Is identifying boredom the answer to controlling the bombardment of notifications on mobile devices?

Rashid Kamal[1,2] · Aimal Rextin[3] · Chris Nugent[1] · Ian Cleland[1] · Paul McCullagh[1]

Abstract

Mobile notifications have become ubiquitous in modern life, yet excessive volumes contribute to interruption overload. This paper investigates intelligent notification management leveraging user context. A three-stage methodology employed a focus group, survey, and in-the-wild data collection app. The focus group ($n = 12$) provided preliminary insights into notification perceptions during boredom which informed survey design. The survey ($n = 106$) probed usage habits across times, days, and app categories. The SeektheNotification app gathered real-world notification data from 20 Android users over 3 months. Analysis revealed social and personal apps dominate notification volumes (91% combined). Shorter response delays occurred on weekends and after 12 pm, suggesting heightened user receptivity during boredom. Random Forest classification achieved 88% accuracy, outperforming 13 other algorithms, underscoring machine learning's potential for context-aware notification systems. Our exploratory findings indicate notifications could be optimized by considering situational factors like boredom. Further research should expand context beyond boredom and employ advanced deep learning techniques. This preliminary study demonstrates the promise of leveraging user psychology and machine intelligence to develop smarter interruption management systems to combat notification overload.

1 Introduction

Effective management of push notifications on mobile devices remains an open challenge despite extensive prior research. Notifications must balance timely delivery of important information against limiting distraction from excessive volumes, particularly as ubiquitous computing intensifies notification overload (Mehrotra et al. 2015; Mehrotra and Musolesi 2020). This highlights the need to explore innovative notification management approaches that empower users to contend with notification deluges more efficiently.

While notifications serve as a crucial communication pathway between providers and users, literature suggests the considerable daily total (which can be often around 100) may impair productivity and well-being (Anderson et al. 2018). Further research into optimizing this channel is therefore imperative.

Although existing studies have analyzed behaviors of specific mobile applications or categories (Fischer et al. 2011), a deeper examination of the role and impact of push notifications in mobile communication is warranted.

✶ Rashid Kamal
 kamal-r@ulster.ac.uk

 Aimal Rextin
 aimal.rextin@seecs.edu.pk

 Chris Nugent
 cd.nugent@ulster.ac.uk

 Ian Cleland
 i.cleland@ulster.ac.uk

 Paul McCullagh
 pj.mccullagh@ulster.ac.uk

1 School of Computing, Ulster University, York St, Belfast BT15 1ED, Northern Ireland, UK

2 Center for the Advanced and Smart Technologies, Faculty of Arts, Science and Technology, University of Northampton, Waterside Campus, University Dr, Northampton NN1 5PH, England, UK

3 School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Scholars Ave, 44000 Islamabad, Pakistan

An intriguing hypothesis proposes targeting notifications during boredom, when individuals display increased receptiveness (Pielot et al. 2015). Rather than mere inactivity, boredom represents a search for stimulation (Eastwood et al. 2012). This study tests this hypothesis, investigating whether leveraging idle moments, with the assistance of pervasive computing, could mitigate excessive notification issues.

We employ a three-stage methodology encompassing an initial focus group, large-scale survey, and in-the-wild smartphone data collection to explain the complex relationship between notifications and user behavior during boredom. The systematic approach provides nuanced understanding of associations between boredom and notification engagement, offering critical insights to enhance notification systems.

Specifically, this research makes the following key contributions:

1. How do notification source (app category) and timing influence user engagement, as measured by click-through rates?
2. What notification response patterns can be identified through exploratory analysis of timing features and app categories?
3. How effective are different machine learning algorithms, for the classification task of predicting user notification engagement?

The remainder of this paper is structured as follows:

Section 2 offers an overview of the relevant literature pertinent to this study. It delves into an examination of concepts related to boredom and breakpoint and critically reviews prior research methodologies employed to determine the most opportune moments for interruptions.

Section 3 provides a detailed exposition of the "Seekthe-Notification" app's technical facets and elaborates on the robust methodology utilised in this study.

Section 4 presents the study's findings along with the analysis of the results. Subsequently, Sect. 5 provides discussion of these findings and outlines potential avenues for future research. This approach aims to contribute significant insights and constructive contributions to the ongoing discourse on the optimization of notification management.

2 Related work

Mobile notifications have become an integral yet distracting part of everyday life. Constant notifications can lead to interruption overload and make it difficult for users to concentrate on tasks (Czerwinski et al. 2004; Adamczyk and Bailey 2004; Horvitz 2001). Excessive notifications on

mobile devices are a source of frustration for many users (Leiva et al. 2012).

To address the issue of notification overload, researchers have proposed various user-aware notification systems that aim to deliver notifications at optimal times based on user context (Mehrotra et al. 2015; Mehrotra and Musolesi 2020; Fisher and Simmons 2011; Pielot et al. 2015, 2017; Anderson et al. 2018). These systems incorporate physical activity recognition, user behavior modeling, and machine learning techniques (Pielot et al. 2015, 2017). However, current systems still have relatively low accuracy, with many reporting 70–80% correct prediction rates (Mehrotra and Musolesi 2020). The lack of large-scale studies and datasets has been identified as a key limitation in developing more robust context-aware notification systems (Mehrotra and Musolesi 2020).

Features like "do-not-disturb" mode allow users to limit notifications. However, research indicates that completely blocking notifications often leaves users unsatisfied and feeling like they are missing out on information (Alt and Boniel-Nissim 2018; Wolniewicz et al. 2018). This phenomenon is linked to the dispositional trait of fear of missing out (FoMO), which refers to anxiety over missing rewarding experiences that others may be having (Wolniewicz et al. 2018). Constant social media feeds and real-time updates experienced by tech-savvy users can exacerbate FoMo (Liao and Shyam Sundar 2022).

Prior work has explored using states of user boredom as opportunities to deliver notifications, based on findings that bored users tend to be more receptive (Pielot et al. 2015). However, research on detecting and leveraging states of boredom is still limited. The "behaviorome" concept suggests that combinations of digital behavioral markers from ambient and wearable sensors could provide insight into user psychosocial states like boredom (Cook 2020; Cook and Schmitter-Edgecombe 2021). But concrete methodologies for inferring boredom remain to be developed and validated.

In terms of responding to notifications, research indicates the timing matters. The concept of "break points," when users switch between tasks, has been identified as a key factor influencing notification acceptance (Okoshi et al. 2014, 2017; Pielot et al. 2014). The overall notification environment, not just individual notifications, affects user behavior (Turner et al. 2019). Mood and usage patterns also influence responsiveness (Heinisch et al. 2022).

A key limitation of prior work is that most studies do not directly account for user perceptions and preferences when designing notification management systems. The subjective user experience is a critical factor that should inform the development of intelligent notification techniques. To

address this gap, our work first conducts a focus group study to understand user perspectives surrounding smartphone use and notifications, specifically investigating usage habits and feelings during boredom. The outcomes inform the design of a survey to probe user perceptions of notifications during bored states. By incorporating user thoughts and context, we aim to develop notification management approaches aligned with subjective user needs.

In our study, we take a holistic approach to investigate mobile notification impact, rather than focusing on individual applications. We conducted a survey of 106 participants and a 3-month notification tracking study of 20 individuals. We developed an application called "SeektheNotification" (SN) based on a research-in-the-large approach (Henze et al. 2011; McMillan et al. 2010; Henze and Pielot 2013) and installed it on participants' Android phones to collect notification data. This study provides insights into the overall impact of notifications on users, informing more effective notification management strategies tailored to user perceptions and preferences.

3 Methodology

The design of our investigation was multifaceted, following ethical regulations. Approval was secured from the COMSATS University research committee to guarantee all procedures complied with institutional and legal parameters. Since the application was installed on the volunteers' personal phones, the goal was to amass data whilst minimally

disrupting the user to prevent them from abandoning the study. Therefore, the application was engineered to operate in the background, minimising disturbance and ensuring user acceptance through limited data sharing. Consequently, we narrowed our data collection to notification-related information, unlike other broader studies (Henze et al. 2011; McMillan et al. 2010; Henze and Pielot 2013).

Our methodology was initiated with a focus group study designed to acquire a preliminary comprehension of user attitudes towards notifications during periods of leisure or boredom. This provided a robust foundation for the subsequent steps. We then distributed a survey using a Likert scale, implemented via Google Forms, to explore mobile phone usage patterns across varying times and days, and to discern the most frequently used app types. This widened our understanding of user behaviors and app usage.

We take a holistic approach to investigate mobile notification impact, rather than focusing on individual applications. We conducted a survey of 106 participants and a 3-month notification tracking study of 20 individuals. We developed an application called "SeektheNotification" (SN) based on a research-in-the-large approach (Henze et al. 2011; McMillan et al. 2010; Henze and Pielot 2013) and installed it on participants' Android phones to collect notification data. This provides insights into the overall impact of notifications on users, informing more effective notification management strategies tailored to user perceptions and preferences.

