https://doi.org/10.2352/EI.2025.37.3.MOBMU-327

OSINT Investigation of Social Media for Classifying Reactions to Cannabis Legalization

Berfin Ebrar Atabey[1], *Resul Bedii Gümüş*[1], *Franziska Schwarz*[2], *Klaus Schwarz*[2,4], *Reiner Creutzburg*[2,3]

[1] Ss. Cyril and Methodius University of Skopje, Boulevard Goce Delchev 9, Skopje 1000, North Macedonia
[2] SRH University, School of Technology and Architecture, Sonnenallee 221c, D-12059 Berlin, Germany
[3] Technische Hochschule Brandenburg, Department of Informatics and Media, Magdeburger Str. 50, D-14770 Brandenburg, Germany
[4] University of Granada, Faculty of Economics and Business, P.◦ de Cartuja, 7, ES-18011 Granada, Spain
Email: eb.atabey@gmail.com, r.bedii.gumus@gmail.com, klaus.schwarz@srh.de, f.schwarz@posteo.de, reiner.creutzburg@srh.de, creutzburg@th-brandenburg.de

Keywords: *Open Source Intelligence (OSINT), Zero-Shot Classification, Cannabis Legalization, Social Media Analysis, Sentiment Analysis, Natural Language Processing (NLP)*

Abstract

This study employs Open-Source Intelligence (OSINT) techniques to analyze social media sentiment regarding cannabis legalization in Germany on April 1, 2024, focusing specifically on Reddit content. The methodology involves data collection, cleaning, clustering, and sentiment analysis using natural language processing. This research utilizes sentence transformer models for text embedding and K-means clustering algorithms to identify distinct perspectives on legalization. Sentiment analysis was conducted using a pretrained BERT-based model developed by Hugging Face. The results revealed seven distinct clusters representing different themes and sentiments regarding cannabis legalization. Public opinion tended to be positive, indicating general support for cannabis legalization despite some expectations. This research demonstrates the value of OSINT in social media analysis while providing valuable insights for researchers and policymakers regarding public sentiment on cannabis legalization.

Introduction

The legalization or decriminalization of cannabis has emerged as a significant global sociopolitical issue, evoking a spectrum of reactions that vary by location, culture, and demographics [1]. In recent years, the debate on cannabis legalization has become more controversial [2], particularly in Germany, where the law on cannabis use changed on April 1, 2024, [3]. Social media platforms such as Reddit, Twitter, and Facebook have become the most important platforms for the public to share their opinions. This has made the diverse sentiments and ideas of the public easily accessible [4]. Open-Source Intelligence (OSINT) techniques have transformed data collection and analysis from digital platforms.

Open-Source Intelligence (OSINT) techniques have transformed data collection and analysis from digital platforms. Researchers can gain valuable insight into public opinion and societal trends by leveraging extensive user-generated online content. OSINT provides powerful analytical tools for examining how different demographic groups respond to proposed legislation and policy changes [5].

This study applies OSINT tools to address two key research questions:

1. What frameworks shape the debate around cannabis legalization in Germany?
2. What is the public sentiment towards cannabis legalization within these identified frameworks?

This research classifies and examines public reactions to this polarizing issue by analyzing Reddit comments. The methodology systematically incorporates data collection, cleaning, clustering, and categorization to analyze social media opinions comprehensively.

This study began by exploring the current landscape of cannabis legalization and its social implications. Then, it reviews the existing literature on OSINT applications in similar research contexts. The following sections detail the data collection and preparation processes, analytical methodology, and results. This study contributes to the broader understanding of public opinion on cannabis legalization and demonstrates the efficacy of OSINT techniques in social media research.

Research Questions

- Under which frameworks is cannabis legalization in Germany debated?
- What is the public sentiment towards cannabis legalization within these identified frameworks?

Literature Review

The legalization of cannabis has been a debatable issue, creating diverse reactions in different parts of society [1]. With the emergence of social media, platforms such as X, Facebook, and Instagram have become important venues for the public to express opinions [4]. OSINT techniques provide powerful tools for gathering and analyzing data from these platforms, enabling researchers to observe and interpret public reactions [6].

This literature review examines the application of OSINT in analyzing social media responses to cannabis legalization. Through an assessment of existing research, this review identifies the methodologies employed, data types collected, and key findings concerning public opinion on cannabis legalization.

The scope of this review encompasses recent peer-reviewed articles. The selected literature comprised studies that employed various OSINT techniques. This review contributes to the body of knowledge on social media analysis, OSINT applications, and public policy research.

Motlagh et al. [11] analyzed Twitter data to predict public opinion on marijuana legalization and track consumption trends across U.S. states. Their methodology combined lexicon-based filtering with an ontology-driven approach, analyzing over 300,000 relevant tweets collected four months before and after the November 2015 Ohio Marijuana Legalisation ballot. The results revealed two key insights: states with recreational marijuana legalization expressed significantly more positive sentiments (67% positive) than states with medical-only legalization or no legalization (23% positive), and states with higher percentages of positive sentiments were more likely to expand marijuana legalization. Additionally, their analysis of marijuana consumption trends identified cannabis (82%) as the most referenced form, followed by marijuana concentrates (73%), demonstrating that social media analysis can serve as a reliable alternative to traditional polling to understand public opinion on drug legalization.

