Bulletin of Electrical Engineering and Informatics
Vol. 13, No. 3, June 2024, pp. 2095~2108
ISSN: 2302-9285, DOI: 10.11591/eei.v13i3.6936

Holistic personas to increase the novice developer productivity

Wahyu Andhyka Kusuma[1,2], Azrul Hazri Jantan[2], Novia Indriaty Admodisastro[2], Noris Norowi[2]
[1]Informatics, Faculty of Engineering, Universitas Muhammadiyah Malang, Malang, Indonesia
[2]Human Computer Interaction, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, Malaysia

Article Info

Article history:

Received Jun 9, 2023
Revised Sep 14, 2023
Accepted Oct 12, 2023

Keywords:

Developer productivity
Holistic persona
Markov chain
User-centric approach
User experience journey

ABSTRACT

A deeper understanding and integration with system users' thoughts and emotional experiences are required for user-engaged development. User experience (UX) journey integrates user requirements and problem-solving approaches. The integration of data-driven techniques and user-centric approaches in software development is investigated in this study. It focuses on using the Markov chain model to predict developer productivity based on data gathered while creating personas across three projects. Organizations can gain valuable insights into user needs and requirements by conducting purposeful activities such as strength, weaknesses, opportunities, and threats (SWOT) analysis, competitor analysis, hypothesis formulation, identification of behavioral variables, mapping interviews, and defining characteristics and objectives. The model has predictive capabilities that allow for more informed decision-making, more efficient resource allocation, and better project planning. The goal of the activity and the model ensure the development of software products that effectively meet the needs of users, resulting in a higher success rate for software development initiatives. This study emphasizes the importance of integrating quantitative and qualitative analysis to drive successful software development projects and increase productivity while meeting user needs. According to the findings of the research conducted from the three projects completed, the proposed methods have similarities, and predictions using the Markov chain can determine the success of novice developers.

Corresponding Author:

Wahyu Andhyka Kusuma
Human Computer Interaction, Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
Serdang, Malaysia
Email: gs63875@student.upm.edu.my

1. INTRODUCTION

Software products play a crucial role in enhancing productivity within software development. This association stems from the imperative for software to effectively address the needs of end users, resulting in profound implications not only for the economy [1] but also across various dimensions encompassing technical and social considerations [2], [3], emotional influences (such as affecting moods and feelings) [4], work-life dynamics [5], [6], culture [7], and collaborative teamwork [8], [9]. Notably, software quality has exhibited a progressive upswing over time, primarily attributable to the implementation of enhanced requirements management practices [10] and the active engagement of users throughout the development process. However, user-involved development necessitates deeper understanding and integration with system users' thoughts and emotional experiences [11].

In software development, developers employ personas to delineate target user groups and gain insights into the characteristics and requirements of system users. Personas are considered an effective

approach for depicting user attributes and desired objectives [12]. Prior research has demonstrated that persona data collection typically involves qualitative and quantitative techniques to facilitate an in-depth exploration of user understanding and creating representative personas [13], [14]. Nevertheless, qualitative methods have limitations stemming from their reliance on the cognitive abilities of the development team [15]. In contrast, quantitative methods may need more contextual details and depth of qualitative approaches [16].

Previous studies propose an alternative approach to characterizing users through using personas in high-level conceptual design, aligning with user-centered development (UCD) methodologies [17]. Within the context of design thinking, personas can be employed to engage users in contemporary development approaches [18]. The persona method encompasses various activities, such as collecting, analyzing, and synthesizing information about the users interacting with the software under development [17].

However, developers need help concerning the proficiency gap between novice and professional developers. Additionally, the user persona method becomes problematic when the integration process becomes overly intricate and subjectively determined by the developer [19]. Furthermore, the identification process must account for the developer's characteristics and capabilities to design and implement usable systems [20], [21]. Consequently, personas become an integral aspect of exploring software requirements.

The persona concept has garnered attention from researchers aiming to offer practical solutions for describing user interactions that can be represented through applications, reflecting user goals [17], [20], [22]. However, personas have been criticized for being abstract, subjective, and misleading [23]. Several personas have been developed to address this issue, including the holistic persona with five dimensions: factual, personality, intelligence, knowledge, and cognitive process [24]. To assess the effectiveness of holistic personas in determining user needs and supporting conceptual design, a series of empirical studies have been designed and conducted. These studies aim to investigate whether including personality traits in the interactions between designers and personas leads to variations in conceptual design. The persona design in this research derives from the user experience (UX) journey framework, which combines user experience and requirements to segment users based on their characteristics. Based on the problems and objectives in the previous paragraph, this research is expected to provide a framework that novice developers or designers can use to fill the gaps in the balance between developers and personas. In addition, developers can identify their own abilities to increase productivity.

The paper follows a well-organized structure that includes a literature review, research questions, methodology, results, discussion, and conclusions. The literature review provides an overview of previous studies, while the research questions and methodology outline the study's objectives and the approach taken to collect and analyze data. The results section presents the findings, supported by visual aids, and the discussion section interprets and analyzes the results in the context of the research questions and relevant theories. The paper concludes by summarizing the main findings, restating the research questions, discussing the implications and contributions of the study, and suggesting potential future research directions. The overall structure ensures a logical progression of information and facilitates a comprehensive understanding of the study's outcomes.

2. METHOD

2.1. Context

To gain insight into how persona influences the user requirement, we conducted a set of empirical studies aimed at addressing the primary research query: do different personas lead to distinct conceptual designs customized for each personality?

In our research, we conducted empirical studies to investigate the influence of personality on the use of personas. Our main research question focused on determining whether personas with different personalities lead to distinct conceptual designs tailored to each personality. Participants were presented with personas characterized by varying personality traits and asked to propose designs addressing the personas' identified personal issues. The collected data was analyzed using a sequential mixed-method approach, combining qualitative, and quantitative methodologies. Initially, a qualitative analysis was employed to categorize the design concepts, followed by quantitative statistical methods to validate or refute the hypothesis. The interpretive aspects of the research were guided by grounded theory, which emphasizes immersion in the subject matter and data-driven analysis. Thematic analysis was employed to analyze and classify the design artifacts generated by participants, involving iterative examination, documentation, and reclassification of the data. The utilization of mixed methods and thematic analysis contributed to the generation of a comprehensive report and enhanced the validity and relevance of the findings [25]–[28].

The research process (Figure 1) began with a thorough competitor analysis to gain insights into the existing market landscape, with detail of each activity as given in Table 1. This analysis helped identify

competitors' strengths, weaknesses, and strategies, providing a foundation for further research. Building on this analysis, a hypothesis was formulated to guide the investigation. The next step involved identifying relevant behavioral variables that would help capture the key aspects of user behavior and preferences. This involved considering factors such as user demographics, attitudes, and needs. Subsequently, in-depth interviews were conducted with target users to gather qualitative data and map their experiences and perspectives. The information obtained from the interviews was then synthesized to identify common characteristics and extract relevant goals or objectives of the target user group. Finally, personas were created to represent distinct user profiles, incorporating the synthesized data to encapsulate the target users' key attributes, motivations, and goals. These personas served as valuable tools for understanding and empathizing with the users throughout the design and development process.

