International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 October 2024* ***Revised:*** *12 November 2024* ***Accepted:*** *30 December 2024*

Customized Privacy Settings: Empowering User Preferences in Social Media Permissions

Aziz Alshehri
Computing Department
College of Engineering and Computing -AlQunfuda, Umm Al-Qura University
Makkah, Saudi Arabia
aaashehri@uqu.edu.sa

Abstract—In the rapidly evolving the digital landscape of social media, user consent and data privacy have emerged as critical facets of social media interaction. This study addresses the complexity of privacy management within social media applications by probing into user preferences for permission requests. With the objective of streamlining the privacy settings process, the research seeks to understand patterns in user consent and to develop an approach that enhances user engagement without compromising data protection.

Utilizing hierarchical clustering and machine learning techniques on a dataset comprising various social media permissions, we identified four principal clusters. These clusters signify distinct user patterns in granting permissions, reflecting diverse attitudes towards privacy that challenge the conventional one-size-fits-all privacy framework.

Our methodology involved condensing the vast array of permissions into a manageable set. By refining the permissions queried from 46 to 10, our predictive model maintained high accuracy while substantially improving the likelihood of users completing the privacy settings process. This reduction led to a more personalized and less cumbersome user experience.

The study's key findings reveal significant variability in user concerns, ranging from pronounced apprehension to relative indifference regarding permissions. These findings hold substantial implications for privacy management, suggesting a need for customizable privacy settings that align with individual user preferences.

The significance of our research lies in its potential to guide app developers and policymakers in enhancing user trust and satisfaction. By aligning privacy practices with user expectations, this study contributes to the broader dialogue on user-centric privacy approaches in social media and presents a pathway to fostering more secure and personalized digital environments.

Keywords- Permissions Control;Privacy Management; User Profiles;Mobile Social Network App

I. INTRODUCTION

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts if possible. True-Type 1 or Open Type fonts are preferred. Please embed symbol fonts, as well, for math, etc. In the current digital era, platforms such as Facebook, Twitter, and Snapchat have emerged, offering users a comprehensive array of privacy permissions aimed at enhancing control over their digital privacy settings [1]. While these options are extensive, they fail to fully accommodate the varied privacy preferences that distinguish one user from another—a complexity that is well-documented in existing literature [2], [3]. Despite a clear demand for control over personal data, a prevalent inclination exists among users to bypass these detailed customization options [5]. The default privacy protections provided by mainstream mobile operating systems, including Android and iOS, underscore this issue [11], yet usability issues, as highlighted by Kelley et al., may prevent users from effectively understanding and managing these permissions on Android devices [11].

The quest to develop frameworks and tools that facilitate user management of privacy data has attracted significant interest. However, these initiatives often assume a uniformity in users' ability to configure settings and privacy needs, overlooking the reality of diverse privacy attitudes and expectations [1]. For example, the sensitivity perceived by individual users towards personal information, such as age or gender on their social network profiles, varies significantly [15]. The prevailing assumption of homogeneity in users' privacy needs is increasingly impractical [14].

Addressing these challenges, developers and researchers are increasingly striving to create tools that simplify permission

settings by automatically adapting privacy settings to match users' unique preferences [6]. Through "privacy clustering" and the nuanced approach of Hierarchical Clustering of Social Media Permissions, this study aims to automate the alignment of system privacy settings with user groupings that share similar preferences [1].

This paper explores two central research questions:

RQ1: How can hierarchical clustering be applied to categorize user privacy preferences into distinct groups within social media apps?

RQ2: What is the efficacy of employing these categorized privacy clusters to streamline initial default settings for users, thereby reducing the complexity of privacy management?

Our research demonstrates the feasibility of segmenting users into distinct clusters based on their privacy preferences, suggesting that utilizing these privacy profiles for initial interface settings could significantly reduce user burden. The application of hierarchical clustering methods reveals that four clusters provide the optimal configuration for such categorization.

The paper is organized as follows: Section 2 reviews related work, setting the stage for our research questions. Section 3 describes the methodology, including data collection and the hierarchical clustering technique used to identify optimal clusters. Section 4 presents the results and contextualizes them within the framework of similar studies. Section 5 discusses the implications of our findings. Section 6 outlines the study's limitations and proposes directions for future research. Finally, Section 7 concludes the paper, summarizing its key contributions to the field of privacy management in social media.

II. RELATED WORK

First The advent of mobile and web applications, coupled with the pervasive nature of social networks, has escalated concerns surrounding user privacy. This section delves into a variety of privacy-enhancing tools and methodologies developed to mitigate potential breaches and manage sensitive information disclosure. Despite the breadth of research in this domain, a singular observation emerges: the privacy solutions proposed thus far tend to offer broad-stroke remedies that often fail to accommodate the intricacies of individual privacy preferences, particularly within mobile environments.

Among the vanguard of these solutions are dynamic analysis tools like TaintDroid [7], which pioneers in monitoring real-time data flow and identifying unauthorized dissemination of sensitive information via mobile apps. Similarly, static analysis tools such as PiOS [6] have been instrumental in uncovering privacy breaches concerning third-party data sharing, albeit with limitations in user engagement and data control. While these tools offer groundbreaking methodologies in data tracking and analysis, their reliance on user-initiated configurations and the absence of intuitive user interfaces significantly diminish their accessibility and practicality for the general populace.

The literature also presents solutions aimed at elevating user awareness regarding privacy concerns. Balebako et al. [5], for instance, leverage the TaintDroid platform to inform users about potential privacy breaches, offering various user interfaces to highlight sensitive data on users' devices. This approach, while educational, stops short of empowering users to specify and protect personal information actively.