Li et al. [12] investigated the differential impacts of verified, regular, and suspended Twitter users on cannabis-related public discourse. Their methodology integrated qualitative thematic analysis and quantitative Latent Dirichlet Allocation techniques for content examination, complemented by sentiment analysis, to evaluate message tone. The results demonstrated that, while all user categories addressed similar cannabis-related topics, their emphasis varied significantly: verified users predominantly focused on legalization issues, suspended users concentrated on promotional content, and regular users discussed a broader range of everyday cannabis-related concerns. The study revealed that suspended users exhibited the highest positive sentiment scores (0.0594), followed by verified users (0.0570), and regular users (0.0402), suggesting different motivations across user groups despite their shared interest in cannabis legalization and health benefits.

Unlu et al. [13] conducted a mixed-methods analysis of Twitter discussions surrounding the Finnish Green Party's 2021 cannabis legalization proposal. Their methodology combines quantitative computational techniques (sentiment analysis and topic modeling via Latent Dirichlet Allocation) with qualitative thematic content analysis to examine over 20,000 cannabis-related tweets. The results show a dramatic spike in Twitter activity following the proposal, with daily tweets increasing from approximately 140 to a peak of 6,600. While sentiment analysis indicated negative attitudes toward cannabis overall, sentiment scores shifted temporarily positive during peak discussion periods. Topic modeling revealed 15 distinct discussion themes, with comparisons between cannabis and alcohol being the most prevalent, followed by policy implications, youth concerns, and health effects. The qualitative analysis found that the most retweeted and engaged-with content came predominantly from public figures opposed to legalization, including politicians, healthcare professionals, and journalists, who emphasized mental health risks. Pro-legalization arguments centered on harm reduction, regulatory benefits, and criminal justice concerns. The researchers noted a potential generational divide, with individual Twitter users showing more positive sentiments than influential figures, suggesting evolving public attitudes despite institutional resistance to cannabis policy reform in Finland.

Rychert et al. [14] examined digital media coverage during New Zealand's 2020 cannabis legalization referendum. Their analysis of six news providers and Facebook campaigns showed digital news was slightly pro-reform (+0.4 on a -2 to +2 scale), with pro-legalization articles receiving better placement, more redistribution, and higher social media engagement. The pro-legalization campaign invested four times more in Facebook advertising than its opponents, dominating the digital space, while anti-legalization efforts concentrated on traditional media. Network analysis of Facebook posts revealed interconnections between digital news outlets and advocacy groups, demonstrating that combining sentiment and network analysis provides valuable insights into referendum debates.

Lim et al. [15] examined the trends and socio-demographic differences in cannabis vaping across the USA and Canada using data from the International Cannabis Policy Study. This study analyzed the prevalence of three distinct cannabis vaping products (dried flower, cannabis oil/liquid, and concentrates) across jurisdictions with different legal statuses. The results showed that in 2019, cannabis oil was the most commonly vaped product across all jurisdictions (8.1% in Canada, 13.7% in U.S. illegal jurisdictions, and 17.4% in U.S. legal jurisdictions), followed by dried flowers and concentrates. The prevalence of all cannabis vaping forms was highest in U.S. jurisdictions with legalized non-medical use, and lowest in Canada. Between 2018-2019, vaping dried flowers decreased in U.S. legal jurisdictions and Canada, whereas the vaping of cannabis oil and concentrates increased significantly across all jurisdictions. Demographically, younger respondents (16-55 years), males, those with some college education, and individuals perceiving daily cannabis vaping as low risk demonstrated significantly higher odds of vaping all cannabis product forms. These findings highlight the impact of legalization on consumption patterns and underscore the importance of collecting detailed cannabis use data to understand the evolving consumption trends and their public health implications.

Mann et al. [16] utilized Twitter data to analyze public reactions to the U.S. House vote on cannabis decriminalization at the federal level. Their methodology employed sentiment analysis, hashtag tracking, and thematic content analysis to examine the social media discourse surrounding the legislative event. The results revealed predominantly positive sentiments toward decriminalization among Twitter users from diverse profile backgrounds, with conversations organized around five key topics: commentary on the House vote itself, concerns about Senate impediments to passage, support for expungement of marijuana-related criminal records, discussions of medical marijuana applications, and debate about the bills' potential social and economic impacts. The researchers extended their analysis to discuss the ethical and regulatory implications of cannabis retailing and marketing practices. This study demonstrates how social media analysis can effectively capture public sentiment and discourse themes surrounding significant cannabis policy developments, thereby providing insights for both policymakers and industry stakeholders operating in this evolving regulatory landscape.