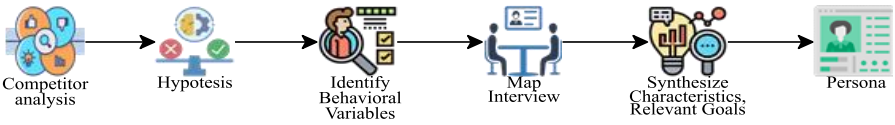


Figure 1. Research process

Table 1. Set activities purposed for personas (new activities in italics)

Activities	Sub activities	Ref.	Goal
Competitor analysis	a. A direct and indirect competitor b. Strength, weaknesses, opportunities, and threats (SWOT) analysis	[29]	Gain insights and understanding of competitors' strengths, weaknesses, strategies, and market positioning to inform decision-making and develop competitive advantages.
Hypothesis	a. Identify possible personas b. Hypothesis explanation	[29]	To propose a testable explanation or prediction that guides research and investigation into a specific phenomenon or relationship.
Identify behavioral variables	a. List of possible behavioral variables b. Variables explanation c. Variables scale	[29]	To identify and analyze the specific actions, attitudes, or patterns of behavior that are relevant to the research topic or study.
Map interview	a. Map interview b. Variables explanation	[29]	To systematically document and analyze the information gathered from interviews in order to identify key themes, patterns, and insights relevant to the research objectives.
Synthesize characteristics and relevant goals	a. Characteristics pattern b. Relevant goal explanation	[29]	To consolidate and integrate the collected information to identify common attributes and objectives that can guide the design and development process.
Persona			To develop a fictional representation of a target user, which helps to understand user needs, behaviors, and preferences in order to design products or services that align with their requirements.

We carried out three empirical studies encompassing three distinct case studies: a booking system for planes or trains, technology aimed at children, and technology designed for the elderly. The participants in these studies primarily consisted of novice developers currently pursuing studies in computer science at our university, with only a few individuals possessing expertise in software requirements and UI/UX. The beginner developers were final-year students who had successfully completed the required competency in requirements engineering as per the association for computing machinery (ACM) curriculum [30]. The total number of participants was 130, although some opted not to work on specific case studies in certain instances. The research was conducted on our university campus between late December 2022 and mid-February 2023. Each participant voluntarily engaged in all stages of the case studies over a two-week period for each individual case study. The participants were provided with general instructions outlining the tasks to be performed. For instance, in the case of the technology designed for the elderly, the participants were instructed: "you are tasked with providing a design solution (a new application) for a problem faced by the elderly. The expected outcome of this project is a useful and suitable solution that can compete with existing competitors". Through this instruction, the researchers aimed to encourage participants to explore their creativity. All participants worked on their downloaded UX journey worksheets and were requested to document their activities, including conducting interviews with potential users, throughout the process [31]. After the assignments, each participant was asked to assess the quality of the requirements gathered using the requirements matrix as a measure of productivity [31]. However, it should be noted that this measurement did not part of the research process, as the data was intended for productivity measurement purposes.

2.2. Modelling process as Markov chain

Numerous studies have been conducted on modeling and analyzing business processes [32], [33]. While existing process models in the literature primarily focus on representing the overall software process, they are unsuitable for quantitatively determining the similarity between processes or capturing individual task steps [32], [33]. The researchers employed Markov chains, commonly used in various domains, to model and analyze stochastic processes [34]. Markov chains facilitate the representation of stochastic processes by utilizing system states and transitions. In the context of a programmer's task process, each task corresponds to a distinct Markov chain, where states represent different conditions, and transition probabilities indicate the likelihood of transitioning between states during task execution. The researchers incorporated transition matrices to calculate the probability of transitioning from one state to another, enabling quantitative analysis and comparison of task processes.

In the proposed methodology, tasks are essential activities that need to be performed. These tasks are part of the UX journey method, as depicted in Figure 1. Developers execute each task to obtain user stories and personas. Each task is modeled using the Markov chain model to better understand the system's behavior over time. In this model, each state represents a specific condition within the process, and the transition probability indicates the likelihood of transitioning from one state to another. This probability quantifies the frequency of state transitions relative to the total number of state transfers. A notable aspect of employing the Markov chain is its application to the task process. Specifically, the Markov chain captures changes in the status or state of the ongoing task process. Each step within the task process corresponds to a state in the Markov chain, and the transition probability reflects the likelihood of moving from one step to another during the execution of the process. By utilizing the Markov chain model for the task process, it becomes possible to analyze and predict the trajectory of the process and comprehend how the preceding step influences the subsequent step. For instance, when a developer performs the "competitor analysis" activity, there is a probability of subsequently undertaking the "hypothesis" activity or the "identify behavioral variables" activity. The probability of transitioning from one activity to another is determined by the ratio of the number of times the developer performs the transition from the first activity to the second activity within the task process to the total number of times the developer advances from the third step and beyond.

2.3. Analysis difference between case study

By utilizing the Markov task process model, the disparities between the two task processes can be elucidated by evaluating the dissimilarity between their respective Markov chains. The Markov chain represents alterations in the state or condition of an ongoing task process. Within this framework, each individual step or activity within the task process is regarded as a distinct state within the Markov chain. At the same time, the transition probability denotes the likelihood of transitioning from one step to another during the execution of the process. The extent of divergence between the two Markov chains can be ascertained by scrutinizing the dissimilarity between the two task processes. This dissimilarity is gauged by comparing the transition probabilities and states manifesting at each Markov chain step. A more pronounced distinction in transition probabilities and states between the two Markov chains indicates a greater disparity between the task processes. Consequently, the Markov task process model is invaluable for quantitatively evaluating and measuring the variations between two task processes. The disparity between the two Markov chains is calculated using the state transition matrix and determining the sum of absolute differences by calculating the Manhattan distance.

3. RESULTS AND DISCUSSION

The outcomes of the comprehensive persona, including a user requirements matrix intended for productivity testing, are publicly accessible in the Indonesian language [35]. The discussion section encompasses multiple subsections, including illustrations of the persona creation process, a statistical analysis of the three case studies, an examination of activity similarities, and an assessment of their impact on productivity.

3.1. Persona creation process

In cooper's study, later updated by Silvia, persona creation was identified as an open and structured activity embedded within a sequence of tasks that involve specific objectives, techniques, and deliverables. However, previous publicly available empirical research suggests that developers need help in formulating hypotheses within the context of creative thinking. As part of hypothesis generation, participants were instructed to identify competitors using two assessment methods: SWOT analysis and competitor analysis. The provided worksheet includes a set of questions for SWOT analysis to assist participants in identifying each component. The competitor analysis section consists of three components: direct competitors, indirect

competitors, and feedback from users or developers. Tables 2 and 3 present illustrative examples of competitor analysis using case studies of train ticket booking systems.