In contrast, solutions like AppFence [10] and ProtectMyPrivacy [1] propose mechanisms that afford users more direct control over their data. AppFence employs a "replacement information" strategy, providing shadow data in lieu of real data to untrusted applications, thereby safeguarding user privacy without compromising app functionality. ProtectMyPrivacy extends this concept by offering fine-grained privacy settings and real-time data access decisions, enhanced by a crowdsourcing system that provides app-specific privacy recommendations.

Further contributions to privacy management include AntMonitor [12] and PrivacyGuard [14], which introduce sophisticated networking analysis techniques to map data packets to apps, utilizing VPN service APIs for traffic interception. This innovation allows for granular privacy settings and offers a novel approach to real-time user protection. Despite these advancements, such tools presuppose a level of user savvy in selecting personal information for system detection, underscoring a recurring theme in privacy research: the challenge of balancing comprehensive privacy protection with user-friendly design.

Recent studies, including those by Frank et al. [9] and Liu et al. [13], pivot towards modeling and predicting users' privacy preferences. By analyzing patterns in permission requests across thousands of apps, these studies aim to uncover commonalities in user behavior and preferences. However, their focus primarily on app permissions, rather than a holistic view of privacy concerns, highlights a gap in the literature: a need for models that encapsulate a wider range of privacy dimensions, including app categories, data usage, and contextual factors.

In light of these observations, our study proposes a novel approach that leverages hierarchical clustering of social media permissions to segment user privacy preferences into distinct clusters. This method aims to provide a more nuanced understanding and tailored management of privacy

settings, catering to the diverse needs and expectations of users.

To better illustrate the contrasts and limitations of current privacy-enhancing tools, Table 1 summarizes their objectives, methodologies, levels of user control, and limitations. This overview not only highlights the need for adaptable privacy management solutions but also contextualizes our study's contribution to this field

TABLE 1: COMPARATIVE OVERVIEW OF PRIVACY ENHANCEMENT TOOLS AND METHODOLOGIES

Tool/Study	Objective	Methodology	User Control	Limitations
TaintDroid [7]	Detect unauthorized data dissemination	Dynamic analysis	High (requires user input for configuration)	Complex for non-technical users; does not address UI usability
Balebako et al. [5]	Increase user awareness of privacy breaches	Utilizes TaintDroid; User interface options for privacy sensitivity	Moderate (informs users but lacks personalization)	Users cannot specify privacy boundaries
PiOS [6]	Examine apps for sensitive information leaks	Static analysis	None (no user interaction)	Lacks user data access and control
AppFence [10]	Protect sensitive data from apps	"Replacement information" strategy; Shadowing and blocking	High (allows user to block or replace data)	Could hinder app functionality; lacks transparency in data use
AntMonitor [12] & PrivacyGuard [14]	Provide fine-grained privacy settings; Analyze network traffic	VPN service API for traffic analysis	High (granular privacy settings)	Assumes users can accurately manage settings; complexity in handling data
ProtectMyPrivacy [1]	Offer anonymized data options and detect breaches in iOS apps	Anonymization; Crowdsourcing for app recommendations	High (real-time decision on data access)	Generalized privacy settings; lacks personalized recommendations
Frank et al. [9]	Model and predict privacy preferences	Probabilistic analysis of app permissions	Not applicable	Focuses on permission patterns without broader privacy context
Liu et al. [13]	Understand user permissions preferences	Machine learning clustering methods	High (profiles based on user choices)	Limited to app permissions, not covering broader privacy needs

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 October 2024* ***Revised:*** *12 November 2024* ***Accepted:*** *30 December 2024*

The table above encapsulates the scope and limitations of existing privacy tools and methodologies, setting the stage for our research's contribution to the field. By addressing the shortcomings identified in prior studies, particularly the need for personalized and user-friendly privacy management solutions, our work aims to enhance the user experience on mobile platforms, providing a tailored approach to privacy settings that align with individual user profiles.

The panorama of privacy solutions thus far presents a fragmented landscape, where the balance between sophisticated privacy protections and user-centric design remains elusive. Our investigation into hierarchical clustering seeks not only to bridge this divide but also to pioneer a pathway towards a more empathetic and user-aligned model of privacy management in the digital age.

III. METHODOLOGY

In this study, we adopted hierarchical clustering to discern the underlying patterns in user preferences regarding social media permissions. Hierarchical clustering is particularly advantageous for its ability to provide a visual summary of the data in the form of a dendrogram, which enables the identification of clusters at various levels of granularity. This method does not require pre-specification of the number of clusters, unlike partitioning methods such as K-means, allowing us to explore the data more freely and potentially uncover a more natural grouping within the user responses.

A. Data Collection and Preparation

Continuing from our prior study [2], we involved 381 participants who completed a comprehensive survey with 46 questions, each addressing a distinct social media permission type. The responses were encoded numerically and normalized to ensure uniformity in scale, which is particularly critical for distance-based clustering methods such as Hierarchical Clustering.

The dataset revealed a comprehensive array of participant responses to permission requests spanning several prominent applications, such as Snapchat, Twitter, Facebook, and YouTube. Each column in the dataset represented a distinct permission type, with rows encapsulating individual participant ratings. These ratings, ranging from 1 to 5, served as indicators of the participants' consent levels to each requested permission.

Further scrutiny of the dataset for missing values uncovered a minimal count of three. To preserve the comprehensive nature of our data, we chose median imputation for addressing these gaps. This method was particularly apt given its robustness against the skewing effects of outliers.

With the dataset thus refined and prepared, we were poised to apply Hierarchical Clustering. This advanced analytical approach was instrumental in uncovering the inherent groupings within the participant responses, shedding light on diverse user preferences in the context of social media permissions. Such insights are invaluable for the development of nuanced and user-focused privacy settings in social media platforms.

