Exploring Correlations between Navigational Behaviors and Psychological Responses in Self-Reported Questionnaires Using Classification Trees

Raúl A. Novelo-Cruz[1], Mauricio G. Orozco-del-Castillo[1,*], Esperanza C. Orozco-del-Castillo[2], Nora L. Cuevas-Cuevas[1], Pedro A. G. Ortiz-Sánchez[1] and Jesús Sandoval-Gio[1]

[1] *Tecnológico Nacional de México / IT de Mérida, Mérida, Yucatán, México*
[2] *Cinvestav-IPN, Departamento de Matemática Educativa, Ciudad de México, México*

Abstract

Self-reported questionnaires (SRQs) are a fundamental tool to gather useful data, particularly in psychology. However, they do not usually gather interaction data, such as navigational behaviors. This study aims to explore the role of navigational metrics and classification trees in interpreting user behavior in a set of SRQs and to examine the influence of psychological factors, such as emotional states, moods, and preferences, on user behavior within digital questionnaires. A total of 95 university students participated in this study. Navigational metrics (such as mouse clicks, mouse movements, and time spent responding questionnaires) were used to predict user response patterns in SRQs, whereas a classification tree model was used to segment users into distinct groups based on their navigational behaviors, revealing specific patterns that are associated with different response tendencies. Our results reveal clear relations between users' emotional states and personal preferences, showing how digital behavior analysis could contribute to a deeper understanding and proactive care of mental health.

Keywords

self-reported questionnaires, navigational behaviors, classification trees, online behavior

1. Introduction

Self-reported questionnaires (SRQs) have evolved into an indispensable instrument within clinical practice, public health research, and epidemiology [1]. They facilitate expansive geographic coverage, navigate sensitive topics with discretion, and present a cost-effective approach to gathering useful data, thereby revolutionizing traditional data collection methodologies [2]. The integration of digital technologies has exponentially expanded their utility and scope, making these questionnaires more accessible across diverse populations. This digital transformation can improve the quality and depth of the data obtained, offering unprecedented insights into complex user behaviors and patterns [3]. While SRQs offer an efficient means for data collection across populations, the integrity of data derived from them is often compromised by inattentive or disengaged responses, occasionally leading to poor-quality data that can significantly bias outcomes [4]. As well, conventional analyses of SRQs rarely gather a wider spectrum of user interaction data, such as navigational metrics from digital applications, which could provide deeper insights into respondent behavior and underlying psychological states. The lack of efficient integration between advanced data analysis techniques and navigational metrics in the study of SRQs stands as a gap in the literature. This gap limits our ability to fully understand and interpret the complex dynamics of user behavior and response patterns in digital environments [3].

This study addresses this gap by integrating navigational metrics and classification trees to interpret the behaviors of users in a set of given SRQs. By employing data analysis and machine learning

ICCBR AI Track'24: Special Track on AI for Socio-Ecological Welfare at ICCBR2024, July 1, 2024, Mérida, Mexico
*Corresponding author.
raul.novelo@maikron.org (R. A. Novelo-Cruz); mauricio.orozco@itmerida.edu.mx (M. G. Orozco-del-Castillo); nora.cc@merida.tecnm.mx (N. L. Cuevas-Cuevas); pedro.os@merida.tecnm.mx (P. Ortiz-Sánchez); jesus.sg@merida.tecnm.mx (J. Sandoval-Gio)
0000-0002-8540-9977 (R. A. Novelo-Cruz); 0000-0001-5793-6449 (M. G. Orozco-del-Castillo); 0000-0002-9421-8566 (E. C. Orozco-del-Castillo); 0000-0002-5477-6176 (N. L. Cuevas-Cuevas); 0000-0002-2466-1837 (P. Ortiz-Sánchez); 0000-0001-5847-3669 (J. Sandoval-Gio)

techniques, particularly classification trees, this research transcends traditional methods, offering a more granular perspective on how navigational patterns correlate with user responses and psychological states, shifting the focus from the responses to the behavior which may be correlated to such responses. The application of classification trees provides a systematic approach to dissecting the complex interactions between various user actions, such as mouse clicks, mouse movements, and time spent responding questionnaires, and their implications for questionnaire responses. This contribution is not only methodological but also theoretical, as it may enrich our understanding of digital user behavior within educational and psychological contexts.

The primary objective of this study is to investigate the extent to which navigational metrics are correlated to user response patterns in the items of a set of given SRQs and to examine the role of classification trees in uncovering the underlying structure of these relationships. Building on the premise that user interaction data might offer insights into their cognitive and emotional states, this study proposes to explore whether there exists a correlation between navigational metrics and the responses provided by users in SRQs, and if classification trees can be utilized to segment users into groups that might reflect distinct navigational behaviors, potentially revealing patterns associated with different response tendencies. The study aims to assess the potential of leveraging navigational metrics alongside machine learning techniques to possibly enhance the interpretation of user data in digital questionnaires.

This article is structured as follows. Section 2 delves into the existing literature surrounding SRQs, highlighting previous studies and identifying the theoretical underpinnings that inform our research approach. Section 3 outlines the technical and procedural details of our study, including the development of the web platform, the design and inclusion of the questionnaires, and the specific navigational metrics and classification trees used in the analysis. Section 4 presents our findings, illustrating the possible correlations and patterns identified through the data analysis and machine learning techniques. Section 5 aims to interpret these findings within the broader context of psychological research, discussing their possible implications for theory and practice. Finally, Section 6 summarizes the key contributions of our study, reflects on its limitations, and suggests avenues for future research in this evolving field.