We deployed a mobile application, the SN app, to non-intrusively capture in-the-wild data regarding notifications

Fig. 1 This figure illustrates the sequential stages of our work, commencing with the focus group study and its outcomes shaping the survey questions. The analysis of survey questions subsequently informed the design of the SeektheNotification Android app, which in turn provided the dataset for classification. Additionally, we conducted in-depth analyses of the survey data, app data, and classification results
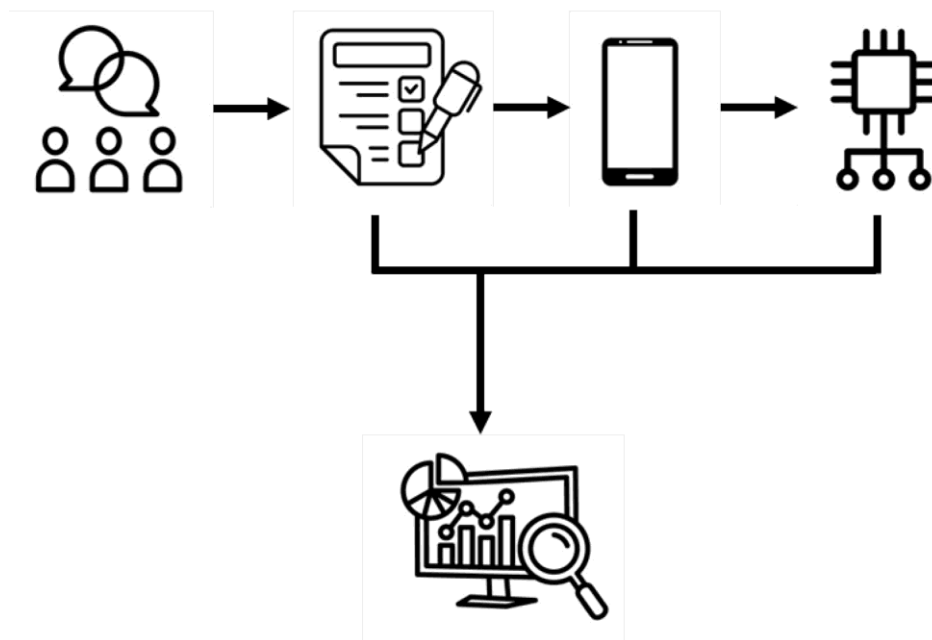
and user responses. The amassed data was subjected to a preprocessing phase, where new features were extracted and selected using keywords in the notification titles to categorize apps. Figure 1 presents a visual representation of the step-by-step progression in our research.

3.1 Focus group study

A focus group study was conducted with the primary aim of gathering an in-depth understanding of participants' attitudes and perceptions towards mobile notifications. This formative research informed subsequent phases. A 12-person group was carefully selected, comprising equal numbers of males and females, students and staff. This heterogeneity provided varied perspectives to effectively test the research questions. The study comprised open-ended discussion centered on two themes: feelings of boredom and preferred activities. A skilled moderator guided the dialogue to maintain focus and encourage participation. The 60-min study allowed sufficient time to explore the topics without fatigue. Resulting qualitative data was analyzed to identify trends related to mobile usage during boredom.

Key insights included use of phones for non-communication purposes and favored app categories when bored. These discoveries informed survey design to quantitatively probe usage habits and attitudes. Focus group patterns and themes guided crafting targeted questions capturing agreement with using phones beyond communication when bored, timing/frequency of boredom, and specific boredom activities like gaming and social media.

Incorporating focus group outcomes in the survey enabled complementary qualitative and quantitative data to examine mobile usage during boredom from multiple angles. The focus group furnished essential preliminary comprehension of user perceptions, laying the groundwork for investigating the research questions through subsequent quantitative methods.

The questions presented in Table A1 were crafted based on the insights gained from the focus group study. This allowed quantitative validation of the qualitative trends identified.

3.2 Recruitment

Recruitment utilized email invitations to computer science students and faculty at COMSATS University, supplemented by social media campaigns. Demographic data was not collected to protect privacy. The sample comprised 106 survey responses and 24 participants providing notification data by installing the SN app. Participants could voluntarily pause, resume, or withdraw during the study. Rigorous steps ensured data confidentiality, including thorough anonymization, validation, verification, and secure university server storage prior to analysis.

Before participation, consent forms outlined the study purpose and expectations. Participants could ask questions and raise concerns. This recruitment approach furnished an appropriately sized sample while upholding ethical standards around informed consent and data privacy.

3.3 Survey

Participants completed a Google Forms questionnaire detailed in Appendix employing 5-point Likert scale questions. This surveyed mobile usage patterns across times and days to investigate engagement habits with device types and frequency.

Additionally, the survey aimed to elucidate the relationship between usage and boredom. Rather than viewing boredom as a purely psychological state, it was conceptualized as a liminal state characterized by passing time through mobile engagement.

Probing usage behaviors and attitudes towards boredom-linked engagement through mobile engagement, for the purposes of this research, provided quantitative data complementing focus group insights. The survey furnished vital understanding of usage habits across app categories, times, days, and subjective mindsets. In Appendix, a detailed presentation of the specific Likert scale questions administered through Google Forms to achieve these research objectives can be found. These questions were designed to delve deeper into participants' experiences and preferences related to mobile phone usage during moments of leisure or boredom, contributing to an understanding of their behaviors and attitudes.

Algorithm I An algorithm for obtaining notification data

```
Require: pN ≥ packageNames
Require: packageName = sbn.getPackageName()
Ensure: AndroidVersion >= LOLLIPOP
   reactionTime = sbn.getPostTime()
   time = System.currentTimeMillis()
   usageStats ← appList
   currentApp ← mySortedMap.get(mySortedMap.lastKey()).getPackageName()
   foregroundTP ← foregroundTaskInfo.topActivity.getPackageName()
   while foregroundTP ≠ 0 do
      if currentApp == packagName then
         ApplicationName ← title
         NotificationPostedTime ← reactionTime
         NotificationReactionTime ← time
         Reaction ← 1
      end if
      ApplicationName ← title
      NotificationPostedTime ← reactionTime
      NotificationReactionTime ← time
      Reaction ← 0
end while
```

3.4 SeektheNotification mobile app

The SN app operates via the NotificationSeeker service for unobtrusive background data collection on Android 6.0+.

Notification data is gathered from all installed apps including title, timestamps, and user responses through a proprietary algorithm. This tracks reactions to notifications from one app while also recording broader cross-app usage.

Algorithm 1 outlines the approach. It defines essential variables like package name and timestamps, initializes the app list, and identifies the target app. A while loop tracks the foreground app, checking if it matches the target. If so, it records the app name, notification time, response time, and reaction flag. Even when unmatched, it logs the app and times without a reaction flag.

This dual recording of target app and holistic usage enables analyzing notification responses across applications. The custom Android build and proprietary algorithm provide controlled, precise in-the-wild data collection.

In summary, the SN app unobtrusively gathers multifaceted notification data leveraging a tailored algorithm. By tracking reactions across applications, it affords the view of real-world user behaviors and responses to notifications. The app acts as an instrumental vehicle for gathering the observational data fundamental to investigating the research questions.

3.4.1 SeektheNotification dataset description

The SN app has been specifically designed to gather data regarding notifications, including details such as the notification's title, the source app that generated the notification, the time when the notification was posted, the timestamp at which the Android operating system initiated the notification, and the user's response time. Notably, the user's reaction time data encompasses whether the user accepted or

Table I An example of the data collected by SeektheNotification

NotificationPostedTime	ApplicationName	UserResponseTime	User response[a]
1494783060	Messenger	1494786420	0
1494786420	WhatsApp	1494786660	1
1494786420	WhatsApp	1494786660	1
1494787380	WhatsApp	1494787680	1
1494787380	Messenger	1494787680	1
1494787380	Instagram	1494787680	1
1494784920	Messenger	1494787680	0
1494786600	Facebook	1494787680	0
1494788760	Instagram	1494789300	0
1494789600	Instagram	1494790620	0

[a] A User Response of 0 indicates the removal of the notification, while a value of 1 signifies that the notification was actively engaged with, specifically through clicking

Table 2 The final set of features after the preparation and selection of features (AppGroup refers to Application Category, NPD refers to Notification Posted Day, NPTD refers to Notification Posted Time of Day (i.e. Morning, Afternoon), URD refers to User Response Day, URTD refers to User Response Time of the Day, URPT refers to User Response Time towards the notification, NPT refers to Notification posted time, User Delay refers to Delay from the posting time to reaction time and finally UserResponse refers to either user clicked the notification or remove the notification

AppGroup	NPD	NPTD	URD	URTD	URPT	NPT	UserDelay	User response
2	6	5	6	5	18.07	17.11	0.96	1
2	6	5	6	5	18.11	18.07	0.039	1
2	6	5	6	5	18.11	18.07	0.039	0
7	6	5	6	5	18.28	18.23	0.05	1
2	6	5	6	5	18.28	18.23	0.05	1
3	6	5	6	5	18.28	18.23	0.05	0

declined the notification, a process that was systematically recorded using Algorithm 1.