Data Collection and Preparation

Across all social media platforms, Reddit was chosen as the dataset because it is widely used and allows us to capture a significant portion of public opinion [7]. Reddit serves as an efficient platform for extracting of substantial volumes of data at no financial cost, rendering it an optimal source for analytical purposes. Additionally, Reddit allows a larger text space than other social media platforms, enabling users to express their thoughts and opinions in greater detail, enriching the quality and depth of the analysis [8].

Data Selection Criteria

The keywords "Germany," "cannabis," and "legalization" were used to identify relevant comments in the Reddit search engine. The first 20 posts in the results offered on July 7, 2024, were chosen by filtering the most relevant results.

Data Acquisition

To collect the data, the Python Reddit API Wrapper (PRAW), a Python library that provides a simple interface for accessing Reddit's API, extracts comments and associated metadata from the first 20 top posts related to cannabis legalization in Germany.

The metadata collected included the following:

- **ID**: Unique identifier for each comment
- **Author**: Username of the comment's author
- **Date**: Timestamp of when the comment was posted
- **Score**: The net upvotes (upvotes minus downvotes)

A total of 13,750 comments were collected through this process.

Preparing Data For Analysis

This section processed the collected data to facilitate effective clustering and classification. Initially, comments labeled as "deleted" were excluded. Subsequently, to concentrate on comments with substantive content, those containing fewer than three words were removed. Short comments often lack the necessary context and depth, which are essential for accurate analysis.

Language detection was conducted for each comment to ascertain the language utilized. The LangDetect library facilitated this process. Comments not identified as English or German were excluded from further analysis. For comments identified as German, the DeepL API was employed to translate them into English. This translation step was imperative to standardize the language of the comments, thereby enabling consistent analysis.

Following the processes of language detection and translation, text normalization was conducted. All comments were converted to lowercase to ensure consistency. Furthermore, URLs and punctuation were eliminated using regular expressions, as URLs are frequently irrelevant to the context of the comments, and punctuation can introduce noise into text analysis.

Emojis, which can also affect text analysis, were removed using regular expressions. This step helped maintain the focus on textual content rather than pictorial representations, which can be misleading without context.

The comments were then tokenized into individual words [9]. This was achieved using the Python nltk library, which provides tools for natural language processing. Stop words, which are common words that do not contribute significantly to the meaning of the text (such as "and", "the", etc.), were removed. Following this, the words were lemmatized using the WordNetLemmatizer from nltk. Lemmatization reduces words to their base or root form, which helps in grouping similar words and increases the accuracy of the analysis [10].

12 122 cleaned comments were then saved with their score for analysis.

Analysis

This chapter analyzes preprocessed data to understand public sentiment on cannabis legalization through social media data. The methodology involved several key steps, leveraging advanced natural language processing techniques and machine-learning models.

The methodology employs a sentence transformer model to convert text data into high-dimensional vectors, capturing the semantic essence of each comment. Subsequently, the K-means clustering algorithm optimized using the elbow method identified distinct clusters within the dataset. This approach enables grouping comments with similar themes or topics, providing a structured framework for data analysis.

For each identified cluster, titles were generated using the large language model

`Mistral-7B-Instruct-v0.3-Q4\_K\_M.gguf.`

This process provided an intuitive understanding of the predominant themes within each cluster, facilitating the interpretation of the clustered data.

- **Cluster 1:** Exploring the Evolution of Weed Culture, Food Industry, and Public Policy: From Prohibition to Legalization and Beyond
- **Cluster 2:** "The Evolution and Controversies Surrounding the Legalization of Cannabis: A Global Perspective"
- **Cluster 3:** "The Long Road to Cannabis Legalization in the Netherlands: Challenges, Controversies, and Perspectives"
- **Cluster 4:** "Exploring the Controversial Landscape of Medical and Recreational Cannabis: Benefits, Risks, Legalization, and Personalized"
- **Cluster 5:** "A Miscellany of Modern Societal Discourse: A Collection of Contemporary Cultural Snippets"
- **Cluster 6:** The Impact and Prevalence of Marijuana Smoke in Urban Environments: A Perspective on Odor, Accessibility, and Social Implications
- **Cluster 7:** A Humorous and Ironic Perspective on Germany's Contemporary Challenges, From SCM to April Fool Jokes and Language Barriers

Sentiment analysis was performed on each comment using a large language model to determine the polarity of the expressed opinions (positive or negative).

Finally, sentiment analysis results were combined with Reddit post scores to compute weighted sentiment scores for each cluster. This integration of sentiment polarity and popularity (indicated by Reddit scores) produced metrics that reflected both the intensity and reach of sentiments expressed within each cluster.

This multifaceted strategy allowed a nuanced analysis of the public's reaction to cannabis legalization and provided a framework for understanding sentiments around different topics discussed on social media [17].

Clustering

This section outlines the detailed methodology for clustering social media data to group comments with similar themes or topics. The process utilized a sentence transformer model for text embedding and a K-means clustering algorithm optimized through the elbow method.