Table 2. Example of SWOT analysis

SWOT	Questions	Example answer
Strength	Advantages?	The train booking system offers a user-friendly interface and seamless integration with multiple payment options for convenient transactions.
	Uniqueness?	The train booking system stands out with its exclusive partnerships with major railway companies, providing access to exclusive train routes and discounts.
	Selling points?	The train booking system's selling points include competitive pricing, comprehensive customer support, and additional services like seat selection and travel insurance options.
	Skills?	The train booking system demonstrates its expertise through skilled customer service representatives, experienced software developers, and proficiency in data analytics.
	Other factors?	The train booking system stands out with its eco-friendly initiatives, accessibility features, and integration with other transportation modes for seamless travel options.
Weaknesses	Limitations?	The train booking system has limitations in terms of the limited availability of regional train services, which may result in fewer options for certain destinations.
	Lack of effort?	The train booking system lacks sufficient effort in providing real-time updates on train delays or cancellations, leading to potential inconvenience for users.
	Problems?	The train booking system experiences occasional technical glitches, such as slow loading times or errors during the booking process, impacting user experience.
	Poor strategy?	The train booking system lacks a clear marketing strategy to effectively target and attract a diverse range of customers, potentially limiting its user base.
	Other factors?	The train booking system faces challenges in adapting to changing customer preferences and demands, such as the integration of emerging technologies or personalized travel recommendations.
Opportunities	Improvements?	Providing a more convenient and accessible booking experience for users.
	Performance?	By optimizing the system's performance and reducing response times, the train booking system can offer a faster and more efficient booking process, improving overall customer satisfaction.
	Opportunities?	Explore partnerships with travel agencies or other transportation services to offer bundled travel packages, expanding its range of services and attracting a wider customer base.
	Consumer behavior?	Promote its environmentally friendly features, such as reduced carbon emissions, to attract environmentally conscious travelers.
	Other factors?	Leveraging emerging technologies, such as artificial intelligence and machine learning, the train booking system can provide personalized recommendations and tailored travel experiences, increasing customer engagement and loyalty.
Threats	External trouble?	Strikes or infrastructure issues pose a threat to the reliability and availability of the train booking system.
	Obstacles?	Increased competition from alternative modes of transportation.
	Trends?	Leading to decreased usage and revenue because of the growing popularity of online travel agencies and aggregator.
	Other factors?	Changes in fare structures or restrictions can pose challenges and uncertainties for the train booking system's operations and profitability.

Table 3. Example of competitor analysis

Direct competitor	Indirect competitor
kai.id	Skyscanner
traveloka.com	Booking.com
TrainBooker	Expedia
Trainline	Kayak
Omio	Rome2rio
GoEuro	
User reviews	
Direct competitor app - "TrainBooker":	
Positive review: "TrainBooker is incredibly user-friendly and offers many train options. The booking process is seamless, and I love how I can easily compare prices and schedules. Highly recommended!"	
Negative review: "I had some issues with the TrainBooker app. The search results were not always accurate, and I encountered errors when completing my booking. Customer support was slow to respond, which was frustrating".	
Indirect competitor app - "Skyscanner":	
Positive review: "Skyscanner is my go-to app for travel bookings, and their train ticket feature is a great addition. Having all my travel options in one place is convenient, and the app always delivers reliable results".	
Negative review: "I found the train ticket booking feature on Skyscanner to be a bit lacking. The search results were limited, and sometimes the prices displayed were outdated. It would be great to see more improvement in this area".	

SWOT and competitor analysis, depicted in Tables 2 and 3, offers a comprehensive assessment of the existing market landscape. Consistent with prior research, the creative thinking process begins with broad considerations and potential divergent thinking, generating various design solutions that are refined into a primary solution through convergent thinking.

Identifying potential users and exploring possible challenges from the findings of the previous stage formed the foundation for constructing hypotheses, as presented in Table 4. This represents an advancement from previous research conducted by Silvia, where hypotheses were derived solely from developer assumptions without a solid basis for their formulation [29]. This limitation poses a significant hurdle for novice developers, particularly when they need to gain prior experience in developing user-centric design solutions. In contrast, our study incorporates an analysis of comparable or similar design solutions, providing developers with a more informed basis. Nevertheless, consistent with earlier research, hypotheses are employed to uncover variables that may differentiate users by identifying distinctive needs and behaviors.

Table 4. Example of hypothesis

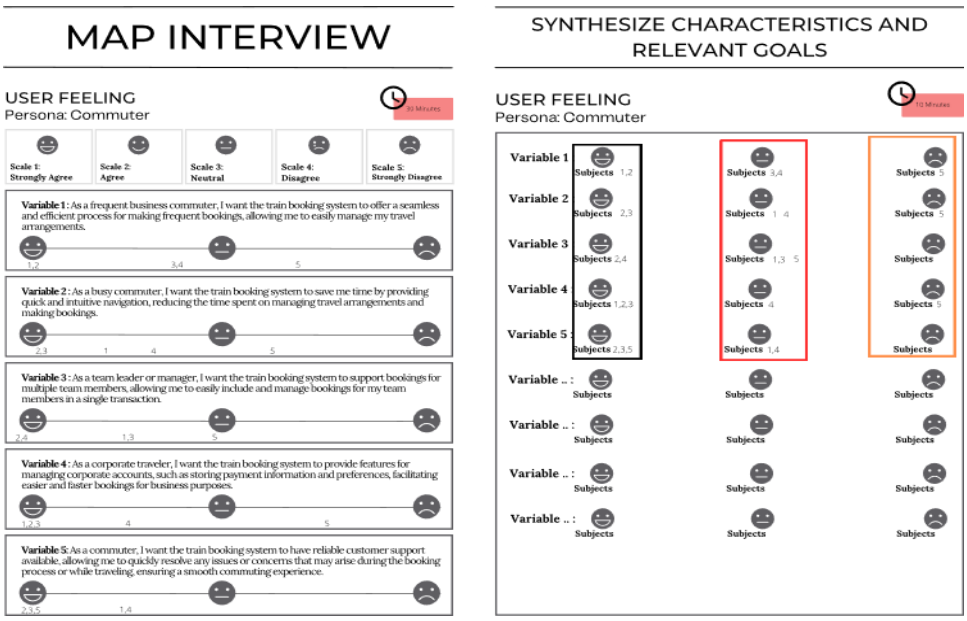
Hipotesis	Persona	Explanation
Ho	Commuter	To find the most convenient and cost-effective train routes for daily commuting to work.
H1	Leisure traveler	To explore new destinations and plan memorable train journey with friends or family.
H2	Business traveler	To efficiently manage business travel arrangements, including train bookings, for himself and his team.