2. Background

The evolution and widespread adoption of SRQs have significantly impacted research methodologies across various fields, including education, psychology, and health sciences [5]. Originating as a practical solution to the logistical and financial challenges posed by face-to-face interviews, SRQs have evolved to become a fundamental tool in data collection efforts, offering a unique combination of efficiency, scalability, and respondent anonymity. This evolution has been further catalyzed by technological advancements, leading to the development of sophisticated online platforms that facilitate a broader reach and more dynamic interaction with diverse populations [6]. Despite their convenience and broad applicability, SRQs are not without limitations, necessitating ongoing research to address issues related to response rates, data quality, and the influence of questionnaire design on respondent behavior [7, 3].

Traditional analysis of SRQs often encounters several challenges and limitations that can significantly affect the quality and reliability of the derived data. One of the primary concerns is the presence of biased responses, where respondents may provide socially desirable answers rather than truthful ones, leading to skewed results and misinterpretations [8]. Additionally, the lack of interaction with an interviewer in self-reported formats can result in higher instances of misunderstanding questions, which can further compromise data accuracy [9]. Moreover, the conventional analysis methods may not adequately address the complex nature of the data, particularly when dealing with high-dimensional data or identifying subtle patterns within the responses [10].

The concept of navigational metrics in digital environments provides a foundational framework for understanding user interactions and movements within web-based or app-based platforms. These metrics offer quantifiable insights into how users navigate through digital spaces, which paths they choose, and how they engage with content [11]. The integration of navigational metrics allows for a more detailed analysis of user behavior. For instance, the study of large unstructured environments through hybrid metric map structures reveals the potential of combining navigational metrics with other data representations to enhance the understanding of user pathways and interactions [12].

Machine learning, a subset of artificial intelligence, has revolutionized the analysis of complex data sets, offering robust methodologies to uncover underlying patterns and relationships. Within machine learning, classification trees stand out for their intuitive structure and interpretability [13], making them particularly suitable for analyzing data from SRQs [14, 15]. These decision-making tools dissect the dataset by recursively splitting it into smaller subsets based on the most relevant attributes, thereby constructing a tree-like model of decisions [15, 16]. This approach not only aids in categorizing responses effectively but also highlights the significance of various questionnaire items in predicting outcomes. Studies leveraging machine learning models, including classification trees, have demonstrated their efficacy in transforming complex datasets into actionable insights, thereby enhancing the understanding of respondent behaviors and characteristics [14, 17].

While SRQs have been extensively used across various disciplines for data collection, traditional analytical methods often fall short in capturing the complex interplay of variables and respondent behaviors. Although machine learning techniques have shown promise in identifying complex patterns within datasets, there remains a gap in the literature regarding the integration of these techniques with traditional analytical methods for analyzing SRQ data. Such integration is crucial, as it combines the interpretability and simplicity of traditional methods with the predictive power and flexibility of machine learning, leading to more comprehensive insights. Studies have highlighted the importance of employing diverse methodologies to improve data quality and response rates in SRQs, suggesting a need for more customized approaches that can handle the intricacies of survey data [3, 4]. Furthermore, the integration of different data collection modes, including electronic and paper-based methods, has been shown to impact data quality, indicating the potential benefits of a multimodal approach in survey design and analysis [6, 8]. The integration of machine learning with traditional methods in the analysis of SRQs holds significant implications for both research and practice. By relying on the predictive capabilities of machine learning, researchers can uncover subtle patterns and relationships within the data that might be overlooked by traditional analyses, leading to richer insights and more accurate predictions. This approach not only enhances the depth and quality of academic research but also has practical applications in areas where understanding respondent behaviors and preferences is crucial.

3. Methodology

The research design of this study was structured around the deployment of a web platform [18, 19], specifically developed for this research to facilitate the detailed capture of user interactions within SRQs. Central to this approach was the adoption of JavaScript as the cornerstone for manipulating browser events, enabling precise monitoring of mouse movements, clicks, and time spent responding to the questionnaires. This enabled a granular analysis of how participants interacted with both the platform and the questionnaires, providing insights into user browsing behaviors. To ensure the fidelity and consistency of the collected data, the platform was engineered to be accessible solely through certain web browsers, excluding mobile devices. This limitation was implemented to guarantee the uniformity of the data collection environment, ensuring that all participants interacted with the platform under comparable conditions and that the recorded metrics were reflective of a standardized user experience.

The participant pool for this study was comprised of 95 university students, with a demographic breakdown of 13 women and 82 men, representing a wide range of academic semesters, all of whom participated voluntarily. The recruitment process was designed to invite a representative sample of the student population, aiming to capture a variety of browsing behaviors and interactions with the SRQs. Participants were thoroughly briefed about the objectives and methodology of the study, in line with the ethical guidelines outlined in the Declaration of Helsinki for research involving human subjects. Informed consent was obtained from all participants prior to their involvement, ensuring that they were fully aware of the study's nature and their role within it.

The battery of questionnaires administered in the study was selected to cover a wide range of psychological constructs. The Generalized Anxiety Disorder Assessment (GAD-7) [20] and the State-Trait Anxiety Inventory (STAI) [21], provided a multifaceted view of anxiety-related symptoms. Concurrently, depressive symptomatology was evaluated through the Beck Depression Inventory (BDI) [22], the Center for Epidemiological Studies-Depression Scale (CES-D) [23], the Carroll Rating Scale for Depression (CRSD) [24], and the Patient Health Questionnaire (PHQ-9) [25]. Additionally, the study sought to explore other relevant psychological and behavioral dimensions. The Reduced Morningness-Eveningness Questionnaire (RMEQ) [26] was employed to assess circadian preferences, an aspect with potential implications for study habits and academic performance. The Rosenberg Self-Esteem Scale (RSES) [27], a critical measure of self-worth, was included to understand its impact on educational achievement and psychological health. Complementing these standardized measures, an internal questionnaire designed by the Tecnológico Nacional de México was integrated to specifically investigate study habits, techniques, motivation, well-being, and assertiveness within the student body [28]. In total, the assessment comprised 176 items.