Sentence Transformer

A sentence transformer model is utilized to convert textual data into high-dimensional embeddings. Sentence transformers generate dense vector representations (embeddings) for sentences, encapsulating semantic meanings in a numerical format. These embeddings enable the comparison and clustering of textual data based on semantic content
citebib18.

The specific sentence transformer model used was

Title	Subreddit	Comments
Cannabis use and cultivation is now legal in Germany! Bubatz legal!	r/europe	244
Germany to legalize marijuana by April	r/europe	379
Germany Moves To Legalize Cannabis, Second Country After Malta In Europe	r/worldnews	1.7K
Cannabis legalization with the recent vote, and why is it so weird?	r/germany	107
Germany to legalize recreational cannabis, say ministers	r/germany	361
About the cannabis legalization?	r/germany	27
Berlin after the Legalization of Cannabis in Germany	r/Damnthatsinteresting	2.1K
Germany Eyes Cannabis Legalization by April	r/worldnews	171
Today, April 1, Cannabis got legalized in Germany. Big smokey meetup at Brandenburg Gate, Berlin	r/MadeMeSmile	1.1K
Germany to legalize cannabis use for recreational purposes	r/news	2.9K
Germany just legalized cannabis	r/GenZ	681
Germany approves partial legalization of cannabis from April	r/worldnews	192
The German Parliament has just voted to legalize Marijuana; people can have up to 3 plants and 50 grams at home for personal use, and 25 grams are allowed to be carried in public. Commercial sale is still illegal though, only...	r/europe	1.1K
Germany to legalize cannabis use for recreational purposes	r/europe	1.5K
Cannabis will be partly legalized in Germany starting April 2024	r/europe	91
Germany Speeds Up The Process To Legalize Recreational Cannabis	r/worldnews	992
Today, April 1, Cannabis got legalized in Germany. Big smokey meetup at Brandenburg Gate, Berlin	r/interestingasfuck	277
Germany partially legalizes cannabis, after April 1st	r/news	134
Germany To Legalize Adult-Use Cannabis in 2024	r/germany	228
Germany to legalize cannabis use for recreational purposes	r/Futurology	233

`all-MiniLM-L6-v2`,

known for its efficiency embeddings [19]. This model is part of the Hugging Face, a widely used library by AI researchers [20].

K-Means Clustering

A K-means clustering algorithm is applied to the embeddings to partition the data into distinct clusters. K-means is a widely used unsupervised learning algorithm that assigns data points to clusters by minimizing within-cluster variance with each point belonging to the cluster with the nearest mean [21]. The implementation utilized the scikit-learn library for K-means clustering.

The embeddings from sentence transformers form a feature space in which each comment is represented as a point. K-means clustering, and then uses distance calculations within this feature space to group comments based on semantic similarity.

The elbow method is employed to determine the optimal number of clusters (K) for K-means clustering. This approach involves running K-means for a range of K values and plotting the sum of the squared distances from each point to its assigned cluster center (inertia) [21].

The optimal K is identified when the inertia decreases more slowly, forming an ëlbowšhape on the plot shown in Fig. 1. At this point, adding more clusters provides diminishing returns in terms of the improved clustering quality.

Once the optimal number of clusters is determined, the K-means algorithm is executed with a value of K. The algorithm iteratively assigns each data point to the nearest cluster center and updates the cluster centers until convergence occurs. This process ensured that the data points within each cluster were as similar as possible, whereas points in different clusters were as dissimilar as possible.

Cluster Titles

Descriptive cluster titles were generated using the

`Mistral-7B-Instruct-v0.3-Q4\_K\_M.gguf`

large language model, which is designed for instruction-following tasks [22]. This process randomly sampled 10 percent of comments from each cluster to create a representative summary. The language model then processed these sampled comments with a prompt to generate concise and informative titles for each cluster, as shown in Fig. 2.

Sentiment Analysis

Sentiment analysis was conducted using Hugging Face's transformers library with the pre-trained

`istilbert-base-uncased-finetuned-sst-2-english`

model. To enhance accuracy, the methodology incorporated a sarcasm indicator consisting of words signaling sarcastic intent, adjusting sentiment classification when these markers appeared in comments.

The weighted sentiment scores for each cluster were determined by integrating the sentiment analysis results with the Reddit post scores. This approach accounts for the popularity of comments, ensuring that more influential comments exert a greater impact on the overall cluster sentiment. Each comment's sentiment was converted into a numerical value (positive=1, negative=-1) and multiplied by its Reddit score to derive a weighted sentiment score. These weighted scores were then aggregated by cluster, and the total sentiment score was normalized to ascertain the percentage of positive and negative sentiments within each group, as illustrated in Fig. 3.