In the study conducted by Silvia, the term "identified behavioral variables" was introduced as a novel concept [29], contrasting with the terminology used in cooper's research [17]. However, we find it more suitable to refer to these variables as "observed behavioral variables" since they will serve as the focus of our observations. While Silva emphasizes the newness of this activity, cooper's research recommends creating a comprehensive list of behavioral variables derived from various aspects of observed behavior. We acknowledge Silva's observation that obtaining information through interviews can challenge developers. However, our research has yet to provide a detailed explanation of the process of forming this list. Hence, our study formulated the list as a compilation of observed insights derived from the SWOT analysis, particularly in the opportunity section, as presented in Table 5.

Table 5. Example of observed behavioral variables

Persona		Variable	Scale
Commuter	a.	Frequency of train bookings for business purposes.	Strongly disagree-strongly agree
	b.	Time spent on managing travel arrangements.	
	c.	The number of team members included in the bookings.	
	d.	Utilization of corporate account management features.	
	e.	Engagement with customer support for issue resolution.	
Leisure traveler	a.	The number of train bookings made for leisure travel.	Strongly disagree-strongly agree
	b.	Variety of destinations chosen.	
	c.	Interaction with personalized recommendations and itinerary customization options.	
	d.	Engagement with social features for sharing travel experiences.	
	e.	Utilization of group booking features for traveling with friends or family.	
Business traveler	a.	Frequency of train bookings for commuting purposes.	Strongly disagree-strongly agree
	b.	Time spent searching for suitable train routes.	
	c.	Utilization of search filters to find convenient schedules.	
	d.	Engagement with features providing cost comparisons and discounts.	
	e.	Feedback or complaints related to the availability and reliability of train services for commuting.	

Prior to conducting the map interview activity with participants, the developer will transform the statements related to the observed behavioral variables into user stories format, following the structure "as (user) I want/need (something) so I can do (something)" [36]. This format enhances the clarity and purpose of the interview process. According to Silvia's research, the precision of the interview itself is of little importance, as the main objective is to explore the relationship between different respondents based on the predefined persona derived from the hypothesis. The outcomes of the interviews are represented using smileys, with labels obtained from defining the observed variables. The developer's worksheet displays these labels in a five-scale format, but respondents can indicate a grayscale that falls between the two predetermined scales. An example of a map interview representation is illustrated in Figure 2(a). Subsequently, significant variables are identified from the map interview activity. The aim is to identify and group respondents who exhibit similar patterns and values based on the proximity of their indicated scales. A group typically comprises six to twelve respondents choosing the same or closely related variables. However, in our research, we limit the closeness to three to six respondents to gather more comprehensive information, without compromising the accuracy of the design solution pursued. A representation of this activity is presented in Figure 2(b). The established groups serve as the foundation for constructing a persona.



(a) (b)

Figure 2. Example of interview: (a) map interview and (b) synthesize characteristics and relevant goals

An in-depth understanding of the respondents' backgrounds and preferences plays a crucial role in guiding developers in defining the objectives of their development efforts. The persona creation process serves as a means to represent potential system users. A key aspect of persona development is prioritizing target users, ensuring that the developer focuses on specific user groups sample of the persona from the method as given in Figure 3. This is achieved through the synthesis of relevant characteristics and goals that help developers gain a deeper understanding of their intended user segments. To facilitate this process, this study proposes a comprehensive set of persona components, as outlined in Table 6, that enable developers to precisely identify the target audience.

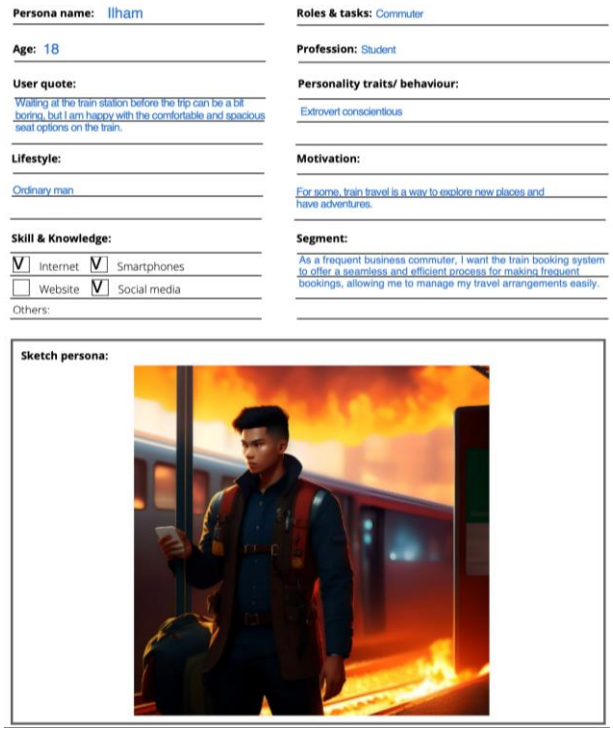


Figure 3. Example of persona

Table 6. Proposed persona structure

Component	Content	Example
Persona name	Target users with specific characteristics, goals, and pain points, enabling designers and developers to create solutions tailored to their needs and preferences	Ilham: identified Ilham's strength in efficiently managing travel arrangements and her expertise in finding cost-effective train routes.
Age	Age information helps create a more realistic and relatable persona that aligns with the target audience's demographics, needs, and behaviors.	18 years: consider any strengths identified in the SWOT analysis that are related to age, such as a persona's familiarity with technology or specific preferences that align with a particular age group.
User quote	User quotes provide direct insights and perspectives from real users, capturing their thoughts, opinions, and experiences.	Waiting at the train station before the trip can be a bit boring, but i am happy with the comfortable and spacious seat options on the train: identify strengths of the product or service that can be highlighted in user quotes to showcase the positive aspects experienced by users.
Lifestyle	Represents the daily routines, preferences, and behaviors of the target user	Ordinary man: if the product offers convenience and time-saving features, it can appeal to users with a busy or fast-paced lifestyle.
Skill and knowledge	Represents the user's proficiency and familiarity with various digital tools and platforms.	If the product offers a user-friendly interface or integrates well with popular social media platforms, it can appeal to users who are already proficient in using such tools.
Role and tasks	Represents the user's specific responsibilities, roles, and tasks in relation to the product or service being analyzed.	Commuter: identify the strengths of the product or service that align with specific user roles or tasks
Profession	Represents the user's occupation, field of work, or professional background	Student: if the product offers advanced data analysis capabilities, it may be well-suited for professionals in data-driven fields such as marketing or finance.
Personality traits/behaviour	Represents the characteristic patterns of thoughts, feelings, and behaviors that influence how a user interacts with a product or service	Extrovet conscientious: if the product offers a user-friendly interface and intuitive navigation, it may appeal to users who prefer simplicity and ease of use.
Motivation	Represents the underlying needs, desires, and goals that drive a user's behavior and decision-making process.	If the product offers time-saving features or cost-effective solutions, it may appeal to users motivated by efficiency or financial considerations.
Segment	Represents the specific target audience or user group with similar characteristics, needs, or preferences.	As a frequent business commuter, i want the train booking system to offer a seamless and efficient process for making frequent bookings, allowing me to easily manage my travel arrangements.
Sketch persona	The visual representation of the persona and includes key information and characteristics	