The data acquisition phase was designed to capture three user interaction metrics when responding the SRQs: mouse clicks (C_M), mouse movements (M_M), and time spent responding to the questionnaires (t). Upon collection, the data were transmitted to a centralized server through an HTTP request, employing Python and Django for backend operations and PostgreSQL for database management [18]. The information was systematically serialized in JSON format, encompassing both the interaction metrics and the questionnaire responses.

A data cleaning process was necessary to prepare the dataset for analysis; this involved a thorough review of the data presented in a tabular format. Each row corresponded to a participant's entry, with columns detailing responses to the questionnaire items. The cleaning process aimed to identify and exclude entries lacking complete information on the key interaction variables, thereby ensuring the dataset's integrity and coherence. From the initial pool of 95 participants, the data cleaning process led to the exclusion of 19 records due to inconsistencies, primarily attributed to the participants' web browser compatibility or specific privacy configurations on their computers. Specifically, 16 samples were removed due to inaccuracies in the recording of time spent responding to the questionnaires, and 3 due to missing mouse movement data. This refinement left a robust dataset of 76 complete and valid samples for subsequent analysis.

Following the data cleaning process, an exploratory data analysis (EDA) was conducted to uncover underlying patterns and relationships within the dataset. The EDA employed a scatterplot matrix visualization for its efficacy in handling multivariate data. The scatterplot matrix served as a comprehensive tool for visualizing pairwise relationships and distributions among the variables and facilitated an intuitive grasp of both linear and nonlinear correlations between variables, enriching our understanding of the data's underlying structure. A particularly noteworthy finding was the discernible positive correlation between mouse movements (M_M) and time spent responding to the questionnaires (t), quantified by a Pearson correlation coefficient of 0.9597. To uncover distinctive patterns, each variable (C_M, M_M, and t) was segmented into tertiles, isolating the lower (T_1) and upper (T_3) extremes of the distribution for detailed analysis. This approach allowed for a concentrated examination of behaviors at the distribution tails.

Along with classification trees, to predict the key browsing behavior variables we also evaluated other machine learning models, particularly Multi-core Support Vector Machines (SVMs), Logistic Regression, k-Nearest Neighbors (k-NN), and Discriminant Analysis. However, these models yielded modest accuracy rates between 50% and 60% through 5-fold cross-validation. In contrast, classification trees emerged as notably more effective, achieving up to 70% accuracy. Despite this value being relatively modest, the potential for further improvement through optimization was encouraging. Classification trees were particularly appealing due to their transparency, interpretability, and suitability for smaller datasets [13], like our focused samples of T_1 and T_3. These qualities minimized the risk of overfitting, a common concern with complex models in small data contexts. Consequently, we opted to concentrate on refining classification trees to enhance performance while preserving the model's interpretability.

The fine-tuning and optimization of the classification trees in this study followed a strict, systematic

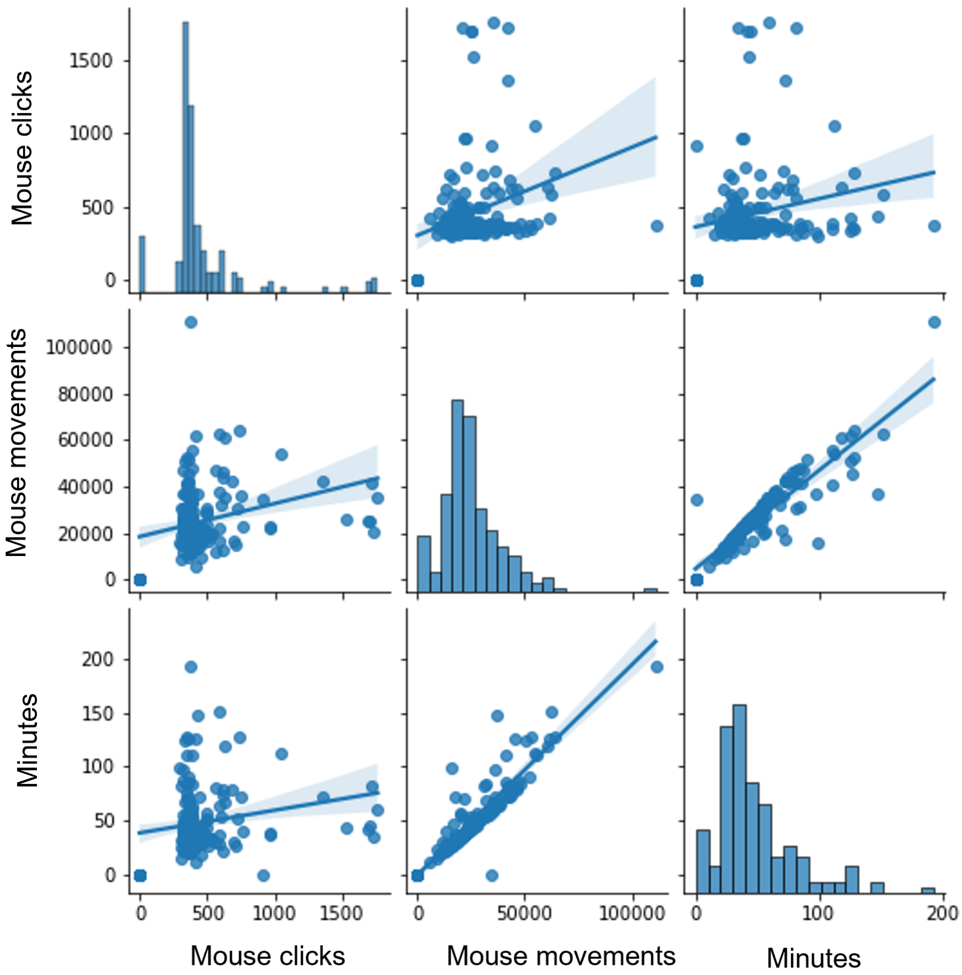

Figure 1: Scatterplot matrix between the number of clicks, mouse movements and time spent responding to the questionnaires