B. Identifying Clusters and Participant Distribution

After applying hierarchical clustering, we analyzed the resulting dendrogram for the optimal cluster definition, guided by visual inspection and statistical measures, such as the inconsistency coefficient, which evaluates the heterogeneity between successive cluster formations.

C. Feature Importance Evaluation

To prioritize social media permission features by their impact on clustering, we employed a Random Forest model, renowned for its effectiveness in handling categorical data and providing interpretable metrics on feature importance.

D. Accuracy Assessment

We assessed model accuracy through a progressive inclusion of the most influential features, as identified by the Random Forest analysis. This process, illustrated in Figure 8, allowed us to gauge the impact of feature selection on the model's predictive accuracy.

E. Permission Reduction

The analysis led to a reduction of survey questions from 46 to 10, focusing on those most predictive of user cluster membership while considering the balance between model accuracy and user engagement in survey completion.

F. Final Assignment

Utilizing responses to this refined question set, we assigned users to one of four clusters, each reflecting a distinct privacy preference profile. This approach was validated by observing negligible gains in accuracy beyond the inclusion of the top 10 features, suggesting an optimal balance between thoroughness and user experience.

IV. RESULTS AND FINDINGS

As mentioned, this study employs advanced clustering analysis and machine learning techniques to deepen our understanding of user preferences concerning social media permissions. Building on the demographic groundwork laid in our previous survey, which revealed a significant youth presence and high general concern for privacy among users, we have concentrated our analytical endeavours on uncovering previously unobserved patterns and relationships within the data.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 October 2024* ***Revised:*** *12 November 2024* ***Accepted:*** *30 December 2024*

This approach not only allows us to identify distinct user groups based on their permission preferences but also aids in understanding the nuanced variations in these preferences.

By applying a Hierarchical Clustering Dendrogram of Mobile App Permission Requests, we have visually articulated the proximities and affiliations between various permissions as influenced by user access trends (see Figure 1). The demographic insights informed the interpretation of these clusters, providing a richer understanding of the context behind user preferences. The clusters depicted in Figure 1 underscore the diversity within our sample, mirroring the varied demographic factors such as age, education level, and IT skills that correlate with distinct privacy concerns and behaviours.

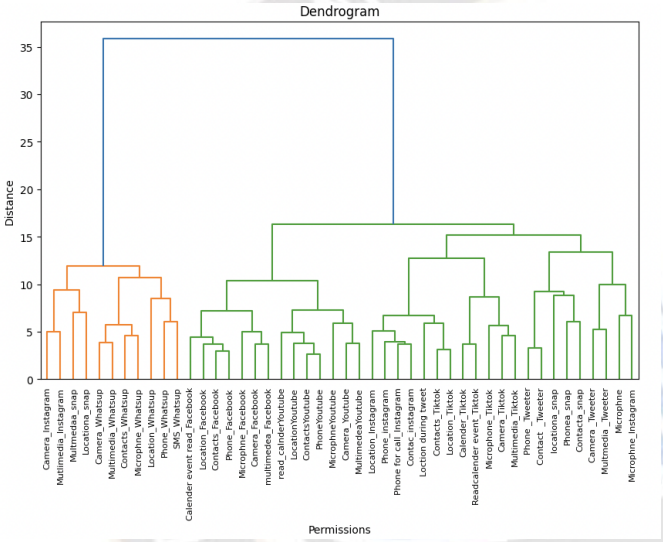

Figure 1. Hierarchical Clustering Dendrogram of Mobile App Permission Requests

Upon applying the Hierarchical Clustering technique to our dataset, we have identified four distinct clusters within the realm of mobile app permission requests, as indicated in Figure 2. These clusters were discerned by analyzing the distance or dissimilarity between permissions, as illustrated in the accompanying dendrogram. Each cluster represents a group of permissions that users tend to grant or deny in conjunction. This clustering allows us to interpret the underlying structure of user preferences, revealing the grouping of permissions that are likely to be accepted or rejected together. The demarcation at a specific distance level, as indicated by the dashed line in the dendrogram, suggests the optimal number of clusters for this particular analysis, which in this case is four. These findings provide significant insights into user behavior patterns and can guide developers and researchers in understanding privacy concerns and permission management strategies.

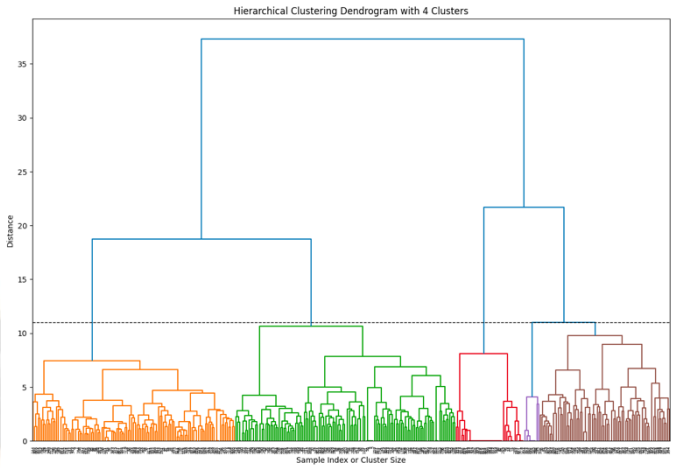

Figure 2. A Hierarchical Clustering Dendrogram with Four Defined Clusters

In our analysis of user attitudes towards app permissions, we have categorized the responses into four segments, as depicted in Figure 3. The pie chart illustrates the proportion of data points corresponding to each user concern cluster. Notably, a significant 34.6% of users fall into the 'Concerned' category, indicating a cautious approach to granting permissions. In contrast, 31.8% of users are 'Very Concerned', demonstrating an even higher level of apprehension regarding data privacy. Meanwhile, 22.8% of users are categorized as 'Unconcerned', showing a more relaxed attitude towards sharing permissions. Lastly, a smaller segment of 10.8% is labeled as 'Very Unconcerned', suggesting a segment of the user base that is largely indifferent to permission requests. This distribution provides valuable insight into the varying levels of user concern regarding privacy and the management of app permissions.