For novice developers, a clear understanding of the target audience is paramount to ensure the development efforts align with user expectations. Unlike previous research conducted by Silvia, which categorized respondents into primary and secondary groups, this study treats all respondents equally and considers them valuable development references. Furthermore, the map interview conducted during the research process allows developers to obtain valuable insights and elicit system requirements directly from the respondents. The outcomes of the identification process outlined in Table 6 are subsequently transformed into personas, which are documented in the UX journey worksheet. It is important to note that personas are fictional representations that may not necessarily correspond to an individual respondent but serve as archetypes that embody the characteristics and needs of specific user groups.

3.2. Statistical analysis

Statistical analysis is a crucial component of research that involves applying various statistical techniques to analyze and interpret data. It plays a significant role in gaining insights, making inferences, and drawing conclusions from research findings.

In project 1 (Table 7), the SWOT analysis revealed a minimum score of 1, a maximum score of 180, an average score of 16, and a standard deviation of 19. This suggests a moderate level of evaluation across the strengths, weaknesses, opportunities, and threats. The competitor analysis indicated a range of 3 to 200, with an average of 20 and a standard deviation of 19, indicating a relatively high level of competition. The hypothesis and identified behavioral variable analysis showed a minimum score of 2, a maximum score of

60, an average score of 11, and a standard deviation of 8.1, indicating a reasonable understanding of user needs and behaviors. The map interview analysis had a range of 2 to 100, an average score of 12, and a standard deviation of 13, suggesting moderate user engagement. The characteristics and relevant goals analysis showed a range of 2 to 90, an average score of 19, and a standard deviation of 16, indicating a good understanding of user characteristics and goals. The persona analysis had a minimum score of 1, a maximum score of 75, an average score of 11, and a standard deviation of 11, suggesting room for improvement in persona development.

Table 7. Time to complete (in minutes)

	Project 1				Project 2				Project 3			
Activity	Min	Max	Avg	STD	Min	Max	Avg	Std	Min	Max	Avg	STD
Swot	1	180	16	19	1	180	16	19	1	180	16	19
Competitor	3	200	20	19	1.2	200	21	21	3	200	20	19
Hypothesis	2	60	11	8.1	1	80	13	12	2	60	11	8.1
Identify behavioral variable	1	60	12	8.8	1	90	14	13	1	60	12	8.8
Map interview	2	100	12	13	1.4	120	15	19	2	100	12	13
Characteristics and relevant goals	2	90	19	16	2	120	19	18	2	90	19	16
Persona	1	75	11	11	1	60	14	11	1	75	11	11

In project 2 (Table 7), the SWOT analysis had a minimum score of 1, a maximum score of 180, an average score of 16, and a standard deviation of 19, indicating a similar evaluation as in project 1. The competitor analysis showed a range of 1 to 200, an average score of 21, and a standard deviation of 21, suggesting a slightly higher competition level than project 1. The hypothesis and identified behavioral variable analysis had a minimum score of 1, a maximum score of 80, an average score of 13, and a standard deviation of 12, indicating a reasonable understanding of user needs and behaviors. The map interview analysis had a range of 1 to 4, an average score of 120, and a standard deviation of 15, indicating a moderate level of engagement with users. The characteristics and relevant goals analysis showed a range of 2 to 120, an average score of 19, and a standard deviation of 18, suggesting a good understanding of user characteristics and goals. The persona analysis had a minimum score of 1, a maximum score of 60, an average score of 14, and a standard deviation of 11, indicating room for improvement in persona development.

In project 3 (Table 7), the SWOT analysis had a minimum score of 1, a maximum score of 180, an average score of 16, and a standard deviation of 19, similar to the previous projects. The competitor analysis indicated a range of 3 to 200, an average score of 20, and a standard deviation of 19, showing a similar level of competition as in project 1. The hypothesis and identified behavioral variable analysis had a minimum score of 2, a maximum score of 60, an average score of 11, and a standard deviation of 8.1, indicating a similar understanding of user needs and behaviors as in project 1. The map interview analysis had a range of 2 to 100, an average score of 12, and a standard deviation of 13, suggesting a similar level of user engagement as in project 1. The characteristics and relevant goals analysis showed a range of 2 to 90, an average score of 19, and a standard deviation of 16, indicating a similar understanding of user characteristics and goals as in project 1. The persona analysis had a minimum score of 1, a maximum score of 75, an average score of 11, and a standard deviation of 11, suggesting a need for improvement in persona development, similar to project 1.

3.3. Analyzing the similarity between activity

Statistical analysis is a crucial component of research that involves applying various statistical techniques to analyze and interpret data. It plays a significant role in gaining insights, making inferences, and drawing conclusions from research findings.

3.3.1. Modelling as Markov chains

The Markov chain process is a mathematical analysis process that models a system's transition from one state to another. It is a stochastic process that satisfies the Markov property, which states that the future state depends only on the current state and is independent of the past states. In a Markov chain, a system can exist in a finite or countably infinite number of states. Transition probabilities govern the transition from one state to another. These probabilities represent the likelihood of moving from one state to another in a single step. The mathematical analysis of a Markov chain involves studying the transition probabilities and using them to calculate various system properties. Some common analyses include determining the steady-state probabilities (also known as the stationary distribution), representing the system's long-term behavior,

calculating expected values and probabilities of reaching specific states, and predicting future events based on the current state (1).

$$P^{(h)} = \left[P_{ij}^{(h)}\right]_{NxN} = \begin{bmatrix} P_{11}^{(h)} & P_{12}^{(h)} & P_{1n}^{(h)} \\ P_{21}^{(h)} & P_{22}^{(h)} & P_{2n}^{(h)} \\ P_{n1}^{(h)} & P_{n2}^{(h)} & P_{nn}^{(h)} \end{bmatrix} \quad (1)$$