approach to enhance model robustness, accuracy, and interpretability. The process began with k-fold cross-validation, where the dataset was divided into k subsets. For each subset, the model was trained on k-1 folds and validated on the remaining fold, with k ranging from 1 to 20, to ensure thorough evaluation and minimize overfitting. An exhaustive parameter search was then conducted to identify the optimal classification tree configurations, focusing on the minimum leaf size and maximum number of splits. This exhaustive search evaluated all parameter combinations through 200 iterations per combination to ensure the robustness of the findings, using k-fold cross-validation with k values from 1 to 5. To finalize the model selection, a holdout validation technique was employed over 300 cycles, randomly splitting the data into training and test sets each time. Classification trees were trained using the optimal parameters found, and accuracy was assessed on both the training and test sets. The minimum of these accuracies was taken to prevent overfitting, with the best-performing model across iterations being selected based on its ability to generalize well to new data.

Several actions were carried out to evaluate the performance of the model and understand its predictive ability in greater depth. These actions allowed obtaining both quantitative and qualitative results. One of the quantitative analyzes consists of calculating the importance of the predictors by applying the technique of Permutation of Importance of Predictors [**?**]. This technique is based on a permutation process in which random perturbations are introduced into the characteristics of the data set. Then, we observe how these perturbations affect the performance of the model. The main idea is to evaluate which characteristics of the data set, when randomly perturbed, have a significant impact on the predictive ability of the model. Those features whose perturbation causes a noticeable decrease in model performance are considered more important, since their presence or absence significantly influences model decisions. This predictor importance analysis provides valuable insights into the specific characteristics of the data set that have a prominent influence on the decisions made by the model. This not only helps to better understand how the model makes its predictions, but can also guide relevant feature selection and model optimization in later stages. In contrast, qualitatively, visualizing the classification tree allows for a more intuitive understanding of how the model makes decisions based on input features.

Figure 2 provides a visual overview of the methodology applied in this study, detailing each step from data cleaning to predictive accuracy enhancement. This diagram illustrates the systematic approach used to ensure the robustness and validity of our analysis.

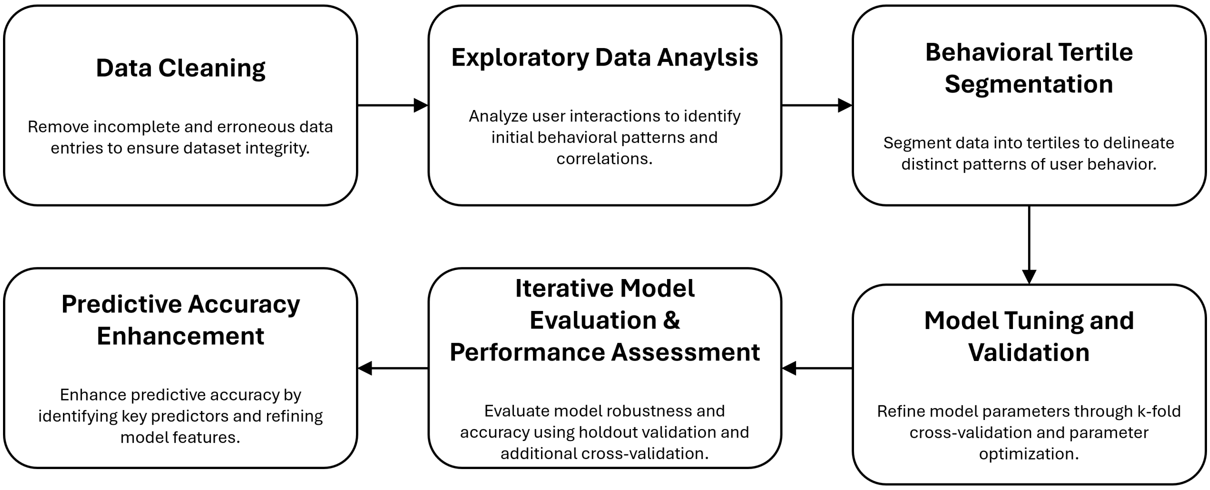

Figure 2: A diagram of the process used in the study, starting with Data Cleaning, where inaccuracies are removed. It follows with Exploratory Data Analysis (EDA), where initial patterns and correlations within user interactions are discovered. Behavioral Tertile Segmentation separates the data into tertiles to highlight distinct behavior patterns. Model Tuning and Validation involve refining the machine learning models through k-fold cross-validation and parameter optimization to ensure optimal settings. Iterative Model Evaluation and Performance Assessment assesses the models' robustness and accuracy to validate their predictive capabilities. Lastly, Predictive Accuracy Enhancement focuses on enhancing the models' predictive accuracy by identifying and prioritizing key predictors and refining model features.

4. Results

We systematically explored the predictive performance of the classification tree models across three critical variables: mouse clicks, mouse movements, and time spent responding to the questionnaires. Our approach for each variable commenced with training the model using predetermined parameters, followed by validation employing k-fold cross-validation to assess baseline performance. In the analysis focused on mouse clicks (C_M), the initial exploration using default model parameters revealed a baseline accuracy of 0.53, accompanied by a considerable standard deviation of 0.24, indicating variability in model performance across different data segments. Through a meticulous process of parameter optimization, an improved configuration was identified with a minimum leaf size of 1 and a maximum of 5 splits, slightly enhancing the model's accuracy to 0.5465. The search for the optimal classification tree configuration indicated a notable improvement, with the model reaching an accuracy of up to 0.96 under the specific conditions described. The qualitative analysis of the classification tree, as illustrated in Figure 3, provides a clear understanding of the factors influencing clicking behavior. The numbers at the end of each branch denote tertiles of user interaction, where '1' represents the lowest tertile (T_1) and '3' the highest tertile (T_3), indicating varying levels of engagement and activity. Central to this model is item 57, which pertains to contemplations of self-harm. Participants who seldom entertain thoughts of being better off dead or injured are predominantly categorized within the highest

