classifier we employed is based on SlovakBERT, a language model specifically tailored for the Slovak language. This model is capable of distinguishing three levels of sentiment:

-1 → Negative sentiment

0 → Neutral sentiment

1 → Positive sentiment

By utilizing this model, we aimed to assign sentiment scores to each comment in our database, providing insights into the overall sentiment expressed within the Instagram data collected during the 2024 Slovakia elections.

2.3.2 Implementation

To implement sentiment analysis on our Instagram data, we utilized the SlovakBERT-based sentiment analysis classifier. This involved processing each comment in our database through the classifier to generate sentiment scores.

Scoring Comments

Below is an example of how the SlovakBERT-based sentiment analysis classifier can be used to score comments:

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

>>> pipe('Som veľmi sklamaný z vašej politiky.')
[{'label': '-1', 'score': 0.9978156089782715}]

>>> pipe('Čakám na nové informácie')
[{'label': '0', 'score': 0.889336347579956}]

>>> pipe('Skvelá práca! Ďakujem za úžasný obsah!"')
[{'label': '1', 'score': 0.9964995384216309}]
```

This example demonstrates the process of scoring comments using the sentiment analysis model. Each comment is passed through the model, which assigns a sentiment score (-1, 0, or 1) based on the sentiment expressed in the comment. The sentiment score is then 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 Slovakia elections, allowing for a more comprehensive analysis of public opinion and attitudes.

Chapter 3

Analysis

In this chapter, we delve into the analysis of various aspects of Instagram posts related to our candidates. We explore the relationship between likes and comments, analyze commenter behavior, and examine the sentiment expressed in comments.

3.1 Likes and Comments

In this section, we will be examining the relationship between likes and comments on Instagram posts over time. We aim to understand how engagement changes over the weeks and if there are any patterns in user interaction with posts.

3.1.1 Trend Analysis

To start, we will plot the number of likes and comments per week. By visualizing this data, we can easily identify any trends or fluctuations in engagement levels over time. This analysis will provide us with valuable insights into the dynamics of user interaction on Instagram and how it evolves week by week.

3.1.2 Peak Analysis

Once we have plotted the data, we will closely examine the trends in likes and comments over the weeks. Are there certain weeks where posts tend to receive more likes or comments? Are there any notable spikes or dips in engagement that coincide with specific events or campaigns? These are the questions we aim to answer through our analysis.

3.1.3 Event Analysis

Furthermore, we will conduct an analysis of engagement around specific events or campaigns. By identifying peaks or anomalies in engagement levels during these periods, we can gain insights into the impact of external factors on user interaction.

Overall, this analysis will provide us with valuable insights into the relationship between likes and comments on Instagram posts over time. By understanding how engagement levels fluctuate and evolve, we can gain a deeper understanding of user behavior and optimize our content strategy accordingly.

3.1.4 Popularity

In the "Popularity" subsection, we combine the total number of likes and comments for each post into a single metric to assess overall post popularity. By summing the likes and comments, we can create a holistic measure of user interaction with each post. We will then plot this combined metric over time to identify trends in post popularity and assess which posts resonate most with the audience. This analysis will help us understand the overall impact of our Instagram content and identify strategies for maximizing engagement.

3.2 Commenters

In this section, we will analyze the behavior of commenters on Instagram posts related to our candidates since we have collected data on all the comments made on these posts, including information on who made each specific comment.

3.2.1 Total Number of Commenters

To begin, we will identify the total number of unique commenters across all posts. This will give us an overview of the size of the commenting audience for our candidates' posts.

3.2.2 Overlap Analysis

Next, we will create Venn diagrams to visualize the overlap between commenters on posts related to different candidates. By comparing the sets of commenters for each candidate, we can identify how many commenters are engaging with multiple candidates' posts. This analysis will provide insights into the level of cross-engagement between different candidates' audiences on Instagram.

3.2.3 Top Commenters Analysis

Furthermore, we will analyze the engagement patterns of the top commenters. We will identify the users who have commented the most frequently across all posts and examine their commenting behavior. Are there certain users who consistently engage with posts from multiple candidates? What types of comments do they typically leave? Understanding the behavior of these top commenters can provide valuable insights into the dynamics of user engagement on Instagram.

Overall, this analysis of commenters will help us understand how users interact with posts related to our candidates on Instagram. By examining the overlap between commenters and identifying patterns in their behavior, we can gain a deeper understanding of audience engagement and tailor our content strategy accordingly.

3.3 Comment Sentiment

In this section, we analyze the sentiment of each comment using the SlovakBERT [15] pipeline with the "text-classification" model "kinit/slovakbert-sentiment-twitter". This pipeline allows us to classify the sentiment of each comment as either positive, negative, or neutral.

3.3.1 Sentiment Analysis Methodology

We use the SlovakBERT pipeline to conduct sentiment analysis on every comment. Specifically, we apply the pre-trained "text-classification" model "kinit/slovakbert-sentiment-twitter" to classify the sentiment of the processed comment text.

3.3.2 Sentiment Label Calculation

For each comment, we calculate a sentiment label indicating the strength of the sentiment expressed. The sentiment label is determined based on the output of the SlovakBERT model, which assigns a numerical value to represent the sentiment of the comment. A positive sentiment label indicates a positive sentiment, while a negative label indicates a negative sentiment.

3.3.3 Sentiment Score

In addition to the sentiment label, we store the sentiment score, which represents the confidence level of the sentiment classification. The sentiment score is a measure of the model's certainty in its sentiment prediction for each comment. A higher score indicates greater confidence in the accuracy of the sentiment classification, while a lower score suggests more uncertainty.

3.3.4 Visualization

To visualize the sentiment analysis results, we present graphs showing the distribution of sentiment label and scores across all comments. These graphs provide an overview of the sentiment distribution within our comments, allowing us to identify any trends or patterns in sentiment expression. By visually inspecting the distribution of sentiment labels and scores, we can gain insights into the overall sentiment dynamics of the comments and understand how users perceive the content on Instagram.

Overall, this analysis of comment sentiment using SlovakBERT allows us to gain insights into the overall sentiment of user comments and understand the sentiment dynamics.

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

[6] POSTING ON Instagram. Instagram. *Online unter: https://www. wissenschaft-skommunikation. de/format/instagram*, 2016.

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

[8] 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.

[9] 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.

[10] kevinzg. Facebook scraper. `https://github.com/kevinzg/facebook-scraper`, 2024.

[11] aandergr. Instaloader. `https://github.com/instaloader/instaloader`, 2024.

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

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

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

[15] 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.

[16] 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.