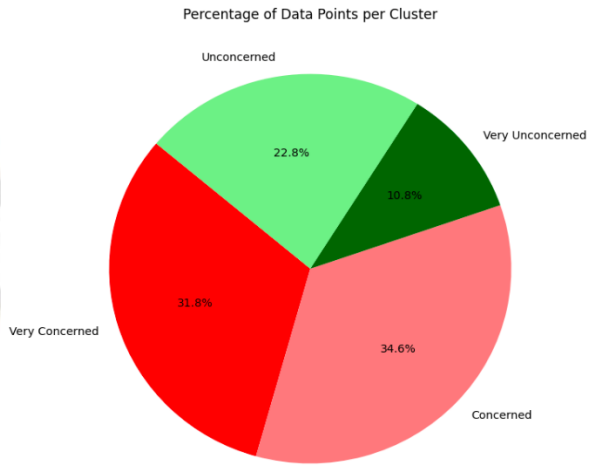

Figure 3. : Percentage Breakdown of Privacy Concern Clusters in Mobile App Permissions

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 October 2024* ***Revised:*** *12 November 2024* ***Accepted:*** *30 December 2024*

Figure 4 presents a heatmap illustrating the varying comfort levels with data sharing across four distinct clusters. The first cluster is characterized by predominantly red and orange hues, indicating a high level of concern among its members about sharing personal information. Participants in this cluster are likely the most cautious and may favor stricter data privacy measures.

In contrast, the second cluster exhibits a mix of yellow, green, and orange, reflecting a more nuanced perception of data sharing, where comfort levels significantly fluctuate depending on the type of data and the context in which it is shared.

The third and fourth clusters show a marked difference, dominated by green hues, which suggest a general comfort and willingness to share data. The lower levels of concern in these clusters may indicate a more trusting attitude toward data sharing practices, possibly reflecting a different assessment of the associated benefits and risks.

Overall, the heatmap in Figure 4 depicts a clear gradient of data sharing comfort levels, from the highly cautious first cluster to the relatively unconcerned third and fourth clusters. This visual representation is crucial for understanding the range of attitudes toward data privacy across different user segments, which could be instrumental in crafting targeted data management strategies.

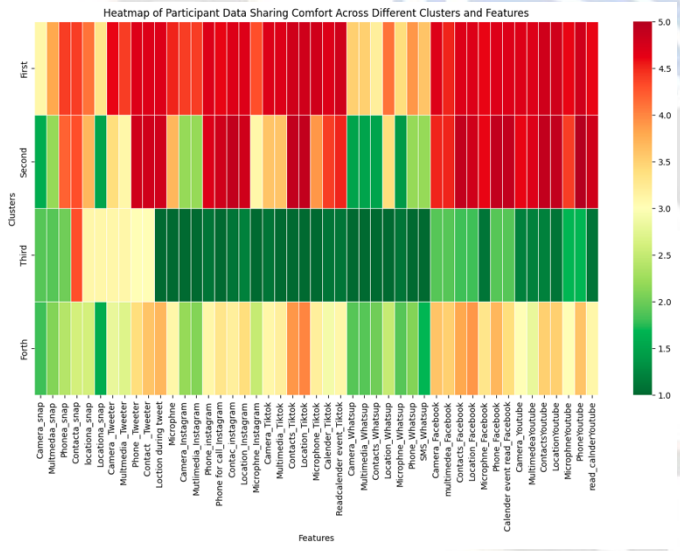

Figure 4. Heatmap of concern level across different clusters and features

Figure 5 provides a comparative visualization of user concerns regarding permissions for various social media platforms across different clusters. The chart shows clear distinctions between the clusters, reflecting diverse concern levels.

Cluster 1 is characterized by a dark red color, signaling the highest level of concern with an average score of 4.4 across all permissions. This deep shade indicates a heightened awareness and apprehensiveness about the handling and access of personal information by social media platforms.

Cluster 2, depicted with a lighter shade of red, represents a slightly lower average concern level of 3.8. The visual differentiation between the reds of Clusters 1 and 2 effectively conveys the relative intensity of concern, with Cluster 2's users being concerned, albeit to a lesser degree than those in Cluster 1.

Cluster 3, shown in dark green, indicates a notably lower average concern, scoring just 1.6. The stark contrast in both color and the relative height of the bars compared to Clusters 1 and 2 emphasizes the significant divergence in concern levels about social media permissions.

Lastly, Cluster 4, with an average concern level of 2.9, is represented by a lighter shade of green. This coloring suggests a moderate level of concern, more relaxed than the heightened vigilance seen in Clusters 1 and 2 but still indicative of some awareness.

In summary, Figure 5 effectively communicates the variance in permission-related concerns among the different user clusters, employing an intuitive color-coding system that aligns with the intensity of user concerns. The vivid red for Cluster 1 points to a vigilant and potentially cautious approach to social media permissions, while the green shades for Clusters 3 and 4 reflect a more relaxed posture. This visual comparison facilitates an immediate understanding of the varying attitudes towards permissions across the user spectrum.