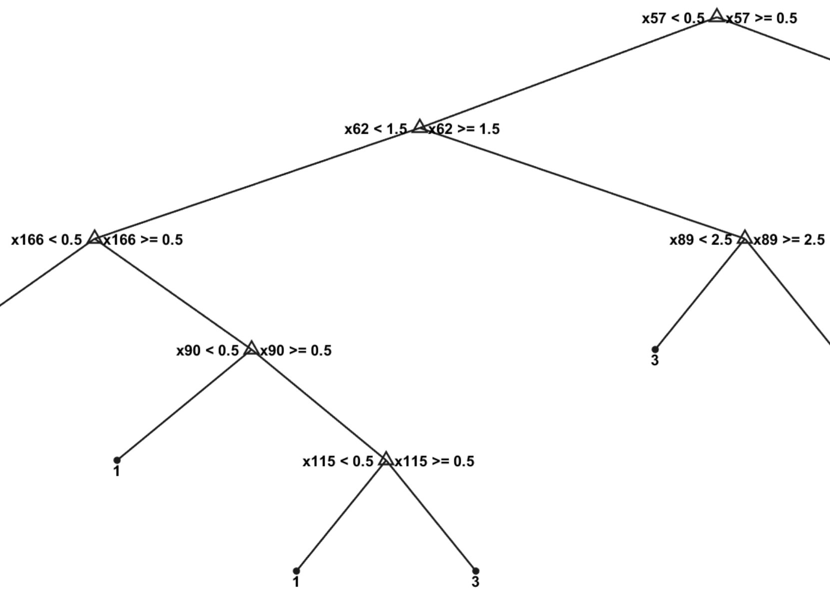

Figure 3: Classification tree model illustrating the influence of psychological factors on mouse clicking behavior in digital questionnaires. Central nodes represent key questionnaire items, with item 57 at the root, reflecting the impact of self-harm thoughts on clicking activity. Branches diverge at items 166 (beliefs in luck), 62 (circadian preferences), 90 (crisis avoidance strategies), 115 (organization of personal items), and 89 (feelings of safety), each contributing to distinct clicking patterns. The numbers at the end of each branch represent tertiles of user interactions, with 1 indicating the lowest tertile (T_1) and 3 the highest tertile (T_3) of activity. The model achieved a significant accuracy of 0.96.

tertile for clicking activity. Diverging at item 166, concerning beliefs in luck, the model suggests that individuals less convinced of their own good fortune (scored below 0.5) demonstrate reduced clicking tendencies, placing them in the lowest tertile. Conversely, the model indicates that self-identified nocturnal individuals (item 62) are inclined towards more extensive clicking behaviors. Further down the left branch, item 90, which addresses strategies for avoiding crises, plays a pivotal role. Meanwhile, the decision pathway through item 115, concerning the organization of personal items in one's study space, bifurcates clicking behaviors significantly. On the far right, item 89, associated with feelings of safety, emerges as a determinant; individuals expressing a lower sense of security tend to engage in more frequent clicking, possibly as an indicator of heightened anxiety or uncertainty during the questionnaire completion process.

In the examination of mouse movements (M_M), the classification tree model initially presented an average accuracy of 0.68, with a standard deviation of 0.24. Parameter optimization revealed that a configuration with a minimum leaf size of 15 and a maximum of 5 splits was optimal, leading to a notable improvement in accuracy to 0.7107. The refined model, with its adjusted parameters, demonstrated a more robust predictive capability. An exhaustive search further enhanced the model's precision, culminating in training and validation accuracies of 0.75 and 0.80, respectively, and an overall accuracy of 0.76. Class-specific metrics revealed an accuracy of 0.8095 for the lowest tertile (T_1) and 0.7241 for the highest tertile (T_3), complemented by recall rates of 0.6800 and 0.8400, respectively. The F1-Scores for T_1 and T_3 stood at 0.7391 and 0.7778, respectively, highlighting the model's balanced efficacy in distinguishing between different levels of mouse movement engagement.

For the variable of time spent responding to the questionnaires (t), the initial assessment of the model demonstrated an accuracy of 0.63, accompanied by a significant standard deviation of 0.30. The optimization phase, mirroring the approach for mouse movements, determined that the optimal settings involved a minimum leaf size of 15 and a maximum of 10 splits, leading to a slight increase in accuracy to 0.6853. Following an exhaustive search to fine-tune the model, the training and validation accuracies reached 0.75, with the overall accuracy settling at 0.76. The model showcased a differentiation between

