Analysing User Reviews for Evaluating Game Playability of Mobile Gaming Apps

by

Swapnil Thakar
202111077

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of

MASTER OF TECHNOLOGY
in
INFORMATION AND COMMUNICATION TECHNOLOGY
to

DHIRUBHAI AMBANI INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGY

July, 2023

Declaration

I hereby declare that

i) the thesis comprises of my original work towards the degree of Master of Technology in Information and Communication Technology at Dhirubhai Ambani Institute of Information and Communication Technology and has not been submitted elsewhere for a degree,

ii) due acknowledgment has been made in the text to all the reference material used.

Swapnil Thakar

Certificate

This is to certify that the thesis work entitled **Analysing User Reviews for Evaluating Game Playability of Mobile Gaming Apps** has been carried out by **Swapnil Thakar** for the degree of Master of Technology in Information and Communication Technology at *Dhirubhai Ambani Institute of Information and Communication Technology* under my/our supervision.

19/7/23

Dr. Saurabh Tiwari
Thesis Supervisor

Acknowledgments

First and foremost, I would like to express the deepest gratitude my supervisor, Dr. Saurabh Tiwari for their invaluable guidance, continuous support, and expertise throughout the research process. His mentorship and encouragement have been instrumental in shaping my research direction and enhancing the quality of my work. He consistently provided unwavering support and expert guidance, readily assisting me whenever I encountered obstacles during my research work.

I would also like to thanks to my parents, my family, and my friends for their continuous support.Their love, patience, and belief in my abilities have been a constant source of motivation.

Contents

Abstract

The playability of a game depends on the players' experience in terms of functionality, usability, and satisfaction. Mobile gaming has recently evolved because of the availability of suitable hardware, configurable mobile devices, and the ability to download games from the Android and iOS platforms. Most online gaming stores allow customers to submit their reviews about gameplay, issues, and functionalities publicly. Game developers can better grasp such consumer issues by examining player feedback and increasing how well-liked a game is among players. We have mapped the playability of Sánchez's model with Schwartz's theory of human values and analyzed 20,346 user/player reviews from the top 15 game apps in the Google Play Store. We have also created a labelled dataset of each playability category of Sánchez's model. Finally, we applied a machine learning model to support the automatic classification of a review to a specific playability category violation. Our analysis shows that 30% of the reviews show human values violations, consequently affecting game playability. We found that *Socialism* is the most violated and *Emotion* is the least violated value category. We also found that only 18% of the user reviews received responses from the game app developers for the value violations. Using fine-grained feature extraction, we found the top 42 functionalities, issues, and concerns for the violations. The analysis results of our study give developers a foundation for creating apps that consider users' values for ensuring better playability of mobile game apps.

List of Tables

List of Figures

Chapter 1
Introduction

1.1 Game Playability

Since smartphones have become commonplace, app development has risen exponentially to simplify people's lives and give them access to everything they need. We all have smartphones today, and due to their functionality, we all love utilizing the many apps on them more and more frequently. With a revenue of $935 billion by 2023, Google Play's popularity has significantly increased over the last ten years. In 2022, there were more than 2.9 million apps in the store, which had 218 billion downloads [19]. Nowadays, people download highly sophisticated games from app stores, and although most are cheap to buy or often free, publishers make a considerable amount of revenue from in-app purchases [13]. The worldwide video game sector's tremendous increase is now driven by mobile gaming. Mobile games income is predicted to reach $100 billion by 2023 as mobile penetration rates and smartphone usage continue to rise globally. Smartphone games made up 45 percent of global video game revenue in 2022 [21].

Playability suggests that it is an idea that may be used on various levels. The simplest level is when a person interacts with media content or technology. The interaction between media creators, distributors, and consumers can be seen as a game or as playful on a more abstract level. Each player in this game may be seen to be employing a specific strategy, which in some instances leads to a highly dynamic interplay between the players. Regardless of where the people belong and their community, anyone can search and download games and start playing right away [22]. Given this enormous financial potential, it is hardly surprising that the gaming market is still very competitive. The market is expanding, and there is still massive potential for success. The development of mobile games on Android has increased the scalability of the mobile gaming industry. Due to the majority of Android users, platforms like Netflix, Facebook, and Instagram started featuring android mobile games before iOS games to employ appealing games to strengthen their advertising strategies [20]. Playability is a term very commonly used to assess the quality of games from different perspectives. Sánchez et al. [44] proposed a definition and framework for playability focusing on the players' experience.

"A set of properties that describe the player experience using a specific game system whose main objective is to provide enjoyment and entertainment by being credible and satisfying, when the player plays alone or in the company."

Sánchez et al. [44] again rephrased the definition of playability by incorporating its usability aspects [1].

"Playability represents the degree to which specified users can achieve specified goals with effectiveness, efficiency and satisfaction and fun in a playable context of use."