The data presented in Table 1 serves as a valuable resource for gaining insight into the notifications received by study participants. This dataset is pivotal for facilitating in-depth analysis and interpretation of the study's findings (Table 2).

3.5 Data preprocessing

The dataset utilised for this investigation comprised data from 20 individuals, as opposed to the initially intended 24 participants, due to the fact that four participants withdrew from the study and failed to furnish any relevant data. The data collected in this study was grouped by application using a keyword-based method, whereby the notification title was used as the primary keyword. As a result of this approach, various categories of applications were identified:

1. Health
2. Personal
3. Social
4. Work
5. Entertainment
6. System
7. Other

The categorisation of the applications was determined by referring to the categories provided by the Google Play Store. Additionally, to further refine the Personal Apps category, input was solicited from study participants on which specific apps they considered to be personal in nature. Based on this feedback, the apps Phone, Messages, and Whatsapp were ultimately included in the Personal Apps category.

The categorization of applications into distinct app groups was a critical component of the methodology, as it enabled the analysis of how notification source impacts user engagement. Specifically, by calculating the click-through rate for each app group, defined as the proportion of notifications received that were actively clicked by users, variability in notification success across sources could be quantitatively examined. The app group categorization provided the foundation for this analysis by dividing notifications into relevant sources. Visualizing the click-through rates for each resulting app group in graphical form then facilitates clear comparison of how different notification sources influence user behavior. This analysis, presented visually rather than through statistical hypothesis testing, serves as a key contribution by elucidating the relationship between notification source and success. Examining this phenomenon is imperative for comprehending what drives user engagement with notifications in real-world contexts. The outcomes of this examination across the defined app groups are detailed in the subsequent Sect. 4.

The goal was to first describe how the app groups were defined, then transition into how analyzing click-through rates by these groups enables investigating the relationship between source and success.

3.5.1 Feature extraction and selection

We derived four novel features from the NotificationPosted-Time and UserResponseTime data. These features include:

1. Notification posted day (NPD),
2. User response day (URD),
3. Notification posted time of the day (NPTD), and
4. User response time of the day (URPTD).

These new features played a crucial role in conducting a more comprehensive analysis of the collected data.

To enhance the granularity of our analysis, we divided the NPTD and URPTD features into six distinct time categories throughout the day:

1. Late night,
2. Early morning,
3. Morning,
4. Noon,
5. Evening, and
6. Night.

This division allowed us to gain deeper insights into the patterns of mobile phone usage among participants. Additionally, we transformed the NPD and URD features into weekday numbers (e.g., Monday) to enable the potential for independent predictions in the future, irrespective of the specific date on which the data was originally collected. This approach enhances the generalizability and applicability of our findings.

The utilization of these derived features greatly enriched our dataset and contributed to a more comprehensive analysis of notification-related behaviors. For reference, please see Table 1.

3.5.2 Class imbalance

The presence of class imbalance within a dataset can have substantial implications for the accuracy and reliability of resulting predictions. In particular, a pronounced class imbalance in the training dataset can lead to reduced precision and recall for minority classes during testing

(Mazurowski et al. 2008). Moreover, the choice of evaluation metric can yield deceptive outcomes in the presence of class imbalance. For example, a straightforward accuracy metric, computed as the ratio of correct to incorrect predictions, may exhibit bias due to class imbalance. Consequently, high accuracy values for majority classes may be achieved independently of the performance of minority classes, which holds lesser statistical significance.

To mitigate the adverse effects of class imbalance, various techniques can be employed. One such method is the Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al. 2002), which functions by generating new samples through the synthesis of variations from existing samples. This approach serves to rebalance the class distribution, thereby enhancing the classifier's performance on the minority class.

The implementation of these strategies is crucial to ensure that predictive models are both accurate and equitable in their treatment of minority classes. For further details, please refer to the citations (Mazurowski et al. 2008) and Chawla et al. (2002).

3.6 Clustering

One of the objectives was to determine the presence of distinct clusters within the dataset based on various features.
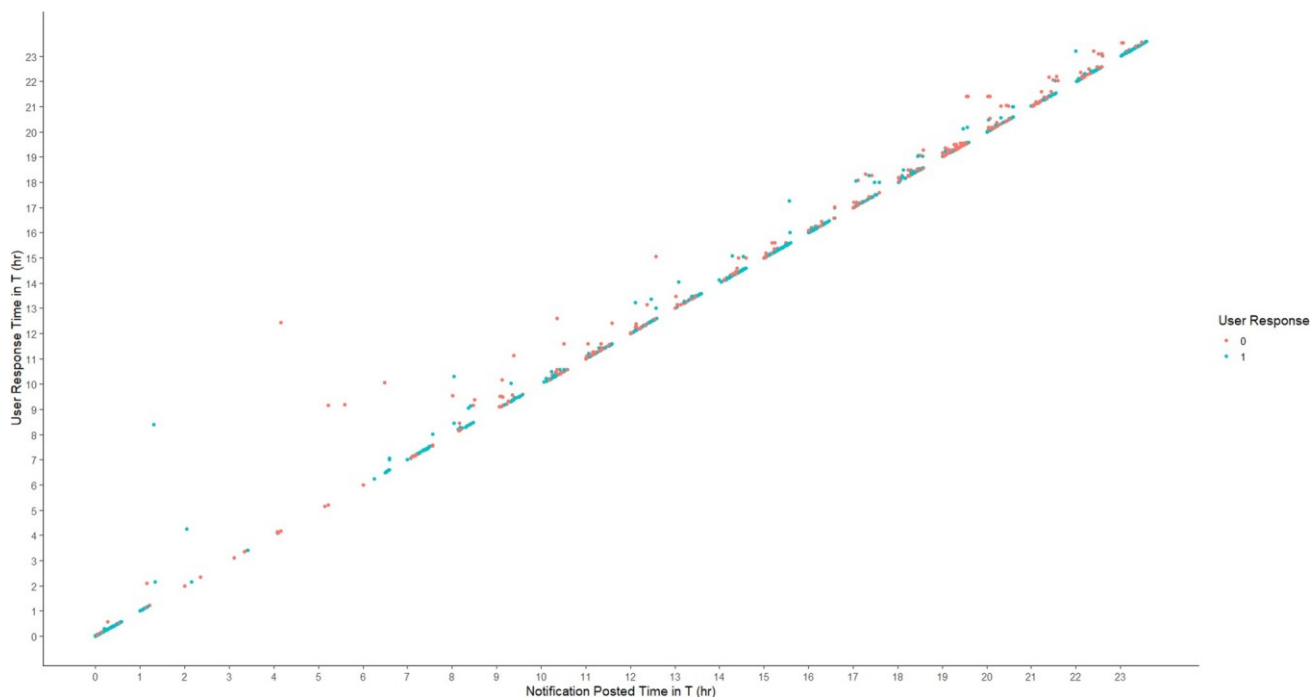

Fig. 2 Data from App Data: the *X*-axis represents the Notification posted hour and the *Y*-axis represents the Notification reaction hour. The color of the point represents the application clicked or ignored. It is interesting to note that when a user ignores a notification, they usually do not do so immediately

Clear evidence of clustering emerged through iterative trials across these features, as elaborated in Sect. 4.2.

The dimensionality reduction technique t-SNE was utilized to achieve this aim. t-SNE is widely acknowledged for visualizing high-dimensional data exhibiting inherent structure like clusters.

The algorithm transforms high-dimensional data into a lower-dimensional representation, typically 2–3 dimensions, while optimally preserving relative data point distances. This enables visualizing patterns not readily apparent in higher dimensions.

In summary, t-SNE afforded critical reduction of the notification dataset into lower dimensions that surfaced underlying clusters undetectable in the native high-dimensional space. By leveraging this technique, intrinsic data structure was unveiled, laying the groundwork for targeted analysis of relationships between key features. The emergence of clusters facilitated nuanced investigation of research questions surrounding notification categories, timing, and user behavior (Fig. 2).

3.7 Classification

In order to classify notification acceptance, we conducted an empirical comparison of several standard classification techniques. We employed fourteen widely used classifiers on the dataset, including ensemble-based methods, linear classifiers, nonlinear classifiers, boosting algorithms, non-parametric methods, probability machine learning, neural networks, and stochastic approaches. As part of the ensemble-based learning approach, we used RFs (RF) (Breiman 2001), which combines multiple decision trees to form a forest and improve the overall accuracy and robustness of the model. Extra Trees (ET) (Geurts et al. 2006) is an extension of RF that randomly samples the feature space at each split, leading to more decorrelated trees. Decision Trees (DT) (Myles et al. 2004) are a simple yet powerful method for classification based on a single decision tree. For linear classifiers, we used Logistic Regression (LR) (Hastie et al. 2001), which models the relationship between a dependent variable and one or more independent variables using a logistic function. Support Vector Machines with Radial Basis Functions kernels (SVM) (Vapnik 1995) is a nonlinear classifier that uses the concept of decision boundaries to separate different classes.