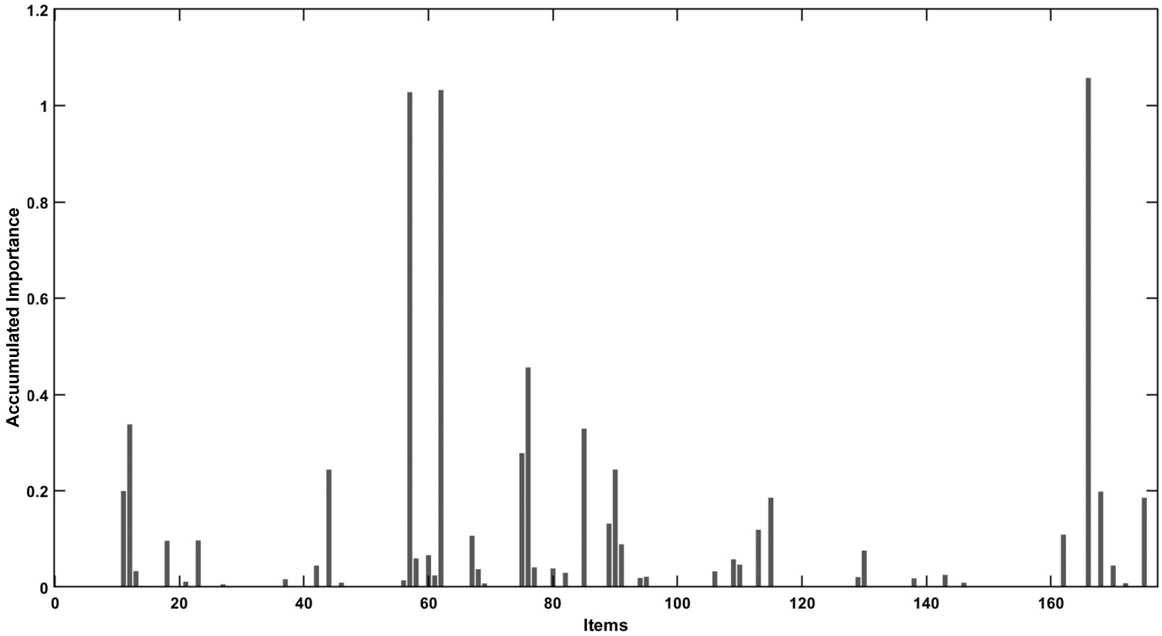

Figure 4: Bar graph showing the accumulated importance of each of the 176 items during the 100 iterations of the classification tree construction process. Three items emerge as the most relevant for the prediction of the mouse clicks variable: 57, 62 and 166

the lowest and highest tertiles of time, evidenced by accuracies of 0.8095 for T_1 and 0.7241 for T_3, respectively. Additionally, recall scores of 0.6800 for T_1 and 0.8400 for T_3, along with F1-Scores of 0.7391 for T_1 and 0.7778 for T_3 were observed.

The integrated analysis of the classification tree model across the three key variables—–mouse clicks, mouse movements, and time spent responding to the questionnaires—–revealed an understanding of user interaction behaviors within the digital questionnaire environment. Initial evaluations showed baseline accuracies that pointed to the inherent variability and complexity of each variable, setting the stage for targeted optimizations. Through meticulous parameter tuning, each variable saw improvements in accuracy, highlighting the importance of tailored approaches in machine learning models. Notably, mouse clicks experienced a remarkable accuracy increase after optimization, highlighting the potential of fine-tuning. The quantitative analysis also evidenced a significant correlation between mouse movements and time spent responding to the questionnaires, as evidenced in the scatterplot matrix (Figure 1), confirming a closely linked behavioral pattern where increased navigation is associated with longer time periods. This correlation is also reflected when analyzing the relevant items through the Permutation of Importance of Predictors, since the same, very straightforward results are obtained for both variables. In both cases, item 173 is the only one to stand out, "Sometimes I don't listen to what they tell me because I am thinking about other things". Those who answer "Almost never" show a distinctive pattern: they make more mouse movements and take longer to complete the questionnaire (T_3), compared to those who answer otherwise (T_1).

It is important to note that while these results are promising, they are context-specific and may not generalize to other settings without similar validation.

5. Discussion

In this section, we delve into the relationship between user interaction metrics in SRQs and the psychological underpinnings these behaviors may signify. Our analysis of mouse click behavior (C_M) suggests possible connections between certain questionnaire items and the psychological states they may indicate, though these findings are preliminary and require further validation. Notably, item 57, concerning thoughts of self-harm, emerged as a significant predictor of click behavior, suggesting that individuals less burdened by such thoughts engage more actively with the questionnaire. This might indicate a more positive or stable emotional state, where engagement with the task at hand is higher. Similarly, circadian preferences, as captured by item 62, highlight how personal productivity cycles influence digital engagement, with nocturnal individuals showing increased click activity. This points to the broader implication that individual differences, such as circadian rhythms, significantly affect how users interact with digital platforms. Furthermore, item 166's exploration into beliefs about good fortune reveals the psychological impact of optimism on digital behavior; individuals with a positive outlook may navigate the questionnaire more efficiently, reflecting a targeted and decisive interaction style. Conversely, a less optimistic view might lead to increased exploratory behavior, as manifested in higher click rates.

The exploration of navigation metrics in SRQs suggests a potential avenue for enhancing mental health assessments, though these findings require further empirical support to establish their broader applicability. The significant correlations between user behaviors, such as mouse clicks and movements, and psychological states offer a unique perspective on mental well-being. For instance, the association between frequent clicking and items related to self-harm thoughts or optimism suggests that beyond explicit questionnaire responses, user interactions with digital platforms can provide valuable, indirect insights into their mental state. This approach holds the potential to identify individuals who may be at risk or require further mental health support, even when conventional questionnaire scores do not indicate such needs. By leveraging these metrics, mental health professionals can gain a more comprehensive understanding of an individual's psychological condition, potentially leading to more timely and tailored interventions. This innovative use of navigation metrics evidences their value as critical components in the evolving area of mental health care, where digital behavior becomes an integral part of understanding and supporting mental wellness.

The integration of navigation metrics into mental health practice represents a significant advancement in patient assessment and care. By analyzing patterns in mouse clicks, movements, and engagement times, practitioners can uncover subtle indicators of psychological distress or well-being that might not be evident through traditional assessment methods. This approach enables a more dynamic and real-time evaluation of patients' mental states, facilitating early detection of potential issues and allowing for prompt, personalized interventions. For example, increased mouse activity coupled with specific item responses could trigger further evaluation or direct conversations about mental health, even if the patient hasn't explicitly expressed concerns. Incorporating these metrics into clinical settings could considerably improve patient monitoring, making it possible to continuously assess mental health in a less invasive manner. As we move forward, developing protocols for the ethical use and interpretation of these digital behaviors will be crucial in ensuring that this innovative approach effectively complements existing mental health practices, enhancing the overall care and support provided to individuals.