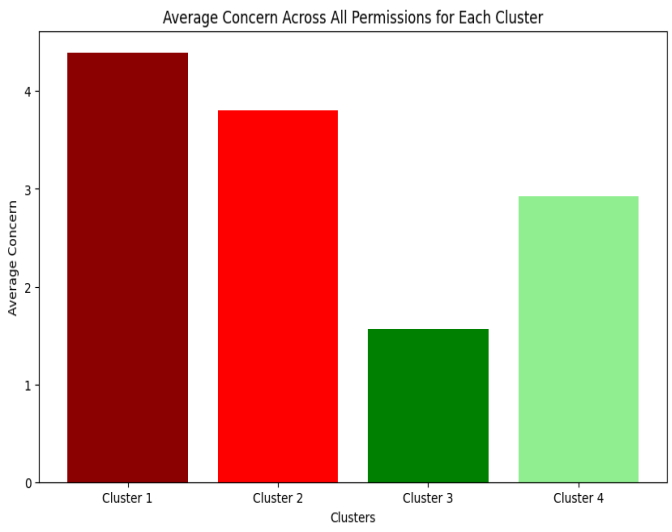

Figure 5. Average Concern Across All Permissions For Each Cluster

Figure 6 depicts a detailed bar and line graph that illustrates the varying levels of concern across multiple data permissions among four distinct clusters. The bars represent the average concern rating for each permission, with distinct colors assigned to each of the four clusters. This visualization makes apparent

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 October 2024* ***Revised:*** *12 November 2024* ***Accepted:*** *30 December 2024*

the notable differences and similarities in privacy concerns among the clusters.

Lines connecting the tops of the bars highlight discernible trends or patterns in the permission concerns across the clusters. For example, Cluster 2 consistently shows the highest levels of concern, especially regarding permissions related to social media activities, such as accessing contacts and location data. Conversely, Cluster 3 tends to demonstrate the lowest levels of concern, indicating a more relaxed attitude toward data privacy.

This graph offers a comprehensive overview of how different user groups value the privacy of their digital data, showcasing the varying degrees of concern across different permissions and social media platforms.

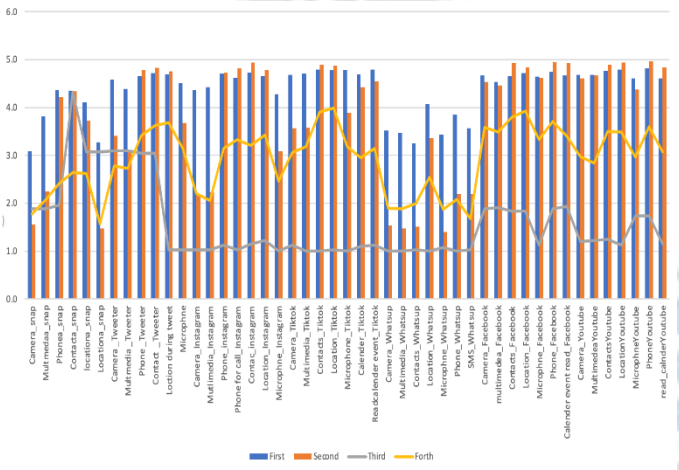

Figure 6. Analysis of User Concerns Across Different Clusters

Table 2 presents an analysis of user concerns across different clusters, revealing distinct patterns in the prioritization of data types and social media applications. Cluster One shows pronounced concerns primarily with TikTok and YouTube, emphasizing permissions like Microphone, Phone, Location, and Contacts on TikTok, along with Phone and Location on YouTube. This reflects a heightened sensitivity towards how these platforms handle personal information.

In contrast, Cluster Two's concerns are more widely distributed across various platforms. The leading concerns are YouTube's Phone permission, several Facebook permissions including Calendar and Contacts, as well as TikTok and Instagram features. This diversity indicates varied app usage and concerns, with a notable focus on how Facebook manages phone and contact information.

Cluster Three demonstrates significantly lower concern levels, particularly for Snapchat permissions such as Contact and Location, and Twitter's Camera and Multimedia features. This trend suggests a different engagement pattern with social media, with a focus on Snapchat and Twitter functionalities.

Cluster Fourth exhibits a moderate level of concern, with attention to Location and Contacts permissions on TikTok and Facebook, along with Phone permissions on Facebook. This pattern points to specific concerns about the management of location and contact information on these platforms.

Overall, the clusters show overlapping concerns, especially with TikTok and Facebook, yet each cluster also displays unique top concerns. Cluster One and Cluster Four are notably concerned with TikTok's permissions, while Cluster Two indicates a broader concern across platforms, and Cluster Three's focus is more on Snapchat and Twitter. These findings underscore the complexity and diversity of user concerns and the need for privacy settings that can cater to varied user sensitivities.

TABLE 2: SOCIAL MEDIA PLATFORM FEATURE RANKINGS BY CLUSTER

Rank	Cluster First	Cluster Second
1	TikTok Microphone (4.8)	YouTube Phone (5.0)
2	YouTube Phone (4.8)	Facebook Calendar (4.9)
3	YouTube Location (4.8)	Facebook Contacts (4.9)
4	YouTube Contacts (4.8)	TikTok Location (4.9)
5	TikTok Location (4.8)	TikTok Contacts (4.9)
6	TikTok Contacts (4.8)	YouTube Location (4.9)
7	TikTok Calendar (4.8)	YouTube Contacts (4.9)
8	Facebook Location (4.7)	Facebook Phone (4.9)
9	Facebook Camera (4.7)	Instagram Contact (4.9)
10	TikTok Multimedia (4.7)	Phone for call Instagram (4.8)
Rank	**Cluster Third**	**Cluster Forth**
1	Snap Chat Contact (4.3)	TikTok Location (4.0)
2	Snap Chat Location (3.1)	Facebook Location (3.9)
3	Snap Chat Multimedia (3.1)	TikTok Contacts (3.9)
4	Twitter Camera (3.1)	Facebook Contacts (3.8)
5	Twitter Multimedia (3.1)	Facebook Phone (3.7)
6	Twitter Phone (3.0)	Location during tweet (3.7)
7	Twitter Contact (3.0)	Twitter Contact (3.6)