Gradient Boosting (Natekin and Knoll 2013) is a boosting algorithm that improves the accuracy of weak models by combining multiple models. K nearest neighbour (Cover and Hart 1967) is a non-parametric, supervised learning classifier that uses the concept of distance-based classification. Naive Bayes algorithm (NB) is a probability machine learning technique that can be applied to a variety of classification

functions. An extension to the Naive Bayes algorithm is the Gaussian Bayes algorithm (GNB). Neural Network is a layer-based approach that uses artificial neural networks to model complex relationships between inputs and outputs. Quadratic Discriminant Analysis (QDA) (using Bayes'rule and fitted class conditional densities) generates a quadratic decision boundary, with each class fitted with a Gaussian density. SGDClassifier is a stochastic approach that uses gradient descent to optimize the parameters of a model. As a machine learning library for gradient-boosted decision trees (GBDT), XGBoost (Shwartz-Ziv and Armon 2022) is widely used on tabular datasets and provides parallel tree boosting. We were interested in investigating how XGBoost would perform on this dataset, as it is also an industry-leading machine learning library for regression, classification, and ranking problems. We conducted an empirical comparison of a wide range of classification techniques based on a variety of problem settings, and a total of 14 widely used classifiers from a variety of domains were examined.

4 Results and discussion

In this section, we undertake a rigorous examination of the dataset obtained from surveys and notifications, partitioning the analysis into three distinct components.

Firstly, we delve into the survey data to comprehend participants' perceptions and attitudes toward notifications. Employing a diverse array of statistical techniques, including descriptive statistics, inferential statistics, and visualization methods, we seek to uncover discernible patterns and trends within the dataset. The overarching aim of this analysis is to provide valuable insights into how participants perceive and engage with notifications.

Secondly, we conduct an in-depth analysis of the notification data to unveil inherent patterns and relationships within the dataset. This phase of analysis involves a range of techniques, including data preprocessing, feature extraction, dimensionality reduction, and visualization methods. Our objective here is to illuminate the patterns and trends governing the usage and responses to notifications by participants.

Lastly, we assess the performance of various machine learning models applied to the dataset and present the results of these evaluations. These models are rigorously trained and tested on the data, with their performance evaluated using established metrics such as accuracy, precision, recall, and F1-score. The primary purpose of this analysis is to identify the most effective models for classifying the data, shedding light on the latent data patterns and relationships effectively captured by these models.
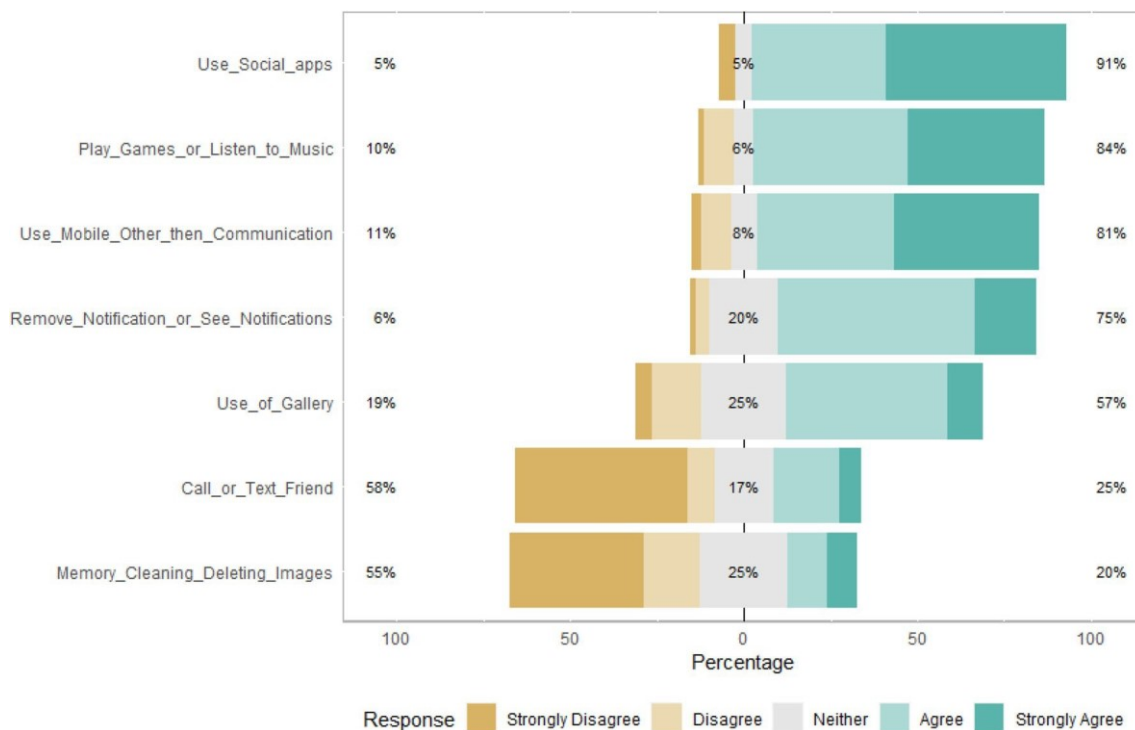
Fig. 3 Data from Survey: results form a questionnaire requesting responses on a Likert scale is shown on the *X*-axis, and questions are shown on the *Y*-axis
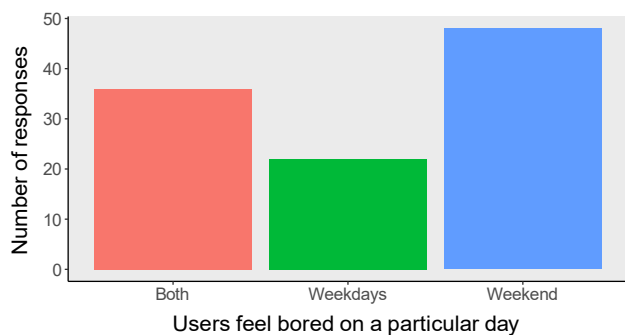

Fig. 4 Data from Survey: Responses to the question that what day do they feel more bored. *X*-axis represents the day of the week and *Y*-axis represents the total number of responses
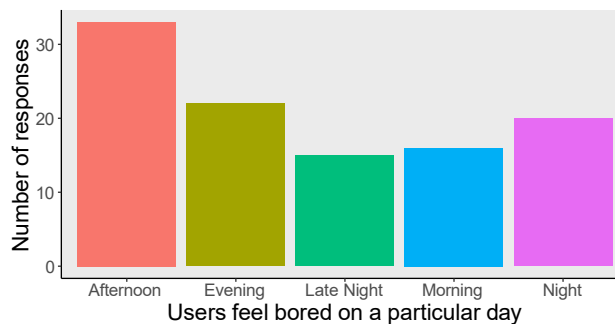

Fig. 5 Data from survey: the *X*-axis shows frequency of responses, while the *Y*-axis shows time of day. During certain times of the day such as (Afternoon, Evening, and Night), users are more likely to feel bored

4.1 Survey

The present study aimed to investigate the behaviour of mobile phone users when experiencing boredom. A total of 106 participants were recruited for the survey, which aimed to identify the activities that individuals engage in to pass the time when bored. The results of the survey, as illustrated in Fig. 3, revealed that the majority of participants reported using social apps as a means of combating boredom. Additionally, a significant proportion of respondents indicated that they engage in activities such as playing games or

listening to music, as well as viewing notifications, when experiencing boredom or having free time.