This matrix is also stochastic because each row represents the conditional distribution. Where P is the transitional matrix, and $\left[P_{ij}\right]_{NxN}$is the transitional matrix with NxN probability in discrete time Markov chain with the conditions $\{X_n, n \geq 0\}$ in the state space $h = \{1, 2, \dots, N\}$ based on space that predict. From this condition $P_{ij} \geq i, j \leq N$and for the first space, $\sum_j^i = 1\ P_{ij}^h = 1, 1 \leq i \leq N$. For the 1-step, 2-step, n-step transition, as (2)-(4):

$$P^{(1)} = P^{(0)} \times P \quad (2)$$

$$P^{(2)} = P^{(1)} \times P = P^{(0)} \times P \times P = P^{(0)} \times P^2 \quad (3)$$

$$P^{(n)} = P^{(n-1)} \times P = \cdots = P^{(0)} \times P^n \quad (4)$$

The transition probabilities in the provided Markov chain describe the likelihood of moving between different process stages. The chain starts with the SWOT stage, where the transition to the competitor stage is probability 1, meaning it always occurs. From the competitor stage, the process always moves to the hypothesis stage with a probability of 1. A sample task process is given in Figure 4, describing each step's flow and probability.

$$[next\_state] = \begin{bmatrix} Matrix\ of \\ transition \\ probabilities \end{bmatrix} [curent\_state] \quad (5)$$

$$X_1 = X_0 \times P \quad (6)$$

$$\begin{bmatrix} A_1 \\ B_1 \\ C_1 \\ D_1 \\ E_1 \\ F_1 \\ G_1 \end{bmatrix} = \begin{bmatrix} 1 & & & & & & \\ 1 & & & & & & \\ & 1 & & & & & \\ & & 0.33 & & & & \\ & & 0.33 & 0.5 & & & \\ & & 0.33 & 0.5 & 1 & & \\ & & & & & 1 & \end{bmatrix} \begin{bmatrix} 16 \\ 20 \\ 11 \\ 12 \\ 12 \\ 19 \\ 11 \end{bmatrix} \begin{bmatrix} 0.158 \\ 0.198 \\ 0.108 \\ 0.118 \\ 0.118 \\ 0.188 \\ 0.108 \end{bmatrix} = \begin{bmatrix} 0.158 + 0.198 \\ 0.198 \\ 0.108 \\ 0.33 \times 0.108 \\ 0.33 \times 0.108 + 0.5 \times 0.118 \\ 0.33 \times 0.108 + 0.5 \times 0.118 + 1 \times 0.188 \\ 0.108 \end{bmatrix} = \begin{bmatrix} 0.356 \\ 0.198 \\ 0.108 \\ 0.0356 \\ 0.0946 \\ 0.282 \\ 0.108 \end{bmatrix}$$

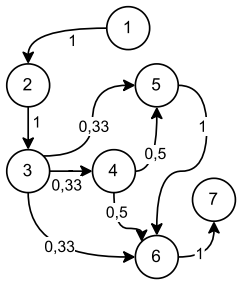


Figure 4. A sample task process of average productivity modeled as a Markov chain

Within the hypothesis stage are three possible transitions, each with a probability of 0.33. This indicates an equal chance of moving to identify behavioral variable, map interview, or characteristics and relevant goals. Once in the identify behavioral variable stage, there is an equal probability of 0.5 to transition either to the map interview stage or the characteristics and relevant goals stage. The transition from map interview to characteristics and relevant goals is certain, with a probability of 1. Markov chain represents a

sequential process where the initial stages (SWOT, competitor) always lead to subsequent stages (hypothesis, identify behavioral variable), and from there, the process branches out with equal probabilities to different stages (map interview, characteristics and relevant goals) before eventually reaching the final stage persona.

3.3.2. Difference between high-average activity

Average productivity refers to the typical or usual level of productivity observed within a given context. It represents the average or mean level of output or efficiency achieved by the entities being considered. Hence, high productivity refers to a level of productivity that exceeds the average or typical performance. It represents a higher-than-normal output or efficiency level achieved by individuals. High productivity implies that the entities are performing exceptionally well and achieving output levels surpassing the average or standard.

In project 1 (Figure 5), the minimum value observed is 22.5, and the maximum value is 490, indicating the range of the data. The first quartile (Q1) is 58, representing the 25th percentile and the third quartile (Q3) is 116, representing the 75th percentile. The median, calculated as 80, represents the midpoint of the data. The mean for project 1 is approximately 100.08, indicating the overall average. The interquartile range (IQR), measuring the spread of the middle 50% of the data, is 58. The lower limit is -29 and the upper limit is 203, defining the expected range for most of the data.

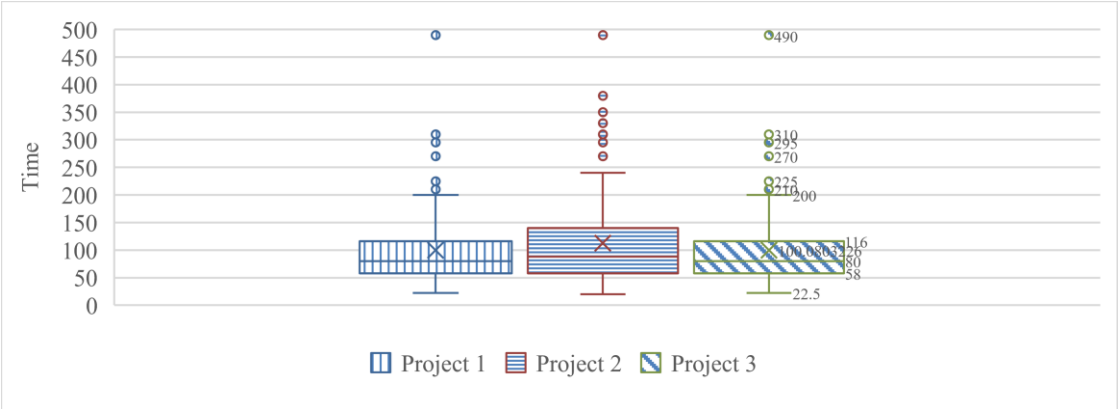


Figure 5. Difference high-average activity

In project 2 (Figure 5), the minimum and maximum values are 20 and 490, respectively, showing the range of the data. The first quartile is 58, the median is 88, and the third quartile is 140. The mean for project 2 is approximately 112.51, slightly higher than that of project 1. The IQR is calculated to be 82, indicating a wider data spread. The lower limit is -65, and the upper limit is 263. Project 3 (Figure 5) statistical measures are identical to those of projects 1 and 2, with the same minimum, maximum, quartiles, median, mean, IQR, lower limit, and upper limit. Considering minimum and maximum values provides insights into the potential presence of outliers or extreme data points. The range of values from 22.5 to 490 indicates the variability within each project's data.

3.5. Discussion

Software development plays a critical role in improving productivity within the realm of software engineering. As the introduction highlights, software products have far-reaching implications across various dimensions, including technical, social, emotional aspects, work-life dynamics, and collaborative teamwork. To ensure that software effectively meets the needs of end users, developers employ personas to understand user characteristics and requirements. The persona activity involves creating representations of target user groups, enabling developers to gain valuable insights into user preferences, behaviors, and goals. Compared to previous research [29], this research provides a comprehensive solution for representing user targets using characteristic and relevant goals. This understanding helps guide the design and development process, ensuring that software meets user needs and expectations. However, user persona development has challenges. Balancing qualitative and quantitative techniques can be complex, and the subjective determination of personas can pose difficulties, particularly when integrating them into the software development process.