Kibbee outlined the essential attributes necessary for developing captivating and engaging business simulation games [23]. These qualities can be summarized as follows:

"A game must be simple to play. This does not mean that it needs to be easy to make good decisions, but the participants should not have to devote considerable time and energy to learning the rules. It requires skill and experience to abstract from the real world those elements of major importance so that a playable game will result."

According to the internet dictionary Dictioinary.com, playability Defines as :

"The quality or state of being playable: The sound and playability of vintage instruments depends on how well they are maintained. Poor graphics and counterintuitive controls negatively affected the playability of the video game." [10]

The term playability has found application in the domain of mathematical optimization and zero-sum games. Leitmann, in a seminal paper published in 1974, explores the concept of playability within the context of differential two-person zero-sum games, drawing connections to classical zero-sum games [28]. By examining the playability of such games, Leitmann provides valuable insights into their fundamental properties and the factors that influence successful gameplay.

The primary goal of every game designer is to create a kind of game that will appeal to and be fun for a large user base [12]. However, creating a game is a complex undertaking that might occasionally take years to complete [24]. A good game takes time to create; thus, game developers have to assess playability periodically. Before the players can engage in a balanced and playable game, it is challenging to evaluate the player experience. Hence, game playability can be assessed at any stage when the player is playing or interacting with the game. However, there is no standard way of analyzing the game's playability. Though several attempts have been made to propose heuristics and manual analysis processes, an efficient methodology is still missing. Specifically from the game player's perspective.

1.2 Motivation

Human values generally refer to the inherent fundamental values that bring out goodness in terms of love, truth, honesty, peace, happiness, etc. Similarly, playability values refer to the game's satisfaction, fun, effectiveness, and so on. Nowadays, many gaming applications are available in the market freely or on a subscription basis. It may be possible that a few of the games may not be playable due to certain violations (like disappointment, frustration, challenges, diversity etc.) of game playability that may distract the user.

Consider a real game example of *Apex Legends Mobile*[1] game app available in Google Play Store. Figure 1.1 shows two recent user reviews from the *Apex Legends Mobile game*, one with a five-star rating and another with a one-star rating. As we can see, the first review has a five-star rating but have negative sentiments and violates human values such as *universalism* and *achievement*. It also affects *socialism* and the *learnability* property of game playability. Though the user is happy with the app, a close look at the review also highlights the prominent concerns, such as *drop fps* (as an issue) and *recent update* (as a concern) faced by the user while playing the game.

Another part of Figure 1.1 shows another review posted by the user. The one rating of the app the user gives shows that he is unhappy with the game and has several issues. The user review analysis shows that it violates human value categories, namely *universalism*, *self-direction*, *power* and *benevolence*. Consequently, in terms of game playability, the properties such as *socialism*, *learnability*, *motivation*, *effectiveness*, and *immersion* are getting affected. We also observed that the *drop fps* and *drop frame* are the issues for which the user wants the solution.

Understanding such violations from the user reviews help the game developers to address the issues and make the game more playable for the user. If there is no reason to prioritize such values-violating app defects, they may go unsolved for a long time. Least-understood user needs can easily violate playability values and business and design decisions that increase profits at the expense of costs and software defects. These problems lead to poor app adoption, confused or dissatisfied users, harm to an organization's assets, and lost customers.

To understand the user concerns and possible violations addressed by the user in terms of user reviews, we conducted a study to analyze the user reviews of