Results

This section delineates the results of the analysis as outlined in the methodology. The comments were systematically cate-

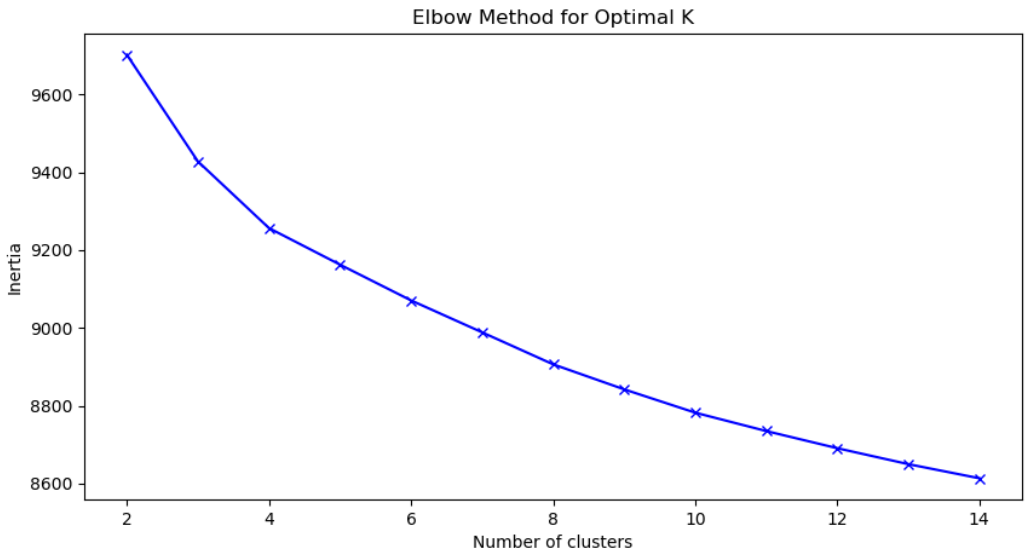


Figure 1. *Elbow Method Plot*

```python
# Function to generate a title for a given cluster
def generate_cluster_title(cluster_texts):
    # Filter out any NaN values that might be present in cluster_texts
    cluster_texts = [str(text) for text in cluster_texts if pd.notna(text)]
    if not cluster_texts:
        return "No valid texts available for this cluster."
    summary = " ".join(cluster_texts)  # Simple concatenation of cluster texts for
    prompt = f"Generate a subject title for the following text: {summary[:2000]}"
    output = llm(
        f"<s>[INST]  {prompt} [/INST]</s>",
        max_tokens=35,  # Limit the length of the generated title
        stop=["</s>"],
        echo=False,
        temperature=0.7
    )
    return output['choices'][0]['text'].strip()

# Read the clusters from the existing CSV file
clusters_df = pd.read_csv("clusters_output.csv")

# Initialize variables
cluster_titles = []
current_cluster_id = None
cluster_texts = []
```

Figure 2. *Generating Cluster Titles*

gorized into seven distinct clusters, each representing a unique facet of the discourse surrounding cannabis legalization, accompanied by corresponding sentiment scores. This approach facilitates a quantitative and nuanced comprehension of this contentious sociopolitical issue. Detailed below are the specifics of each cluster, including the percentages of positive and negative sentiments, along with explanations of the thematic content.

- **Cluster 1:** Exploring the Evolution of Weed Culture, Food Industry, and Public Policy: From Prohibition to Legalization and Beyond
  Positive Sentiment Percentage: 56.4%
  Negative Sentiment Percentage: 43.6%
  This cluster includes comments about the cultural and policy evolution of cannabis, showing its journey from prohibition to legalization. Comments have a balanced perspective, with slightly more positive sentiment, expressing an appreciation for the progress made and its applications to various industries.
- **Cluster 2:** The Evolution and Controversies Surrounding

```python
# Calculate sentiment scores for each cluster
cluster_sentiment_scores = []

for cluster_id in clusters_df['Cluster ID'].unique():
    cluster_data = clusters_df[clusters_df['Cluster ID'] == cluster_id]
    total_score = cluster_data['Score'].abs().sum()  # Sum of absolute scores
    if total_score == 0:
        continue
    total_weighted_sentiment_score = cluster_data['Weighted Sentiment Score'].sum()
    # Calculate the positive and negative sentiment percentages
    positive_percentage = (total_weighted_sentiment_score / total_score) * 50 + 50
    negative_percentage = 100 - positive_percentage
    cluster_sentiment_scores.append({
        'Cluster ID': cluster_id,
        'Positive Sentiment Percentage': positive_percentage,
        'Negative Sentiment Percentage': negative_percentage
    })


cluster_sentiment_df = pd.DataFrame(cluster_sentiment_scores)
cluster_sentiment_df.to_csv("cluster_sentiment_percentages2.csv", index=False)

print("\nSentiment analysis and scoring complete. Results saved to 'test_cluster_sentiment_percentages.csv'.")
```