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 October 2024* ***Revised:*** *12 November 2024* ***Accepted:*** *30 December 2024*

8	Twitter Phone (2.0)	YouTube Phone (3.6)
9	Snap Chat Camera (1.9)	Facebook Camera (3.6)
10	Facebook Calendar (1.9)	YouTube Contacts (3.5)

- Cluster 1 demonstrates the highest level of concern, particularly for TikTok, evidenced by a deep red color correlating with an average concern score of 4.78. YouTube and Facebook also elicit high concern within this cluster, with scores closely following TikTok. Conversely, Snapchat registers the lowest concern within this cluster but still maintains a moderate score of 3.85.
- Cluster 2's primary concerns are Facebook and YouTube, indicated by the most intense shades of red on the heatmap, with scores of 4.78 and 4.75, respectively. WhatsApp, however, exhibits a significantly lower concern level, denoted by a light green color and a score of 1.95.
- Cluster 3 shows lower overall concern levels compared to Clusters 1 and 2, with YouTube presenting the highest concern at a mid-level score of 3.14. Instagram and WhatsApp are the least concerning for this cluster, both represented by the lightest green on the heatmap with scores of 1.04.
- Cluster 4 exhibits the highest level of concern for TikTok at a score of 3.42, followed by Facebook at 3.60. WhatsApp reflects the least concern within this cluster, with a score of 2.00, denoted by light green.
- Cross-cluster comparison highlights:
- TikTok commands significant attention within Clusters 1 and 4, with Cluster 1 showing the most concern.
- Facebook and YouTube consistently raise high levels of concern within Clusters 1 and 2.
- WhatsApp's concern is lowest within Clusters 2, 3, and 4, contrasting with its higher concern level in Cluster 1.
- Instagram's concern level is moderate within Clusters 1 and 4 but minimal in Cluster 3.
- Twitter, while not the highest concern, maintains a moderate to high level across all clusters except Cluster 3, where it drops significantly.

This comparative analysis underscores that while certain platforms like Facebook and YouTube are universally significant concern areas, each cluster presents distinct concern patterns. Cluster 1 is generally the most concerned across all platforms, while Cluster 3 exhibits the least concern, reflecting the diverse priorities and sensitivities related to data privacy and app permissions.

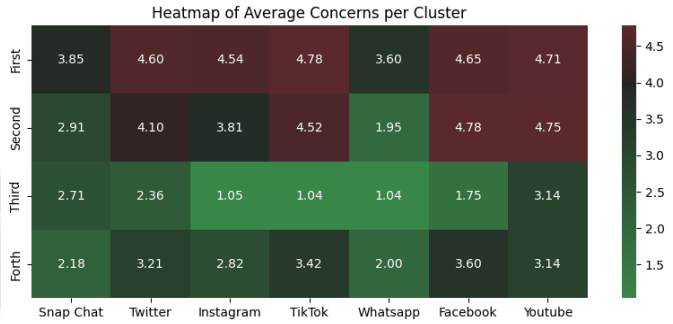

Figure 7. Figure 1: Heatmap Of Average Concerns Across Apps Per Cluster

Prioritizing App Permissions Across Social Media Clusters

In the process of optimizing user privacy and app functionality across various social media clusters, a systematic approach was employed to prioritize app permissions. This methodology involved the evaluation of permissions based on their assigned scores within each cluster, using statistical analysis to determine which permissions should be enabled ('on') and which should be disabled ('off') by default.

Methodology Overview

The core of our prioritization strategy hinged on the use of median scores as a threshold. For each cluster, we calculated the median score of all permissions. This median score served as a critical pivot: permissions rated above the median were considered essential or high-priority and were therefore recommended to be turned 'on', indicating they should be actively granted for the app's optimal functionality and user experience. Conversely, permissions scoring at or below the median were categorized as lower priority or non-essential, recommended to be turned 'off' by default, enhancing user privacy by minimizing unnecessary data access.

This approach ensured that the prioritization of permissions was tailored to the unique characteristics and user preferences within each cluster, reflecting a nuanced understanding of the importance and sensitivity of different permissions in diverse contexts.

Implementation Steps

Median Calculation: For each cluster, we identified and calculated the median score of all permissions. This step was crucial for establishing a benchmark that reflects the central

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 October 2024* ***Revised:*** *12 November 2024* ***Accepted:*** *30 December 2024*

tendency of permission scores within the cluster, providing a balanced perspective that is less influenced by outliers.

Threshold Application: Using the calculated median as a threshold, we systematically evaluated each permission against this benchmark. Permissions with scores above the median were marked 'on', while those at or below the median were marked 'off'.

Consolidated View: To facilitate a comprehensive overview and ease of comparison, we compiled the decisions into a consolidated table, showcasing the 'on'/'off' status of permissions across all clusters alongside their respective median scores. This table not only provided a clear visual representation of our methodology and its outcomes but also allowed for straightforward cross-cluster comparisons and analyses.

Feature Importance Analysis

The goal of our analysis was to identify the most important features within a dataset related to permissions for various social media platforms and applications. By understanding which permissions are most influential, we can streamline our focus, reduce the dimensionality of the dataset, and potentially improve the performance and interpretability of our models. This is particularly crucial in contexts where permissions might impact user privacy, app functionality, or security concerns.

To achieve this objective, we employed a Random Forest classifier due to its ability to handle high-dimensional data and provide insight into feature importance. The Random Forest model evaluates the significance of each feature in predicting the target variable, allowing us to rank permissions based on their impact.

Below is a table summarizing the top 20 features, as determined by their importance scores. These scores reflect the contribution of each permission to the model's ability to accurately classify or predict the target variable. By focusing on these permissions, we aim to minimize the complexity of our questions and analyses, targeting the most critical aspects of the data.