This study was conducted with the primary objective of examining the behavior of mobile phone users in the context of experiencing boredom. Furthermore, a substantial proportion of respondents reported engaging in activities such as playing games, listening to music, and attending to notifications, particularly when experiencing boredom or during leisure time.
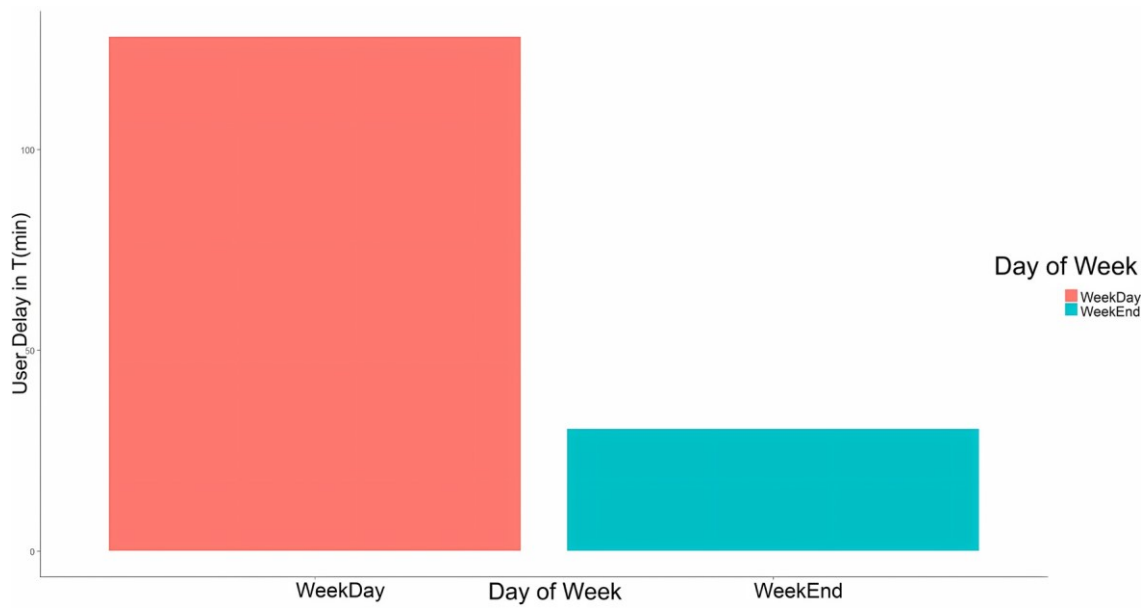
Fig. 6 The *X*-axis represents the day of the week and the *Y*-axis represents the user delay toward the notification in minutes. The user delay toward the notification is shorter on weekends than on weekdays
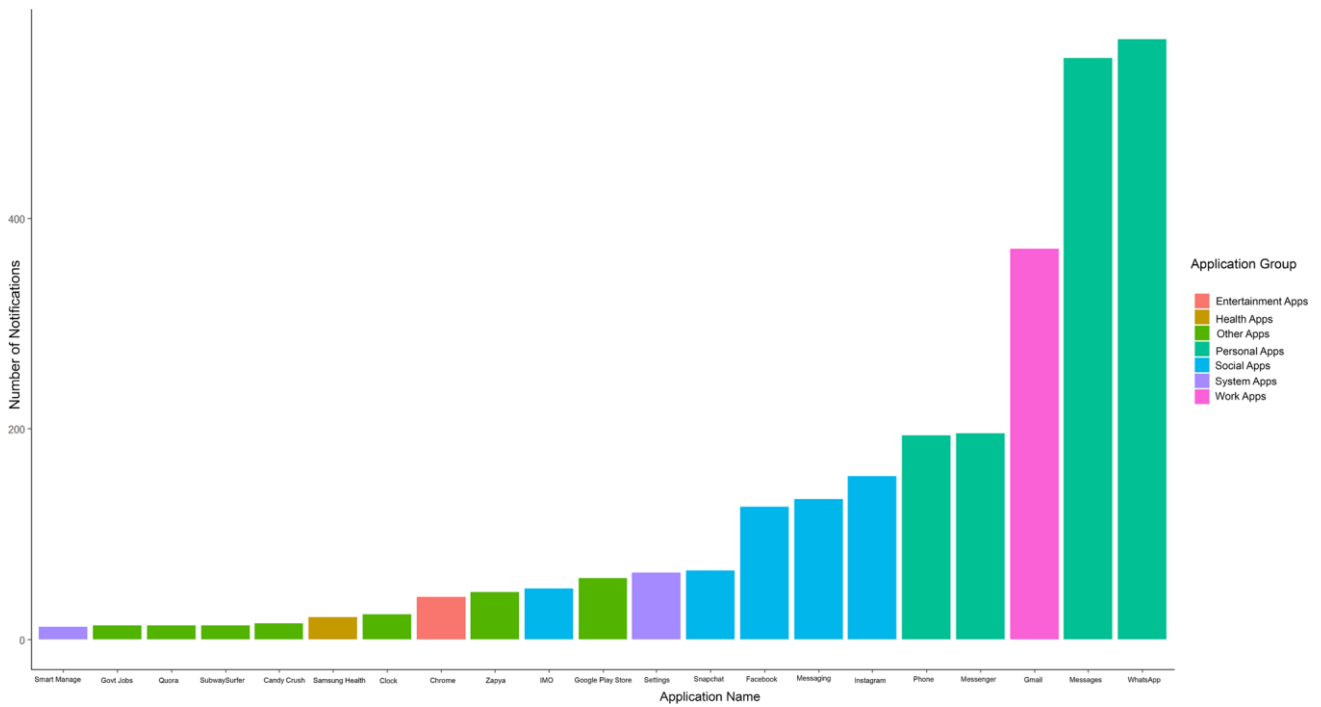

Fig. 7 The *X*-axis represents the name of the application and the *Y*-axis represents the number of notifications. The colour of the graph represents the group of the application. The most notifications were sent by the social application group and the personal application group
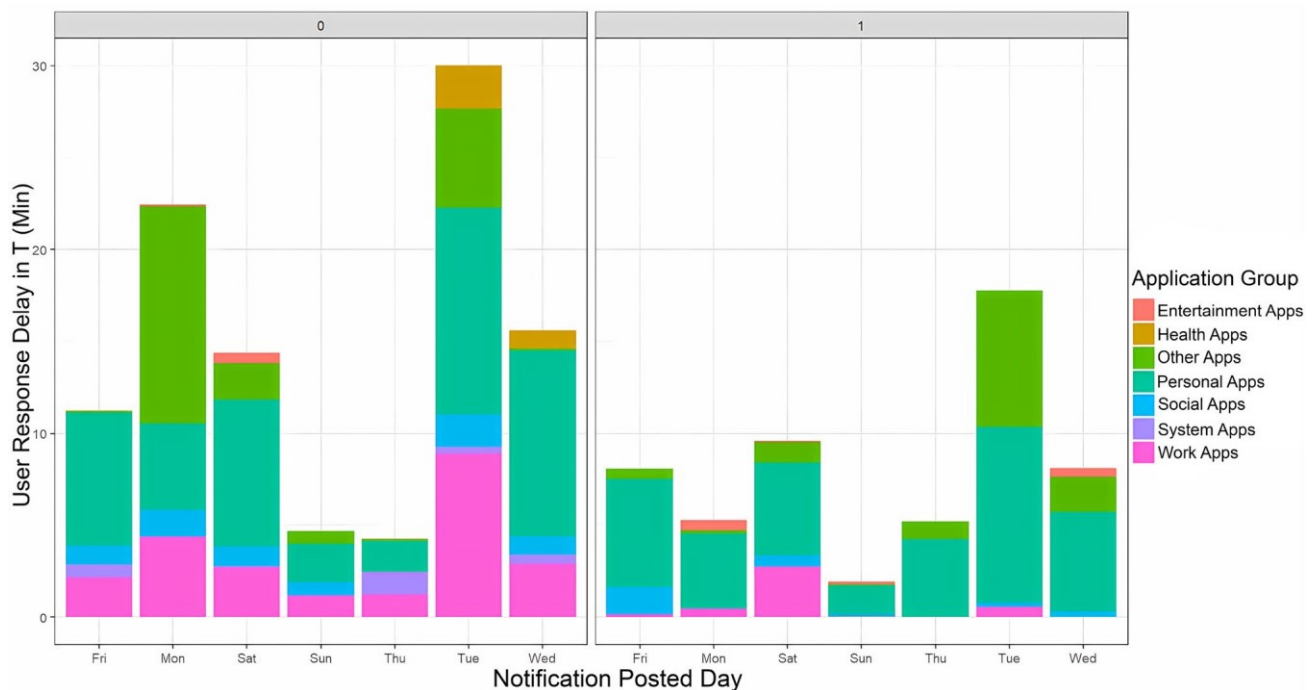
Fig. 8 The left side of the figure represents clicked notifications and the right side represents ignored or removed notifications. The *X*-axis represents the application group and the *Y*-axis represents the average user delay towards the notification in minutes

One intriguing implication derived from the survey findings is the potential for mobile phone notifications to be intelligently designed to detect user boredom and deliver notifications accordingly. Additionally, the results underscore that respondents were less inclined to utilize their mobile phones for direct communication (e.g., calling or texting) when experiencing boredom, highlighting the preference for social and entertainment applications during breaks or leisure periods.

Furthermore, this study revealed that the day of the week and time of day play a noteworthy role in determining user availability and susceptibility to boredom. Figure 4 illustrates a higher incidence of reported boredom on weekends, which may be linked to increased free time or reduced daily activities. Moreover, as demonstrated in Fig. 5, users tend to experience more pronounced episodes of boredom in the afternoon and evening, with a notable surge in boredom occurrences after 12 pm.

In summary, these findings suggest that the nature of mobile phone users' responses to boredom is influenced significantly by their application preferences. Furthermore, the effectiveness of notifications may hinge on their timing, with user receptivity varying according to the time of day and day of the week.

4.2 Notification data analysis

This study was conducted with the overarching goal of delving deeper into patterns and relationships within a dataset pertaining to notifications. To achieve this objective, we employed a dimensionality reduction technique known for its ability to facilitate data visualization in a two-dimensional space, as previously established in the literature (Van der Maaten and Hinton 2008). This technique enabled us to gain a visual perspective on the dataset's intricate structures.