Figure 3. *Sentiment Score Calculation*

Title	Positive Sentiment Percentage	Negative Sentiment Percentage
Exploring the Evolution of Weed Culture, Food Industry, and Public Policy: From Prohibition to Legalization and Beyond	56.4%	43.6%
The Evolution and Controversies Surrounding the Legalization of Cannabis: A Global Perspective	52.9%	47.1%
The Long Road to Cannabis Legalization in the Netherlands: Challenges, Controversies, and Perspectives	51.6%	48.4%
Exploring the Controversial Landscape of Medical and Recreational Cannabis: Benefits, Risks, Legalization	58.9%	41.1%
A Miscellany of Modern Societal Discourse: A Collection of Contemporary Cultural Snippets	47.4%	52.6%
The Impact and Prevalence of Marijuana Smoke in Urban Environments: A Perspective on Odor, Accessibility, and Social Implications	57.4%	42.6%
A Humorous and Ironic Perspective on Germany's Contemporary Challenges, From SCM to April Fool Jokes and Language Barriers	56.8%	43.2%

the Legalization of Cannabis: A Global Perspective
Positive Sentiment Percentage: 52.9%
Negative Sentiment Percentage: 47.1%
The comments in this cluster focus on cannabis legalization from a global perspective. They address the controversies and differing approaches and laws across countries. The sentiment percentages are relatively balanced, with a slight positive lean. This suggests that, despite existing challenges, there is general optimism over the state of cannabis regulations worldwide.

- **Cluster 3:** The Long Road to Cannabis Legalization in the Netherlands: Challenges, Controversies, and Perspectives
Positive Sentiment Percentage: 51.6%
Negative Sentiment Percentage: 48.4%
This cluster focuses on the experience of cannabis legalization in the Netherlands, reflecting on its journey filled with difficulties and controversies. The ideas are nearly evenly split, indicating mixed reactions on the topic.
- **Cluster 4:** Exploring the Controversial Landscape of Medical and Recreational Cannabis: Benefits, Risks, Legalization, and Personalized
Positive Sentiment Percentage: 58.9%
Negative Sentiment Percentage: 41.1%
Here, the focus is mainly on the medical and recreational use of cannabis, including its benefits and risks. The sentiment is mostly positive, indicating that the public is generally supportive of cannabis for both medical and recreational purposes.
- **Cluster 5:** A Miscellany of Modern Societal Discourse: A Collection of Contemporary Cultural Snippets
Positive Sentiment Percentage: 47.4%
Negative Sentiment Percentage: 52.6%
This cluster includes a variety of cultural discussions, comparing the countries to each other, but this is not exclusively

focused on cannabis. The sentiment is slightly negative, showing a negative perspective on societal differences.

- **Cluster 6:** The Impact and Prevalence of Marijuana Smoke in Urban Environments: A Perspective on Odor, Accessibility, and Social Implications
  Positive Sentiment Percentage: 57.4%
  Negative Sentiment Percentage: 42.6%
  Discussions in this cluster revolve around marijuana usage in urban environments, including odor, accessibility, and social effects. The sentiment is largely positive, showing a positive view of cannabis integration into urban settings, but with some concerns around odor.
- **Cluster 7:** A Humorous and Ironic Perspective on Germany's Contemporary Challenges, From SCM to April Fool Jokes and Language Barriers
  Positive Sentiment Percentage: 56.8%
  Negative Sentiment Percentage: 43.2%
  This cluster includes humorous and ironic take on various modern issues in Germany, including cannabis legalization. The sentiment is primarily positive, reflecting a lighthearted and accepting funny attitude towards the change of cannabis legalization.

Overall, the sentiment tends to be more positive across most clusters, indicating general support and optimism towards cannabis legalization.

Conclusion

This study adopts open-source intelligence (OSINT) techniques to analyze public reactions to cannabis legalization in Germany, focusing on comments from Reddit. Through data collection, cleaning, clustering, and sentiment analysis, the research identified distinct themes and sentiments expressed by the public regarding this polarizing issue.

First, the clustering analysis demonstrated that discussions on cannabis legalization are multifaceted, touching on cultural evolution, global perspectives, medical and recreational use, societal discourse, urban implications, and even humorous takes on contemporary issues. Each cluster provided another angle on the topic, reflecting diverse public opinion and the complexity of the issue.

Sentiment analysis across these clusters showed a general trend toward positive sentiments, indicating broad support and optimism for cannabis legalization. However, the balanced sentiments in some clusters highlight ongoing controversies and mixed reactions.

Integrating Reddit scores with sentiment analysis offers a nuanced understanding of the influence and reach of various comments. This approach ensured that the sentiments reflected not only the content, but also the popularity and impact of the discussions.

This study underscores OSINT's efficacy of OSINT in capturing and analyzing public sentiment on social media. Processing large volumes of user-generated content allows researchers to gain valuable insight into societal trends and public opinion.

This study presents a framework for data processing, including methods for sarcasm detection and weighted scoring systems, to support future investigations. By providing a detailed and replicable methodology, this approach can assist researchers in deriving meaningful conclusions from large datasets, while improving the accuracy of sentiment analysis.

In conclusion, a positive sentiment trend suggests a supportive public stance, although mixed reactions indicate that the discourse is far from settled. Future research could expand this work by exploring reactions over a more extended period or across different social media platforms, providing a more comprehensive view of public sentiment. The findings of this study not only enrich the academic discussion on cannabis legalization, but also provide valuable insights for policymakers, stakeholders, and advocates involved in shaping future cannabis-related policies.