This is where the application of a Markov chain model becomes relevant. By modeling developer productivity and predicting the probabilities of the next states, the Markov chain offers a powerful tool for

optimizing software development processes. The model considers historical or real-time data, estimating transition probabilities based on past observations. This predictive capability enables project managers and team leads to plan resources effectively, allocate tasks efficiently, and identify potential areas for improvement.

Integrating the Markov chain model with the persona activity provides a comprehensive approach to software development. By incorporating user characteristics, behaviors, and goals derived from personas, the model can predict the next state of developer productivity more accurately. This correlation between persona insights and the Markov chain's predictive power allows a more nuanced understanding of how user requirements and preferences impact developer productivity. However, it is important to note that the accuracy of the Markov chain model relies heavily on the quality and relevance of the data used to estimate transition probabilities. Additionally, the model assumes that future developer behavior depends solely on their current state and is independent of their history. These assumptions, while simplifying the modeling process, may only sometimes capture the complexity of real-world scenarios. Integrating the Markov chain model with the persona activity offers a powerful approach to optimizing software development processes. By leveraging persona insights and predicting probabilities of the next states, organizations can enhance resource allocation, identify areas for improvement, and ultimately deliver software that better meets the needs and expectations of end users. Further research and refinement of the model and the continued use of personas can contribute to more accurate predictions and improved software development outcomes.

4. CONCLUSION

Integrating the Markov chain model in the software development process offers valuable insights into predicting the next state of developer productivity. By following a structured sequence of activities, such as conducting a SWOT analysis, analyzing competitors, formulating hypotheses, identifying behavioral variables, mapping interviews, and creating personas, organizations can gather relevant data to feed into the Markov chain model. This predictive approach enables better resource allocation and decision-making, ultimately improving project planning and successful software development outcomes. By harnessing the power of the Markov chain and leveraging the information obtained from the persona-building process in projects 1, 2, and 3, organizations can enhance their ability to forecast developer productivity and optimize their software development endeavors. Integrating quantitative analysis and qualitative understanding provides a valuable framework for improving efficiency and productivity in software development projects.

ACKNOWLEDGEMENTS

This project was funded by the Ministry of Higher Education (Kementerian Pengajian Tinggi) and Research Management Centre (RMC), Universiti Putra Malaysia for supporting/funding this article under its Fundamental Research Grant Scheme (FRGS)–project code 08-01-20-2319FR–5540451 and Universitas Muhammadiyah Malang for the APC support.

REFERENCES

[1] L. Goldberg *et al.*, "Usability and accessibility in consumer health informatics: Current trends and future challenges," *American Journal of Preventive Medicine*, vol. 40, no. 5 SUPPL. 2, pp. S187–S197, May 2011, doi: 10.1016/j.amepre.2011.01.009.

[2] M. Yilmaz, R. V. O'Connor, R. Colomo-Palacios, and P. Clarke, "An examination of personality traits and how they impact on software development teams," *Information and Software Technology*, vol. 86, pp. 101–122, Jun. 2017, doi: 10.1016/j.infsof.2017.01.005.

[3] K. W. Wagner and W. Dürr, "Design failure cost as a measure of a process measurement system (a method for building the system and evaluating the measure)," in *Software Engineering and Advanced Applications, 2005. 31st EUROMICRO Conference*, IEEE, 2005, pp. 214–221, doi: 10.1109/EUROMICRO.2005.20.

[4] D. Graziotin, X. Wang, and P. Abrahamsson, "Do feelings matter? On the correlation of affects and the self-assessed productivity in software engineering," *Journal of Software: Evolution and Process*, vol. 27, no. 7, pp. 467–487, Jul. 2015, doi: 10.1002/smr.1673.

[5] A. N. Meyer, G. C. Murphy, T. Zimmermann, and T. Fritz, "Design recommendations for self-monitoring in the workplace: Studies in software development," *Proceedings of the ACM on Human-Computer Interaction*, vol. 1, no. CSCW, pp. 1–24, Dec. 2017, doi: 10.1145/3134714.

[6] S. Parashivamurthy and N. G. Cholli, "Software aging prediction – a new approach," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 2, pp. 1773–1781, Apr. 2023, doi: 10.11591/ijece.v13i2.pp1773-1781.

[7] W. A. R. W. Mohd Isa, A. I. Hakim Suhaimi, N. Noordin, and R. A. Hashim, "Exploring mobile learning development lifecycle with cultural context," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 2, pp. 903–909, Nov. 2020, doi: 10.11591/ijeecs.v20.i2.pp903-909.

[8] I. Fatema and K. Sakib, "Factors Influencing Productivity of Agile Software Development Teamwork: A Qualitative System Dynamics Approach," in *Proceedings - Asia-Pacific Software Engineering Conference, APSEC*, IEEE, Dec. 2017, pp. 737–742, doi: 10.1109/APSEC.2017.95.