TABLE 3: FEATURE IMPORTANCE WEIGHT

Rank	Feature	Importance
0	Camera Access on Instagram	0.066147
1	Multimedia Access on Instagram	0.049138
2	Microphone Access on WhatsApp	0.046735
3	Multimedia Access on WhatsApp	0.042329
4	Location Access on Snapchat	0.037948
5	Calendar Event Read on Facebook	0.037876
6	Read Calendar on YouTube	0.036621
7	Contact Access on Instagram	0.036585
8	Camera Access on WhatsApp	0.036284
9	Contact Access on WhatsApp	0.025621
10	Phone Access on Twitter	0.025533
11	Phone Access on Instagram	0.023825
12	Camera Access on Snapchat	0.022822
13	Location Access on YouTube	0.022811
14	Phone Access on WhatsApp	0.022743
15	Location Access on Instagram	0.022363
16	Contact Access on TikTok	0.020245
17	Phone Access on Facebook	0.020083
18	Call Phone Access on Instagram	0.019772
19	Multimedia Access on TikTok	0.019326

This analysis underscores the varied impact of permissions across different platforms and functionalities. By narrowing our

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 October 2024* ***Revised:*** *12 November 2024* ***Accepted:*** *30 December 2024*

focus to the most significant permissions, we can refine our approach to privacy and functionality concerns, ensuring that our models are both efficient and effective in addressing the key aspects of user permissions.

Optimal Feature Selection for User Profiling

In the context of user profiling, particularly when permissions related to social media are involved, striking a balance between user engagement (measured by their willingness to respond to questions) and the accuracy of the predictive model is crucial. Our analysis, as shown in Figure 8, serves as a guide to make an informed decision on the number of features.

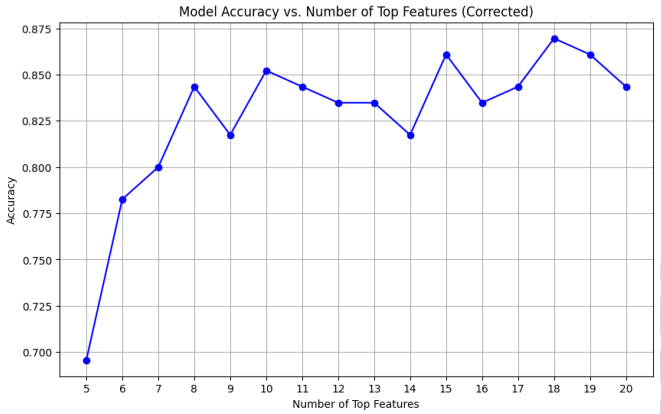

Figure 8. Impact of Feature Selection on Model Accuracy

After reviewing the accuracy curve, we can observe that the model reaches a relatively high level of accuracy with the top 10 features and that the marginal gains in accuracy beyond this point do not justify the additional complexity and user effort. Therefore, it seems reasonable to select 10 features for the initial user questions. Here's why:

- **User Experience:** Asking users too many questions right away can be overwhelming and might lead to reduced engagement or even abandonment of the process. A set of 10 questions strikes a good balance, providing a user-friendly experience.
- **Model Simplicity and Efficiency:** A model with fewer features is simpler, easier to interpret, and often more robust to changes and noise in the data. It's also computationally more efficient, which can be crucial when dealing with large user bases.
- **Accuracy Retention:** The graph indicates that a model with the 10 most important features retains a high accuracy level, suggesting that we capture the most significant data with a reduced question set.

By focusing on the top 10 permissions, we can craft a concise yet effective questionnaire that gathers essential data to accurately assign users to the appropriate cluster without overburdening them with too many questions. This approach respects users' time and attention while still collecting the necessary information to understand their social media usage and privacy preferences.

Implementing this strategy will allow us to create a streamlined user onboarding process, enhancing user satisfaction and engagement while maintaining a high standard of data-driven insights.

V. Discussion

The core of our analysis centers around optimizing the initial user inquiry for social media applications, specifically focusing on permissions. The primary challenge lies in balancing the need for comprehensive user data to drive accurate clustering and the desire to enhance user experience by limiting the number of questions asked. Our research indicates that by adopting a strategic approach to feature selection based on feature importance, we can significantly reduce the number of permission-related questions without compromising the accuracy of user cluster assignments.

The results of our hierarchical clustering analysis and Random Forest feature importance evaluation have paved the way for this balance. The data revealed distinct behavioral patterns among users concerning social media permissions, which we categorized into four major clusters. Each cluster encapsulates a set of permissions that users are inclined to grant or deny collectively, thus allowing us to infer the nuanced variations in user preferences.

VI. Limitations and Future Work

Despite the promising outcomes, this study has limitations that warrant consideration. The feature importance evaluation is dependent on the dataset and the Random Forest model's performance, which may vary with different data or models. Furthermore, user behaviors and attitudes towards privacy are dynamic and can evolve, suggesting that the identified clusters and feature importance rankings may require periodic reassessment.

Future work should consider real-time analysis of user behavior to dynamically adjust the questions asked, keeping the model and user profiling current. Additionally, exploring other machine learning techniques and models could yield even more efficient clustering and feature selection methods. Further research could also delve into the qualitative aspects of user preferences to enhance the quantitative methods used in this study. The ultimate goal is to create a privacy framework that is both user-centric and adaptable to the changing digital landscape.

VII. CONCLUSION

This study offers a nuanced understanding of user engagement with social media permissions, demonstrating the potential of data-driven approaches in enhancing privacy management. By employing hierarchical clustering and machine learning techniques, our research has discerned four principal clusters that categorize users based on their permission-granting patterns. These clusters indicate a spectrum of user attitudes towards privacy, from significant apprehension to relative unconcern, challenging the notion of a uniform approach to user privacy.