References

[1] A. Bahji and C. Stephenson, "International Perspectives on the implications of cannabis legalization: A systematic review & thematic analysis," International Journal of Environmental Research and Public Health, vol. 16, no. 17, p. 3095, Aug. 2019. doi: 10.3390/ijerph16173095.

[2] H. Stöver, I. I. Michels, B. Werse, and T. Pfeiffer-Gerschel, "Cannabis regulation in Europe: Country report Germany," *Transnational Institute*, 2019. https://www.tni.org/files/publication-downloads/cr_german_10062019.pdf

[3] E. Dauke, J. Gesley, Foreign Law Specialist, "Germany: New Cannabis Act Enters into Force," *Law Library of Congress*, Apr. 2024. [Online]. Available: https://www.loc.gov/item/global-legal-monitor/2024-04-18/germany-new-cannabis-act-enters-into-force/

[4] S. Stieglitz and L. Dang-Xuan, "Social media and political communication: a social media analytics framework," *Social Network Analysis and Mining*, vol. 3, pp. 1277–1291, doi: 2013.10.1007/s13278-012-0079-3.

[5] D. Omand, J. Bartlett, and C. Miller, "Introducing social media intelligence (SOCMINT)," in *Intelligence & National Security*, Routledge, pp. 77–94, 2012. doi: 10.1080/02684527.2012.716965.

[6] H. Gibson, "Acquisition and preparation of data for OSINT investigations," in *Open Source Intelligence Investigation: From Strategy to Implementation*, Springer, 2016, pp. 69–93.doi: 10.1007/978-3-319-47671-1_6

[7] World Population Review, "Reddit Users by Country 2024," [Online]. Available: https://worldpopulationreview.com/country-rankings/reddit-users-by-country. [Accessed: April 6, 2025].

[8] Reddit, "Formatting Guide," [Online]. Available: https://support.reddithelp.com/hc/en-us/articles/360043033952-Formatting-Guide. [Accessed: April 6, 2025].

[9] A. Mullen, L. Benoit, K. Keyes, D. Selivanov, and J. Arnold, "Fast, Consistent Tokenization of Natural Language Text," *Journal of Open Source Software*, vol. 3, no. 23, p. 655, 2018. doi: 10.21105/joss.00655

[10] P. Lama, "Clustering system based on text mining using the k-means algorithm," *Turku University of Applied Sciences Thesis, Information Technology*, 2013. [Online]. Available: https://www.theseus.fi/bitstream/handle/10024/69505/Lama_Prabin.pdf. [Accessed: April 6, 2025].

[11] F. G. Motlagh, S. Shekarpour, A. Sheth, K. Thirunarayan, and M. L. Raymer, "Predicting public opinion on drug legalization: social media analysis and consumption trends.," *Proceedings Of The 2019 IEEE/ACM International Conference On Advances In Social Networks Analysis And Mining*, pp. 952-961, 2020. doi: 10.1145/3341161.3344380

[12] M. Li, N. Kakani, C. Li, and A. Park, "Understanding cannabis information on social media: Examining tweets

from verified, regular, and suspended users," *2020 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 1-10, 2020. doi: 10.1109/ICHI48887.2020.9374387.

[13] A. Unlu and A. Hupli, "Twitter activity surrounding the Finnish green party's cannabis legalisation proposal: A mixed-methods analysis," *Journal of Digital Social Research*, vol. 40, no. 6, pp. 625-645, 2023. doi: 10.1177/14550725231171022.

[14] M. Rychert, C. Wilkins, R. van der Sanden, and J. Prasad, "Exploring digital news, advocacy networks and social media campaigns 'for' and 'against' cannabis legalisation during New Zealand's cannabis legalisation referendum," *Drugs: Education, Prevention and Policy*, vol. 30, no. 5, 2023. doi: 10.1080/09687637.2022.2090897.

[15] C. C. W. Lim *et al.*, "Trends and Socio-Demographic Differences of Cannabis Vaping in the USA and Canada.," *International Journal Of Environmental Research And Public Health*, vol. 19, 2022. doi: 10.3390/ijerph192114394.

[16] M. Mann, W. Ginder, and S.-E. Byun, "Highs and Lows of Cannabis Decriminalization: Twitter Analysis and Ethical and Regulatory Implications for Retailing and Marketing.," *Journal Of Global Marketing*, vol. 34, 57-75, 2022. doi: 10.1080/08911762.2021.1958971.

[17] M. Asgari-Chenaghlu, N. Nikzad-Khasmakhi, and S. Minaee, "Covid-transformer: Detecting COVID-19 trending topics on twitter using universal sentence encoder," *arXiv preprint arXiv:2009.03947*, 2020.