Twenty-four participants were initially recruited for this study, with complete data provided by 20 of them, forming the basis of our analysis.

Our survey data revealed a notable trend: participants often reported feelings of boredom during the weekend. This observation led us to hypothesize that response times to notifications might exhibit corresponding variations during these periods. Our analysis substantiated this hypothesis, highlighting significantly shorter response times to notifications on weekends compared to weekdays (as depicted in Fig. 6).

Another noteworthy insight emerged from our survey data, emphasizing the variability in user attention across different application groups (Fig. 7). Our analysis demonstrated that response times to notifications were influenced by the
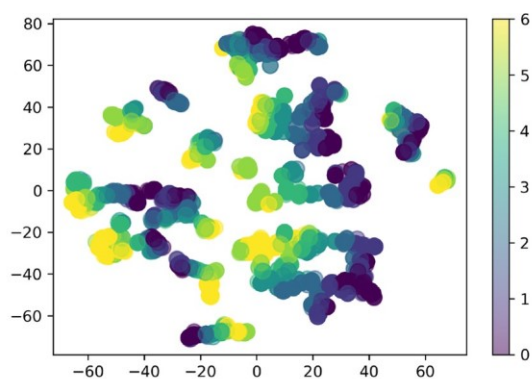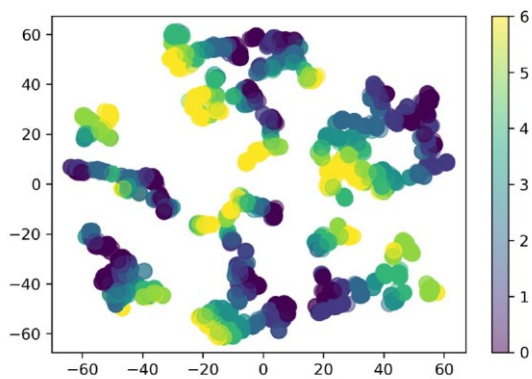
(a) The X and Y axis depict the converted data from high-dimensional to low-dimensional space. Clusters are colour-coded based on the notification posted day
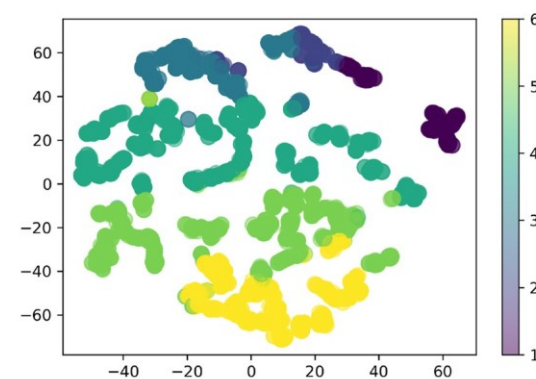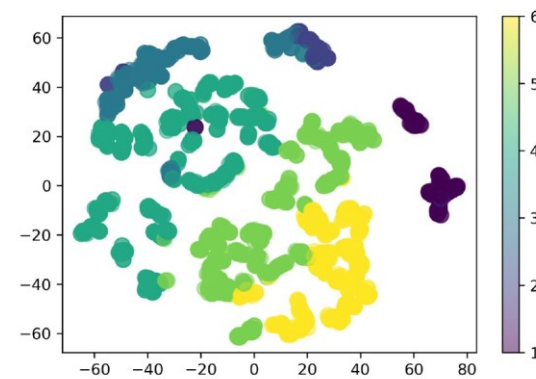
(b) The X and Y axis depict the converted data from high-dimensional to low-dimensional space. Clusters are colour-coded based on the notification posted time of the day

Fig. 9 Comparison of visualisation of the data in a two-dimensional space using t-SNE dimensionality reduction. The left side of the figure depicts the clusters formed by the notification posted day with the scale ranging from 0 to 6 indicating the day of the week (Monday as 0, Tuesday as 1), while the right side of the figure depicts the clusters formed by the notification posted time of the day with the scale ranging from 0 to 6 indicating time of the day (Late Night as 1, Early Morning as 1)

(a) The X and Y axis depict the converted data from high-dimensional to low-dimensional space. Clusters are colour-coded based on the notification reaction day, with the scale ranging from 0 to 6 indicating the day of the week (Monday as 0, Tuesday as 1).

(b) The X and Y axis depict the converted data from high-dimensional to low-dimensional space. Clusters are colour-coded based on the notification reaction time of the day, with the scale ranging from 0 to 6 indicating time of the day (Late Night as 1, Early Morning as 1).

Fig. 10 Comparison of Visualisation of the data in a two-dimensional space using t-SNE dimensionality reduction. The left side of the figure depicts the clusters formed by the notification reaction day, while the right side of the figure depicts the clusters formed by the notification reaction time of the day
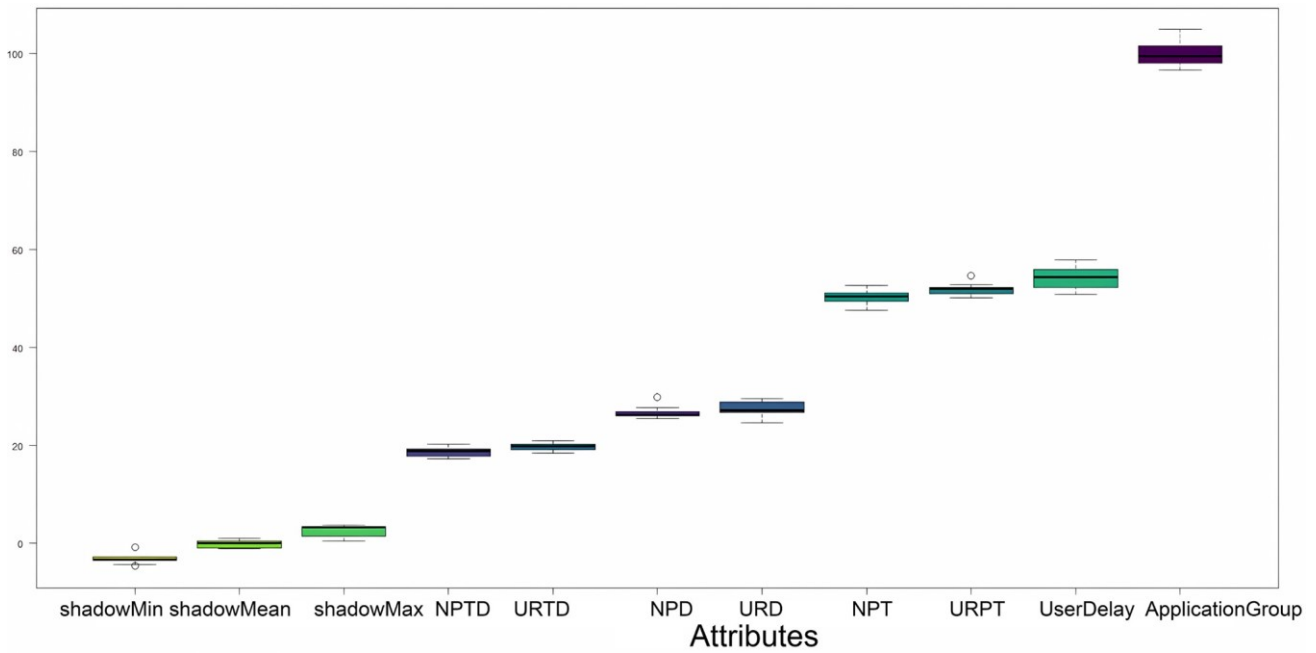
Fig. 11 The *X*-axis of the graph represents the names of the features under consideration, while the *Y*-axis depicts the corresponding values for feature importance, as computed through the implementation of the Boruta algorithm
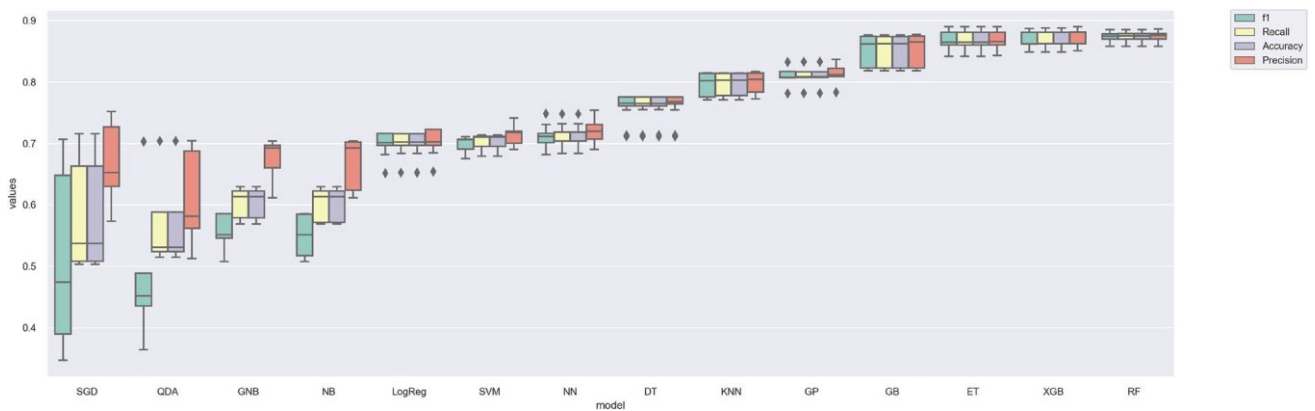

Fig. 12 The *X*-axis of the plot displays the names of the classifiers used in the analysis. Each group of four box plots corresponds to a single classifier, with each individual box plot labeled accordingly. The color scheme used in the plot is as follows: the green box plot represents the f1-score, the yellow box plot corresponds to recall, the purple box plot represents accuracy, and the red box plot is associated with precision. The *Y*-axis denotes the values of the respective classification scores

source of the notification. As illustrated in Fig. 8, discernible differences in delay times were observed depending on whether users engaged with the notification (by clicking or viewing) or chose to ignore it. Within our dataset, health apps garnered the shortest delay when the notification was engaged with, followed by work apps, personal apps, and social apps, in that order. Remarkably, system app notifications were not engaged with by any of our study participants.