[9] H. Asfa and T. J. Gandomani, "Software quality model based on development team characteristics," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 1, pp. 859–871, Feb. 2023, doi: 10.11591/ijece.v13i1.pp859-871.
[10] S. Kujala, "Effective user involvement in product development by improving the analysis of user needs," *Behaviour and Information Technology*, vol. 27, no. 6, pp. 457–473, Nov. 2008, doi: 10.1080/01449290601111051.
[11] M. Miller and A. Hadwin, "Scripting and awareness tools for regulating collaborative learning: Changing the landscape of support in CSCL," *Computers in Human Behavior*, vol. 52, pp. 573–588, Nov. 2015, doi: 10.1016/j.chb.2015.01.050.
[12] Y. N. Chang, Y. K. Lim, and E. Stolterman, "Personas: From theory to practices," in *ACM International Conference Proceeding Series*, New York, NY, USA: ACM, Oct. 2008, pp. 439–442, doi: 10.1145/1463160.1463214.
[13] J. Brickey, S. Walczak, and T. Burgess, "A comparative analysis of persona clustering methods," *16th Americas Conference on Information Systems 2010, AMCIS 2010*, vol. 1, pp. 705–715, 2010.
[14] S. Faily and I. Flechais, "Persona cases: A technique for grounding personas," in *Conference on Human Factors in Computing Systems - Proceedings*, New York, NY, USA: ACM, May 2011, pp. 2267–2270, doi: 10.1145/1978942.1979274.
[15] L. Laporte, K. Slegers, and D. De Grooff, "Using correspondence analysis to monitor the persona segmentation process," in *NordiCHI 2012: Making Sense Through Design - Proceedings of the 7th Nordic Conference on Human-Computer Interaction*, New York, NY, USA: ACM, Oct. 2012, pp. 265–274, doi: 10.1145/2399016.2399058.
[16] J. Brickey, S. Walczak, and T. Burgess, "Comparing semi-automated clustering methods for persona development," *IEEE Transactions on Software Engineering*, vol. 38, no. 3, pp. 537–546, May 2012, doi: 10.1109/TSE.2011.60.
[17] A. Cooper, R. Reimann, and D. Cronin, "About Face 3: The essentials of interaction design," *Information Visualization*, vol. 3, no. 3, p. 610, 2007.
[18] N. Gould and D. Atkin, "Towards a semantic layer to support road and public transport user decision-making," in *Proceedings - 15th IEEE International Conference on Computer and Information Technology, CIT 2015, 14th IEEE International Conference on Ubiquitous Computing and Communications, IUCC 2015, 13th IEEE International Conference on Dependable, Autonomic and Secure Computing, DASC 2015 and 13th IEEE International Conference on Pervasive Intelligence and Computing, PICom 2015*, IEEE, Oct. 2015, pp. 1498–1503, doi: 10.1109/CIT/IUCC/DASC/PICOM.2015.224.
[19] J. A. Ejiwale, "The Changing Global Context of Virtual Workforce," *Journal of Education and Learning (EduLearn)*, vol. 6, no. 4, p. 231, Sep. 2012, doi: 10.11591/edulearn.v6i4.167.
[20] T. Adlin and J. Pruitt, *The Essential Persona Lifecycle: Your Guide to Building and Using Personas*. Elsevier, 2010, doi: 10.1016/C2009-0-62475-2.
[21] J. S. Pruitt and T. Adlin, *The Persona Lifecycle: Keeping People in Mind Throughout Product Design*. Elsevier, 2006, doi: 10.1016/B978-0-12-566251-2.X5000-X.
[22] D. D. K. Goodwin, "Age Create Hum," *Centered Prod. Serv*, 2009.
[23] R. Gudjonsdottir and S. Lindquist, "Personas and scenarios: Design tool or a communication device?," *8th International Conference on the Design of Cooperative Systems*, pp. 165–176, 2008.
[24] F. Anvari and H. M. T. Tran, "Holistic personas and reflective concepts for software engineers," *Proceedings of the 8th European Conference on Information Management and Evaluation, ECIME 2014*, pp. 20–28, 2014.
[25] S. Cronholm and A. Hjalmarsson, "Experiences from sequential use of mixed methods," *Electronic Journal of Business Research Methods*, vol. 9, no. 2, pp. 87–95, 2011.
[26] J. W. Creswell, *Research design qualitative, quantitative and mixed methods approaches*. 2003.
[27] B. G. Glaser and A. L. Strauss, *Discovery of grounded theory: Strategies for qualitative research*. Routledge, 2017, doi: 10.4324/9780203793206.
[28] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, Jan. 2006, doi: 10.1191/1478088706qp063oa.
[29] S. T. Acuña, J. W. Castro, and N. Juristo, "A HCI technique for improving requirements elicitation," *Information and Software Technology*, vol. 54, no. 12, pp. 1357–1375, Dec. 2012, doi: 10.1016/j.infsof.2012.07.011.
[30] CC2020 Task Force, *Computing Curricula 2020: Paradigms for Global Computing Education*. New York, NY, USA: ACM, 2020, doi: 10.1145/3467967.
[31] W. A. Kusuma, "UX Journey-The process to optimize your resource for the quality User Requirement," 2023.
[32] H. K. Meena, I. Saha, K. K. Mondal, and T. V. Prabhakar, "An approach to workflow modeling and analysis," in *Proceedings of the 2005 OOPSLA Workshop on Eclipse Technology eXchange, eclipse'05*, New York, New York, USA: ACM Press, 2005, pp. 85–89, doi: 10.1145/1117696.1117714.
[33] A. Raj, A. Agrawal, and T. V Prabhakar, "Transformation of Business Processes into UML Models: An SBVR Approach," *International Journal of Scientific & Engineering Research*, vol. 4, no. 7, p. 2013, 2013.
[34] B. Sericola, "Markov Chains: Theory and Applications," *Markov ChainsTheory AndApplications*, 2013.
[35] W. Kusuma, A. Hazri, N. I. Admodisastro, and N. Norowi, "UX Journey Novice Developer," *Mendeley Data*, 2023, doi: 10.17632/6bvbhy3wxp.1.
[36] S. Nasiri, A. Adadi, and M. Lahmer, "Automatic generation of business process models from user stories," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 1, pp. 809–822, Feb. 2023, doi: 10.11591/ijece.v13i1.pp809-822.

BIOGRAPHIES OF AUTHORS

Wahyu Andhyka Kusuma is a lecturer at Universitas Muhammadiyah Malang. He has published articles on Human-Computer Interaction, User Experience, and Solo Software Processes. He graduated with first-class honors B.Cs. degree in Universitas Muhammadiyah Malang, Malang, Indonesia in 2012, and a Master's in Computer Science from Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. He has been studying Ph.D. Human-Computer Interaction at Universiti Putra Malaysia, Malaysia. He can be contacted at wahyukusuma@umm.ac.id and gs63875@student.upm.edu.my.

Azrul Hazri Bin Jantan is a professor at the Department of Multimedia, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia, where he has been a faculty member since 2002. Azrul graduated with a first-class honors B.Eng degree in Management Information Systems (UTM, Johor), Malaysia in 2001, and an Information Technology (University of Adelaide, Australia), Australia in 2004. Completed his Ph.D in Web Engineering (Design and Development) (USM, Penang), Malaysia 2009. His research area and specialization include Computational Hypermedia and Human-Computer Interaction, where he is the author/co-author of over 50 research publications. He can be contacted at email: azrul@upm.edu.my.

Novia Indriaty Admodisastro is an Associate Professor at the Faculty of Computer Science and Information Technology at Universiti Putra Malaysia. She holds a Ph.D. in Computing from Lancaster University, UK (2012), a Master of Software Engineering from Universiti Malaya (2003), and a Bachelor of Computer Science from Universiti Putra Malaysia (1999). She has published a number of papers in journals, book chapters, and proceedings in the areas of software engineering, cloud engineering, and human-computer interaction, as well as cross-discipline studies in education and medical technologies. She can be contacted at email: novia@upm.edu.my.

Noris Mohd Norowi is a Senior Lecturer at the Department of Multimedia, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia. She has a Ph.D in Computer Music from the University of Plymouth, United Kingdom, as well as a Master of Science in Multimedia Systems and a degree in Computer Science (Majoring in Multimedia), both from the Universiti Putra Malaysia. Her field of study is Human-Computer Interaction. The focus of her research includes mobile technologies, immersive technologies, music interaction, and interaction design. She can be contacted at email: noris@upm.edu.my.