[18] R. Devika, S. Vairavasundaram, C. S. J. Mahenthar, V. Varadarajan, and K. Kotecha, "A Deep Learning Model Based on BERT and Sentence Transformer for Semantic Keyphrase Extraction on Big Social Data," *IEEE Access*, vol. 9, pp. 165252-165261, 2021. doi: 10.1109/ACCESS.2021.3133651.

[19] C. Galli, N. Donos, and E. Calciolari, "Performance of 4 pre-trained sentence transformer models in the semantic query of a systematic review dataset on peri-implantitis," *Information*, vol. 15, no. 2, p. 68, 2024. doi: 10.3390/info15020068.

[20] Hugging Face, "Hugging Face," [Online]. Available: `https://huggingface.co/`. [Accessed: April 6, 2025].

[21] E. U. Oti, M. O. Olusola, F. C. Eze, and S. U. Enogwe, "Comprehensive review of K-Means clustering algorithms," *Criterion*, vol. 12, pp. 22-23, 2021. doi: 10.31695/IJASRE.2021.34050.

[22] A. Q. Jiang *et al.*, "Mistral 7B," *arXiv preprint arXiv:2310.06825*, 2023.

Author Biography

Berfin Ebrar Atabey is a master's student in the CyberMACS program, an Erasmus Mundus initiative involving Kadir Has University (Istanbul), SRH University (Berlin), and UKIM University (North Macedonia). In 2022, she completed a three-month internship at Maynooth University, Ireland, focusing on cybersecurity and Haskell.
She graduated in 2022 with a bachelor's degree in Software Engineering from Atılım University, Ankara. Berfin is dedicated to advancing her cybersecurity expertise and aims to contribute significantly.

Resul Bedii Gümüş received his BS in Mathematics from Boğaziçi University and his MS in Mathematics from Boğaziçi University. He is pursuing a second Master's in the CyberMACS Joint Erasmus Mundus Master program.

Franziska Schwarz received her M.Sc. in Computer Science from Technische Hochschule Brandenburg (Germany) in 2022. Since 2021, she has worked in cyber security consulting with clients in the public and private sectors at different big four consulting companies. Her research focuses on Cybersecurity and Management, Data Protection, IoT, and Smart Home Security.

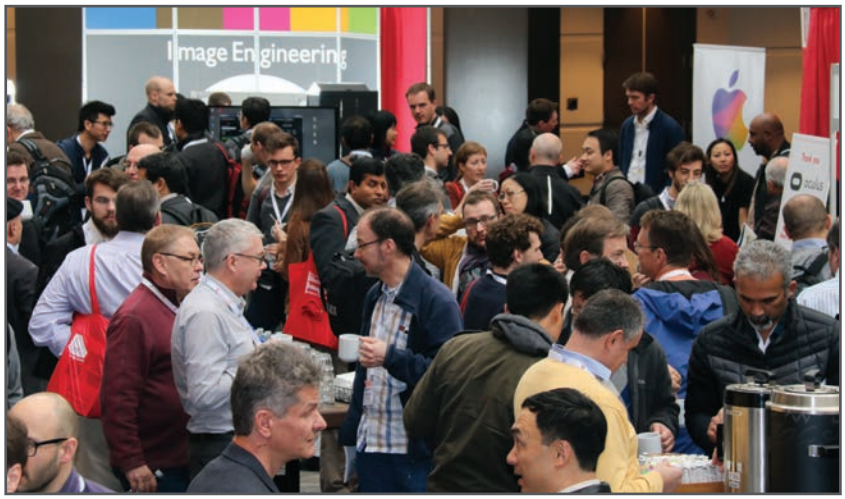
Klaus Schwarz is a Lecturer at SRH Berlin University of Applied Sciences specializing in cybersecurity, embedded systems, and AI applications. He holds B.Sc. and M.Sc. degrees in Computer Science from Brandenburg University and is pursuing his Ph.D. at the University of Granada. His research interests include OSINT, cybersecurity, cloud security, and AI-enhanced security solutions. He has worked as an AI & Cybersecurity Manager at EY Consulting and has published research on secure embedded systems, OSINT methodologies, disaster management systems, and AI-based security applications. At SRH Berlin, he developed the "Applied Mechatronic Systems" undergraduate engineering program.

Reiner Creutzburg is a Retired Professor for Applied Computer Science at the Technische Hochschule Brandenburg in Brandenburg, Germany. Since 2019, he has been a Professor of IT Security at the SRH Berlin University of Applied Sciences, Berlin School of Technology. He is a member of the IEEE and SPIE and chairman of the Multimedia on Mobile Device (MOBMU) Conference at the Electronic Imaging conferences since 2005. In 2019, he was elected a member of the Leibniz Society of Sciences to Berlin e.V. His research interests include Cybersecurity, Digital Forensics, Open Source Intelligence (OSINT), Multimedia Signal Processing, eLearning, Parallel Memory architecture, and Modern Digital Media and Imaging Applications.

JOIN US AT THE NEXT EI!

electronic IMAGING

Imaging across applications . . . Where industry and academia meet!



• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

ADVANCING IMAGING SCIENCE 75 1947 IS&T 2022 imaging.org