The adoption of navigation metrics in mental health assessments brings forth important ethical and privacy considerations that must be meticulously addressed. The collection and analysis of digital behavior data, such as mouse clicks and movements, raise questions about user consent and data confidentiality. It is imperative to establish transparent protocols that inform users about the data being collected, its intended use, and the measures in place to protect their privacy. Ensuring anonymity and secure data handling practices is crucial to maintaining trust and upholding ethical standards. Furthermore, the interpretation of these metrics must be conducted with sensitivity and professional discretion, recognizing the potential implications for individuals' privacy and mental health.

The potential of navigation metrics in mental health opens several avenues for future research. Further studies could explore the relationship between specific patterns of digital interaction and a wider range of psychological conditions, extending beyond the initial findings related to depression and anxiety. Investigating the applicability of these metrics in diverse populations and settings will help determine their universality and potential cultural influences on digital behavior. Advanced analytical techniques, including machine learning and artificial intelligence, could refine the accuracy and predictive power of navigation metrics, offering more personalized and precise mental health assessments. Particularly, more powerful machine learning models than decision trees, such as artificial neural networks, could better identify complex patterns between responses and navigational behavior; however, actions must be taken to tackle the lack of interpretability of such models. Additionally, longitudinal studies examining changes in digital behavior over time could provide insights into the progression of mental health conditions and the impact of therapeutic interventions.

While the study provides insightful observations, it is important to recognize its limitations, which are crucial for interpreting the findings accurately and understanding their generalizability. One of the primary limitations is the sample size, which, while sufficient for initial explorations, may not fully capture the wide variability in human behavior across different demographics and cultural backgrounds. Additionally, the study is limited by its specific context—the use of digital questionnaires within an educational environment—which may not directly translate to other settings or types of digital interaction. Furthermore, the reliance on self-reported data through digital platforms introduces a layer of complexity, as such data can be influenced by various factors including user comfort with technology, privacy concerns, and the potential for response bias. These factors could affect the validity of the correlations observed between navigational metrics and psychological states. The classification trees, while useful for identifying patterns, are also inherently limited by their simplicity and the risk of overfitting, especially with smaller datasets. Thus, the conclusions drawn must be viewed as tentative, highlighting the need for further research to replicate and extend these findings.

6. Conclusion

Our exploration of user interactions within digital questionnaires provides initial insights into how psychological factors might influence engagement metrics such as mouse clicks, mouse movements, and time spent on questionnaires. Our preliminary findings suggest potential associations between behaviors, such as the frequency of clicks and the range of movements, and emotional states and personal rhythms, with certain responses—pertaining to self-harm thoughts, luck perceptions, and evening preferences—standing out as significant predictors. This relationship between digital behavior and psychological dimensions emphasizes the potential of digital platforms not just as tools for data collection but as mirrors reflecting the user's mental health. Such findings advocate for a deeper integration of psychological understanding into digital interactions, suggesting that these platforms could play a crucial role in mental health assessment through everyday digital engagements.

Understanding how emotional states and cognitive tendencies influence behaviors like mouse clicks, movements, and time spent on tasks allows an interpretation of digital interactions as reflections of mental well-being. By aligning digital interactions with psychological profiles, there's an opportunity to enhance mental health monitoring and support, creating a space where digital behavior analysis contributes to a deeper understanding and proactive care of mental health.

Building upon our study's insights, future work could explore a wider range of psychological variables, deepening our comprehension of their influence on how individuals engage with digital questionnaires. The adoption of advanced data analysis techniques, including machine learning, offers the potential to reveal more intricate patterns of behavior, providing a more detailed picture of an individual's mental state as they navigate digital platforms. Investigating how these interaction patterns vary across different digital contexts could uncover new insights into user behavior, offering a more holistic view of mental health in digital environments. A particularly exciting direction is the development of digital interfaces that adapt in real time to the user's psychological cues, offering a more personalized and supportive online experience. This line of research not only promises to broaden our understanding of the nexus between psychology and digital behavior but also to pioneer new approaches for mental health assessment and support.

By advocating for the fusion of psychological understanding with human-computer interaction research, we aim to spark further investigations that elevate the importance of emotional and cognitive states in digital engagements. This approach opens up innovative avenues for mental health assessment, suggesting that valuable information about an individual's psychological well-being can be derived from their interaction patterns with a questionnaire, not solely from the questionnaire's content. Such insights could revolutionize the way we interpret digital behaviors, offering a richer perspective on individual mental health that complements traditional assessment methods, thereby enriching our understanding and support of psychological well-being in the digital age.

References

[1] M. G. Orozco-del Castillo, E. C. Orozco-del Castillo, E. Brito-Borges, C. Bermejo-Sabbagh, N. Cuevas-Cuevas, An Artificial Neural Network for Depression Screening and Questionnaire Refinement in Undergraduate Students, in: M. F. Mata-Rivera, R. Zagal-Flores (Eds.), Telematics and Computing. WITCOM 2021. Communications in Computer and Information Science, volume 2, Springer Nature Switzerland AG 2021, 2021, pp. 1–13. URL: https://link.springer.com/10.1007/978-3-030-89586-0_1. doi:10.1007/978-3-030-89586-0_1.

[2] S. Parikka, J. Ikonen, H. Tolonen, A. Lundqvist, Comparison of response patterns in national surveys among finnish adult population in 2020 and 2022, The European Journal of Public Health 33 (2023). doi:10.1093/eurpub/ckad160.898.

[3] J. S. M. Belisario, et al., Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods, The Cochrane database of systematic reviews 7 (2015) MR000042. doi:10.1002/14651858.MR000042.pub2.