Regarding notification frequency, WhatsApp and Messages emerged as the leading applications in terms of sending notifications. The 91% of notifications originated from the Social and Personal app groups, as evidenced by Fig. 7.

Additionally, our analysis unveiled an intriguing finding: the time of day at which notifications were posted formed more distinct clusters compared to the day of the week on which notifications were sent, as depicted in Fig. 9. A similar pattern emerged concerning the time of day when users responded to notifications, compared to the day of the week when the responses were received, as

shown in Fig. 10. These findings suggest that user interactions with notifications are influenced by specific temporal patterns.

In summary, our study yields valuable insights into user interaction patterns with various application notifications. The weekend, characterized by reported feelings of boredom, was associated with significantly shorter response times to notifications, suggesting heightened user engagement during these periods. Notably, user attention exhibited variability across different application groups, with health apps receiving the swiftest responses. The prevalence of notifications from WhatsApp and Messages, along with the Social and Personal app groups, underscores their substantial contributions to users' notification influx. Furthermore, our findings emphasize the potential importance of time of day in shaping user interaction with notifications, surpassing the significance of the day of the week. This understanding of user behavior can serve as a foundation for refining notification delivery strategies, ultimately enhancing user engagement and satisfaction.

4.3 Classification results

This study endeavors to address a binary classification problem by deploying a diverse array of algorithms, establishing performance benchmarks. In particular, we harnessed the scikit-learn implementations of 14 algorithms, as elaborated by Pedregosa et al. (2011). These algorithms were rigorously applied to a dataset that underwent meticulous preprocessing and feature extraction, resulting in the utilization of a concise set of eight features.

Our dataset initially exhibited an imbalance between the positive 60.26% and negative 39.74% classes. To rectify this class imbalance, we employed the Synthetic Minority Oversampling Technique (SMOTE), as elucidated in Sect. 3.5.2. Additionally, we conducted feature ranking using the Boruta algorithm, which operates as a wrapper around a Random Forest (RF) classification algorithm, as introduced by Kursa and Rudnicki (2010). Notably, based on Boruta's feature ranking analysis, the 'AppGroup' feature emerged as the most influential, followed by 'User Delay,' 'User Response Time,' 'Notification posted time,' and the 'Details of the feature.' Comprehensive details of the feature rankings are available in Fig. 11.

To facilitate the comparative evaluation of model performance, we assessed the F1, precision, recall, and accuracy scores achieved by each model on a reserved test set. As showcased in Fig. 12, our results highlight the Random Forest (RF) model as the top performer, closely followed by the XGBoost model, which exhibited a slightly higher precision score. In contrast, the Stochastic Gradient Descent (SGDClassifier) model demonstrated the least favorable performance.

This paper contributes insights into the contemporary challenges associated with notification management systems, emphasizing the potential of detecting user boredom as a means to address the issue of notification overload. While our study primarily focuses on boredom as a primary factor, it is pertinent to acknowledge the potential for expansion to encompass a broader spectrum of psychological contexts and leverage advanced deep learning techniques. Our analysis of key features and the evaluation of machine learning algorithms underscore the efficacy of these methodologies in addressing binary classification problems in the domain of notifications. It is imperative, however, to confront challenges like class imbalance and feature significance, as aptly addressed through techniques such as SMOTE and feature ranking.

Our findings prominently underscore the promising performance of the Random Forest and XGBoost models in this dataset, positioning them as robust candidates for future notification management systems. Nevertheless, the potential of these models could be further harnessed by extending their scope beyond boredom to encompass a more comprehensive understanding of user context. In summary, our preliminary investigation accentuates the potential of machine learning techniques in crafting intelligent notification systems attuned to user states. However, it is incumbent upon future research to explore broader psychological constructs beyond boredom. Our findings lay the groundwork for forthcoming endeavors, seeking to amalgamate rich user context and advanced algorithms, thus advancing the trajectory towards more sophisticated interruption management systems designed to alleviate notification overload.

5 Conclusion and future work

In this study, we conducted in-the-wild data collection involving 20 volunteers over a duration of up to 10 days using a custom-built Android application, SN, from which we gathered 2265 notifications. Our key findings illuminate that notifications derived from health, personal, and social apps exhibit the highest engagement levels. Nonetheless, as delineated in Fig. 8, it is evident that a significant proportion of notifications were deemed either irrelevant or distracting, thereby interfering with the users' tasks and diminishing

their efficiency. Such repeated disruptions, we hypothesize, could exacerbate the mental load of users, thereby further degrading their productivity.

Our investigation into the optimum time for notification delivery, presented in Fig. 2, suggests that user delay in response to notifications noticeably drops after 12 pm, as compared to earlier times in the day. Interestingly, the impact of the day of the week on user delay is also non-negligible, with Fig. 6 demonstrating lower user delays on weekends than on weekdays.

We also endeavored to discern the critical features of a notification management system. Utilizing the Boruta algorithm (Kursa and Rudnicki 2010) as a wrapper for the RF classification algorithm, we identified 'AppGroup' as the most consequential feature, followed by 'User Delay', 'User Response Time', 'Notification posted time', and other feature details, as showcased in Fig. 11.

To assess various classification models' efficacy, we compared their F1, precision, recall, and accuracy scores using a held-back test set. As per the results displayed in Fig. 12, the RF model outperformed other models, with the XGBoost model trailing slightly behind, albeit exhibiting marginally superior precision. However, the SGDClassifier model displayed subpar performance.

In conclusion, the present study underscores the need to factor in both psychological and situational contexts while developing an effective notification management system. Future work will seek to incorporate phone event data, including battery life, phone usage, and sensor data from the accelerometer and gyroscope, to refine the behavioral model's accuracy further. In addition, we envisage employing a machine learning model to identify the optimal times for delivering notifications or interrupting users. Moreover, we recommend designing notification management systems to prioritize notifications and only interrupt users upon reaching specific "break points".

Appendix: Survey questions

This appendix presents a comprehensive account of the survey questionnaire and the corresponding options offered to the participants in response to the queries posed (Table 3).

Table 3 Survey questions

Question	Answer
Do you agree with the statement that you use mobile phone other than communication purpose when you get bored	Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree
When are you more likely to get bored	Weekend, Weekday, and Both
What time of day are you likely to feel bored	Morning (Timing 06:00 AM–11:59 AM), Afternoon (Timing 12:00 PM–5:59 PM), Evening (Timing 06:00 PM–08:59 PM), Night (Timing 09:00 PM–11:59), and Late Night (Timing After 12 PM)
To which extend do you agree with the following statements when you get bored. Use of gallery when you get bored?	Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree
To which extend do you agree with the following statements when you get bored. Remove notifications or see notifications from notification centre?	Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree
To which extend do you agree with the following statements when you get bored. Play Games or listen to music?	Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree
To which extend do you agree with the following statements when you get bored. Use Facebook, Instagram etc	Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree
To which extend do you agree with the following statements when you get bored. Call or text a friend	Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree
To which extend do you agree with the following statements when you get bored. Memory cleaning (deleting images, songs or text) when you get bored or nothing to do	Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Adamczyk, P.D., Bailey, B.P.: If not now, when? the effects of interruption at different moments within task execution. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 271–278 (2004)

Alt, D., Boniel-Nissim, M.: Links between adolescents' deep and surface learning approaches, problematic internet use, and fear of missing out (fomo). Internet Interv. **13**, 30–39 (2018)

Anderson, C., Hübener, I., Seipp, A.-K., Ohly, S., David, K., Pejovic, V.: A survey of attention management systems in ubiquitous computing environments. Proc. ACM Interact. Mobile Wearable Ubiquit. Technol. **2**(2), 1–27 (2018)

Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)

Cook, D.: Digitally mapping the human behaviorome. In: 2020 IEEE International Conference on Pervasive Computing and Communications (PerCom), p. 1. IEEE (2020)

Cook, D.J., Schmitter-Edgecombe, M.: Fusing ambient and mobile sensor features into a behaviorome for predicting clinical health scores. IEEE Access **9**, 65033–65043 (2021)

Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE Trans. Inf. Theory **13**(1), 21–27 (1967). https://doi.org/10.1109/TIT.1967.1053964

Czerwinski, M., Horvitz, E., Wilhite, S.: A diary study of task switching and interruptions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 175–182 (2004)

Eastwood, J.D., Frischen, A., Fenske, M.J., Smilek, D.: The unengaged mind: defining boredom in terms of attention. Perspect. Psychol. Sci. **7**(5), 482–495 (2012)

Fischer, J.E., Greenhalgh, C., Benford, S.: Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications. In: Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, pp. 181–190 (2011)

Fisher, R., Simmons, R.: Smartphone interruptibility using density-weighted uncertainty sampling with reinforcement learning. In: 2011 10th International Conference on Machine Learning and Applications and Workshops, vol. 1, pp. 436–441. IEEE (2011)

Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. Mach. Learn. **63**(1), 3–42 (2006)

Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer Series in Statistics, Springer, New York, NY, USA (2001)

Heinisch, J.S., Gao, N., Anderson, C., Deldari, S., David, K., Salim, F.: Investigating the effects of mood & usage behaviour on notification response time. arXiv preprint arXiv:2207.03405 (2022)

Henze, N., Pielot, M.: App stores: external validity for mobile hci. Interactions **20**(2), 33–38 (2013)

Henze, N., Pielot, M., Poppinga, B., Schinke, T., Boll, S.: My app is an experiment: experience from user studies in mobile app stores. Int. J. Mobile Hum. Comput. Interact. (IJMHCI) **3**(4), 71–91 (2011)

Horvitz, E.C.M.C.E.: Notification, disruption, and memory: Effects of messaging interruptions on memory and performance. In: Human–Computer Interaction: INTERACT, vol. 1, p. 263 (2001)

Kursa, M.B., Rudnicki, W.R.: Feature selection with the boruta package. J. Stat. Softw. **36**, 1–13 (2010)

Leiva, L., Böhmer, M., Gehring, S., Krüger, A.: Back to the app: the costs of mobile application interruptions. In: Proceedings of the 14th International Conference on Human–computer Interaction with Mobile Devices and Services, pp. 291–294 (2012)

Liao, M., Shyam Sundar, S.: Sound of silence: does muting notifications reduce phone use? Comput. Hum. Behav. **134**, 107338 (2022)

Mazurowski, M.A., Habas, P.A., Zurada, J.M., Lo, J.Y., Baker, J.A., Tourassi, G.D.: Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance. Neural Netw. **21**(2–3), 427–436 (2008)

McMillan, D., Morrison, A., Brown, O., Hall, M., Chalmers, M.: Further into the wild: running worldwide trials of mobile systems. In: International Conference on Pervasive Computing, pp. 210–227. Springer (2010)

Mehrotra, A., Musolesi, M.: Intelligent notification systems. Synth. Lect. Mobile Perv. Comput. **11**(1), 1–75 (2020)

Mehrotra, A., Musolesi, M., Hendley, R., Pejovic, V.: Designing content-driven intelligent notification mechanisms for mobile applications. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 813–824 (2015)

Myles, A.J., Feudale, R.N., Liu, Y., Woody, N.A., Brown, S.D.: An introduction to decision tree modeling. J. Chemometr. **18**(6), 275–285 (2004)

Natekin, A., Knoll, A.: Gradient boosting machines, a tutorial. Front. Neurorobot. **7**, 21 (2013)

Okoshi, T., Nakazawa, J., Tokuda, H.: Attelia: sensing user's attention status on smart phones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, pp. 139–142 (2014)

Okoshi, T., Tsubouchi, K., Taji, M., Ichikawa, T., Tokuda, H.: Attention and engagement-awareness in the wild: A large-scale study with adaptive notifications. In: 2017 IEEE International Conference on Pervasive Computing and Communications (percom), pp. 100–110. IEEE (2017)

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)

Pielot, M., Church, K., De Oliveira, R.: An in-situ study of mobile phone notifications. In: Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services, pp. 233–242 (2014)

Pielot, M., Dingler, T., Pedro, J.S., Oliver, N.: When attention is not scarce-detecting boredom from mobile phone usage. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 825–836 (2015)

Pielot, M., Cardoso, B., Katevas, K., Serrà, J., Matic, A., Oliver, N.: Beyond interruptibility: predicting opportune moments to engage mobile phone users. Proc. ACM Interact. Mobile Wearable Ubiquit. Technol. **1**(3), 1–25 (2017)

Shwartz-Ziv, R., Armon, A.: Tabular data: deep learning is not all you need. Inf. Fus. **81**, 84–90 (2022)

Turner, L.D., Allen, S.M., Whitaker, R.M.: The influence of concurrent mobile notifications on individual responses. Int. J. Hum. Comput Stud. **132**, 70–80 (2019)

Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. J. Mach. Learn. Res. **9**(11) (2008)

Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, New York (1995)

Wolniewicz, C.A., Tiamiyu, M.F., Weeks, J.W., Elhai, J.D.: Problematic smartphone use and relations with negative affect, fear of missing out, and fear of negative and positive evaluation. Psychiatry Res. **262**, 618–623 (2018)

Chris Nugent was awarded a first class honours in BEng Electronic Systems and a PhD in Biomedical Engineering both from the University of Ulster. His research interests include the development and evaluation of technologies to support pervasive healthcare within smart environments. Specifically, this has involved research in the topics of mobile based reminding solutions, activity recognition and behaviour modelling and technology adoption modelling.

Rashid Kamal is a dedicated Lecturer in AI and Data Science at the University of Northampton. He holds a PhD in Computing from Ulster University, a Master's in Computer Science from COMSATS University, and a Bachelor's in Computer Science from IQRA University. Kamal has been the recipient of the Best Researcher award in 2018 and 2019 and has been actively involved in various research projects through significant grants. His recent accomplishment includes the Early Career Research Award for his project titled "Advancing Machine Discernment of Linguistic Cues: A Stylometric Approach to AI-Synthesized Text Detection". His research, spanning Machine Learning, HCI, Generative AI, Optimization, Synthetic Data, and digital health, is documented in multiple peer-reviewed articles, contributing to the academic discourse in his specialized fields.

Ian Cleland is a Senior Lecture within the School of Computing at Ulster University, where he leads the research theme of Human Computer Interaction within the Pervasive Computing Research Centre. He received his BS.c in Biomedical Engineering in 2009 and a PhD in Computer Science in 2012, both from Ulster University. His research combines wearable, pervasive, and mobile computing with data science and artificial intelligence to produce innovative digital solutions, mainly in the healthcare domain. Areas of particular interest include sensor-based activity recognition and the use of synthetic data to support data driven modelling. He has published his research in over 100 peer reviewed scientific publications in top forums in digital health and health informatics. Ian has secured externally funded research, as both Principal Investigator and Co-Investigator, to the value of over £7.89M. Most notably he is a Co-I on the Connected Health Innovation Center (CHIC-Phase 2 £3.36M) and the PwC Advanced Engineering and Research Centre (ARC £3.13M). Ian is Co-Chair for the International Conference of Ambient Intelligence (UCAmI2024). He is currently Vice Chair of the Executive Committee for the Alzheimer's Association Technology and Dementia Professional Interest Area.

Aimal Rextin received his Ph.D. from the University of Limerick, Ireland, in 2010. He has been associated with various universities and is currently an Associate Professor at the National University of Science and Technology. He is also collaborating with OneScreen Solutions on various research projects. His research interests include various domains, such as graph-theoretic algorithms, human–computer interaction, applied AI and data analysis.

Paul McCullagh received a BSc (1979) and a PhD (1983) in Electrical Engineering from Queen's University of Belfast. He is a Reader in the School of Computing at Ulster University. He specializes in the teaching of Computer Networking, Internet of Things and Health Informatics. His research interests include Biomedical Signal and Image Processing, Data Mining, Brain Computer Interfaces, and Assisted Living applications. Project include EU FP7 BRAIN project for e-inclusion using the brain-computer interface, EU FP7 Michelangelo for assessing autistic children in a smart environment, EPSRC SMART 2 for self management of chronic disease, TSB NOCTURNAL for assisting people with dementia, and ESRC New Dynamics of Ageing, Design for Ageing Well; and Marie Curie networks: ACROSSING, a multidisciplinary, cross-sector, pan-European training network aimed at Smart Homes Early Stage Researchers, and REMIND, a RISE network that promoted computational techniques to improve compliance to reminders within smart environments, where he led the Ethics WP. He is a member of the Faculty of Computing & Engineering Ethics Filter committee at Ulster University.