The strategic reduction of permission queries from 46 to a focused set of 10, determined by feature importance metrics, has maintained the predictive accuracy of our model while substantially increasing user engagement. This streamlined approach suggests that users are more likely to engage with privacy settings when presented with a distilled array of relevant questions. The study supports the concept that a more personalized and less daunting initial interface enhances user experience and trust, potentially fostering a more favorable perception of the platform's privacy practices.

Our findings hold substantial implications for the field. They advocate for the adoption of tailored privacy frameworks that resonate with diverse user preferences. The study underlines the importance of developers and policymakers in recognizing the multiplicity of user concerns, advocating for adaptable privacy measures that can be customized at the outset based on user interaction.

In essence, our study delineates a method for anticipating user preferences concerning privacy settings, enabling a more informed and user-centric configuration of social media applications. This proactive stance on privacy management is instrumental in cultivating user trust and satisfaction, advancing the discourse on privacy practices in the digital age.

References

[1] Agarwal, Y., & Hall, M. (2012). ProtectMyPrivacy: Detecting and Mitigating Privacy Leaks on iOS Devices Using Crowdsourcing Categories and Subject Descriptors. Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services, 6(September), 97–110.

[2] Alshehri, A., & Alamri, S. (2022). Exploring Social Media Privacy Preferences in Saudi Arabia. 22(1).

[3] Alshehri, A., & Alotaibi, F. (2019). Profiling Mobile Users Privacy Preferences. International Journal of Digital Society (IJDS), 10(1), 1436–1441. https://infonomics-society.org/wp-content/uploads/Profiling-Mobile-Users-Privacy-Preferences.pdf.

[4] Anton, A. I., Earp, J. B., & Young, J. D. (2010). How Internet Users' Privacy Concerns Have Evolved. IEEE Privacy & Security, 1936(February), 21–27. https://doi.org/10.1109/MSP.2010.38.

[5] Balebako, R., Jung, J., Lu, W., Cranor, L. F., & Nguyen, C. (2013). "Little Brothers Watching You": Raising Awareness of Data Leaks on Smartphones. SOUPS '13: Proceedings of the Ninth Symposium on Usable Privacy and Security, 12:1--12:11. https://doi.org/10.1145/2501604.2501616.

[6] Egele, M., Kruegel, C., Kirda, E., & Vigna, G. (2011). PiOS Detecting privacy leaks in iOS applications. Proceedings of the 18th Annual Network & Distributed System Security Symposium, NDSS 2011, 11.

[7] Enck, W., Gilbert, P., Chun, B.-G., Cox, L. P., Jung, J., McDaniel, P., Sheth, A. N., Han, S., Tendulkar, V., Chun, B.-G., Cox, L. P., Jung, J., McDaniel, P., & Sheth, A. N. (2014). TaintDroid: an information-flow tracking system for real-time privacy monitoring on smartphones. ACM Transactions on Computer Systems (TOCS), 32(2), 5. https://doi.org/10.1145/2494522.

[8] Federal Trade Commission. (2013). Mobile privacy disclosures - Building trust through transparency. February 29.

[9] Frank, M., Dong, B., Felt, A. P., & Song, D. (2012). Mining permission request patterns from Android and Facebook applications. Proceedings - IEEE International Conference on Data Mining, ICDM, 870–875. https://doi.org/10.1109/ICDM.2012.86.

[10] Hornyack, P., Han, S., Jung, J., Schechter, S., & Wetherall, D. (2011). These Aren't the Droids You're Looking for: Retrofitting Android to Protect Data from Imperious Applications. In Proceedings of the 18th ACM Conference on Computer and Communications Security, 639–652. https://doi.org/10.1145/2046707.2046780.

[11] Kelley, P. G., Consolvo, S., Cranor, L. F., Jung, J., Sadeh, N., & Wetherall, D. (2012). A conundrum of permissions: installing applications on an android smartphone. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7398 LNCS, 68–79. https://doi.org/10.1007/978-3-642-34638-5_6.

[12] Le, A., Irvine, U. C., Langhoff, S., & Shuba, A. (2015). AntMonitor: A System for Monitoring from Mobile Devices. 1, 15–20.

[13] Liu, B., Lin, J., & Sadeh, N. (2013). Reconciling Mobile App Privacy and Usability on Smartphones: Could User Privacy Profiles Help? http://reports-

archive.adm.cs.cmu.edu/anon/anon/home/ftp/usr0/ftp/2013/CMU-CS-13-128.pdf.

[14] Song, Y. (2015). PrivacyGuard: A VPN-Based Approach to Detect Privacy Leakages on Android Devices. 15–26. https://doi.org/10.1145/2808117.2808120.

[15] Song, Y., & Hengartner, U. (2015). PrivacyGuard: A VPN-based Platform to Detect Information Leakage on Android Devices. Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices - SPSM '15, 15–26. https://doi.org/10.1145/2808117.2808120.

[16] TRUSTe. (2016). 2016 TRUSTe/NCSA Consumer Privacy Infographic - US Edition | TRUSTe. https://www.truste.com/resources/privacy-research/ncsa-consumer-privacy-index-us/.

[17] Yang, Z., Yang, M., Zhang, Y., Gu, G., Ning, P., & Wang, X. S. (2013). AppIntent: analyzing sensitive data transmission in Android for privacy leakage detection. Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security - CCS '13, 1043–1054. https://doi.org/10.1145/2508859.2516676.

[18] Zhou, Y., Zhang, X., Jiang, X., & Freeh, V. W. (2011). Taming information-stealing Smartphone Applications (on Android). 4th International Conference on Trust and Trustworthy Computing, 2011, 93–107.