[4] M. Hyman, A. Kostyk, W. Zhou, L. Paas, Novel approaches for improving data quality from self-administered questionnaires, International Journal of Market Research 61 (2019) 552 – 555. doi:10.1177/1470785319870622a.

[5] K. Hudmon, N. Elkhadragy, K. Hilts, A practical guide for clinical pharmacists conducting health survey research, Journal of the American College of Clinical Pharmacy 4 (2020) 525 – 534. doi:10.1002/jac5.1368.

[6] A. Artino, J. S. La Rochelle, K. J. Dezee, H. Gehlbach, Developing questionnaires for educational research: Amee guide no. 87, Medical Teacher 36 (2014) 463 – 474. doi:10.3109/0142159X.2014.889814.

[7] V. Díaz de Rada, Influence of the questionnaire design in self-administered surveys, Sociology International Journal (2019). doi:10.15406/SIJ.2019.03.00163.

[8] A. Bowling, Mode of questionnaire administration can have serious effects on data quality, Journal of public health 27 3 (2005) 281–91. doi:10.1093/PUBMED/FDI031.

[9] P. Edwards, Questionnaires in clinical trials: guidelines for optimal design and administration, Trials 11 (2010) 2. doi:10.1186/1745-6215-11-2.

[10] E. Gibson, J. Zhang, J. Yan, L. G. Chillrud, J. Benavides, Y. Nunez, J. Herbstman, J. Goldsmith, J. N. Wright, M. Kioumourtzoglou, Principal component pursuit for pattern identification in environmental mixtures, Environmental Health Perspectives 130 (2021). doi:10.1289/EHP10479.

[11] Z.-Y. Lim, L.-Y. Ong, M. Leow, Cluster-n-engage: A new framework for measuring user engagement of website with user navigational behavior, IEEE Access 11 (2023) 112276–112292. doi:10.1109/ACCESS.2023.3322958.

[12] J. Guivant, E. Nebot, J. Nieto, F. Masson, Navigation and mapping in large unstructured environments, The International Journal of Robotics Research 23 (2004) 449–472. doi:10.1177/0278364904042203.

[13] H. M. Ruiz-Juárez, J. Castillo-Araujo, M. G. Orozco-del Castillo, N. L. Cuevas-Cuevas, F. Cárdenas-Pimentel, R. Cáceres-Escalante, A Decision Tree as an Explainable Artificial Intelligence Technique for Identifying Agricultural Production Predictor Variables in Mexico, 2023, pp. 1–14. URL: https://link.springer.com/10.1007/978-3-031-45316-8_1. doi:10.1007/978-3-031-45316-8_1.

[14] C.-Y. Ko, F.-Y. Leu, Analyzing attributes of successful learners by using machine learning in an undergraduate computer course, in: 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA), IEEE, 2018, pp. 801–806. doi:10.1109/AINA.2018.00119.

[15] W.-Y. Loh, Classification and regression trees, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1 (2011) 14–23. doi:10.1002/widm.8.

[16] N. Speybroeck, Classification and regression trees, International Journal of Public Health 57 (2012) 243–246. doi:10.1007/s00038-011-0315-z.

[17] A. Maxwell, T. Warner, F. Fang, Implementation of machine-learning classification in remote sensing: an applied review, International Journal of Remote Sensing 39 (2018) 2784–2817. doi:10.1080/01431161.2018.1433343.

[18] C. I. Moo-Barrera, M. G. Orozco-del Castillo, M. R. Moreno-Sabido, N. L. Cuevas-Cuevas, C. Bermejo-Sabbagh, Web Platform for the Analysis of Physical and Mental Health Data of Students, in: M. F. Mata-Rivera, R. Zagal-Flores (Eds.), Telematics and Computing. WITCOM 2022. Communications in Computer and Information Science, Springer Nature Switzerland AG 2022, 2022, pp. 139–156. URL: https://link.springer.com/10.1007/978-3-031-18082-8_9. doi:10.1007/978-3-031-18082-8_9.

[19] I. Madera-Torres, M. G. Orozco-del Castillo, S. N. Moreno-Cimé, C. Bermejo-Sabbagh, N. L. Cuevas-Cuevas, Detection of Mental Health Symptoms in the Written Language of Undergraduate Students Using a Microblogging Platform, volume 1, 2023, pp. 473–486. URL: https://link.springer.com/10.1007/978-3-031-45316-8_30. doi:10.1007/978-3-031-45316-8_30.

[20] R. Spitzer, K. Kroenke, J. Williams, B. Löwe, A brief measure for assessing generalized anxiety disorder: The gad-7, Archives of Internal Medicine 166 (2006) 1092–1097.

[21] C. Spielberger, Manual for the State-Trait Anxiety Inventory STAI (form Y) ("Self-Evaluation Questionnaire"), Consulting Psychologists Press, 1983.

[22] A. Beck, R. Steer, G. Brown, The beck depression inventory-ii, San Antonio 78 (1996) 490–498.

[23] L. Radloff, The ces-d scale: A self-report depression scale for research in the general population, Applied Psychological Measurement 1 (1977) 385–401.

[24] B. Carroll, J. Fielding, T. Blashki, The carroll rating scale for depression, The British Journal of Psychiatry 138 (1981) 194–200.

[25] K. Kroenke, R. Spitzer, J. Williams, The phq-9: Validity of a brief depression severity measure, Journal of General Internal Medicine 16 (2001) 606–613.

[26] A. Adan, V. Natale, Chronotype and personality factors in the daily consumption of alcohol and psychostimulants, Addiction 107 (2012) 439–447.

[27] M. Rosenberg, Society and the adolescent self-image (1965).

[28] Tecnológico Nacional de México, Encuesta de hábitos de estudio, https://www.tuxtla.tecnm.mx/wp-content/uploads/encuesta_habitos_de_estudio.docx, 2024. (Accessed on 03/24/2024).