[1]https://play.google.com/store/apps/details?id=com.ea.gp.apexlegendsmobilefps

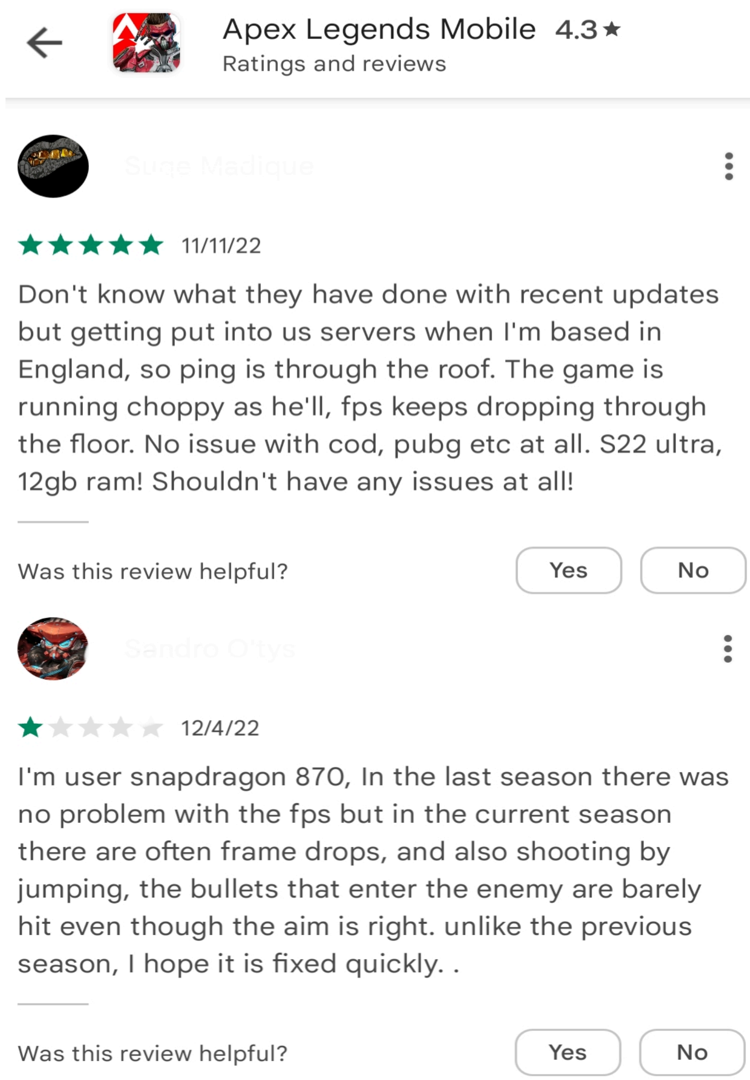

Figure 1.1: User Reviews on *Apex Legends Mobile Game App*

mobile game apps found in the Play Store. The aim is to identify the violations, functionality, issues, or concerns in mobile game apps that affect game playability and how developers respond to such violations and issues. Hence, we formulate four research questions (RQs) for the study.

- **RQ1:** What are the most and least violations in mobile game apps affecting game playability?
- **RQ2:** To what extent and how do the game developers respond to the violations?
- **RQ3:** What functionalities, issues, or concerns affect game playability?
- **RQ4:** Can we automatically detect the violations of the playability categories documented in the user/player reviews?

Figure 1.2: Mapping of Playability with Human Values

1.3 Mapping of Playability-Value and Human-Value Violations

According to Schwartz's theory, human values are classified into ten categories: *Stimulation*, *Hedonism*, *Self-direction*, *Achievement*, *Power*, *Security*, *Benevolence*, *Universalism*, *Conformity*, and *Tradition* [48]. These ten categories were then further divided into 58 values. For example, the value category of *Hedonism* includes values (or properties) such as Pleasure, Self-indulgent, and Enjoying life; *Stimulation* includes properties such as excitement in life, a varied life, and daring. Similarly, all other remaining eight values category consists of individual properties to evaluate the corresponding human value category. On the other hand, the playability model proposed by Sánche et al. [44] proposed seven different properties to assess game playability. The proposed seven categories are *Satisfaction, Learnability, Effectiveness, Immersion, Motivation, Emotion,* and *Socialisation*. In the playability model, the *"Satisfaction"* denotes enjoyment, disappointment, and attractiveness.

After closely examining the proposed human values and playability models, we found that both properties are related. As a result, we linked the playability value category to the human value categories. The playability 'Satisfaction' category (Fun, Disappointment, and Attractiveness) is closely related to Stimulation

Table 1.1: Mapping of Playability-Value and Human-Value Violations [44][48]

Playability Category	Human Value Category	Properties & Description
Satisfaction	Stimulation, Hedonism	Fun, Disappointment, Attractiveness Excitement, novelty, and challenge in life, Pleasure or sensuous gratification for oneself
Learnability	Self-Direction, Achievement	Game Knowledge, Skill, Difficulty, Frustration, Speed, Discovery, Independent thought and action - choosing, creating, exploring Personal success through demonstrating competence according to social standards
Effectiveness	Power, Security	Completion, Structuring, Social status and prestige, control or dominance over people and resources, Safety, harmony, and stability of society, of relationships, and of self
Immersion	Benevolence	Conscious Awareness, Absorption, Realism, Dexterity, Socio-Cultural Proximity, Preserving and enhancing the welfare of those with whom one is in frequent personal contact
Motivation	Self-Direction	Encouragement, Curiosity, Self-improvement, Diversity, Independent thought, and action - choosing, creating, exploring
Emotion	Stimulation	Reaction, Conduct, Sensory Appeal, Excitement, novelty, and challenge in life
Socialisation	Universalism	Social Perception, Group Awareness, Personal Implication, Sharing, Communication, Interaction, Understanding, appreciation, tolerance, and protection for the welfare of all people and nature

and Hedonism. The goal of Stimulation values is excitement, novelty, and challenge in life, which are related to fun. Hedonism is pleasure, enjoying life, and self-indulgence related to Attractiveness and disappointment in the satisfaction category. We have now associated satisfaction with Stimulation and Hedonism. Similarly, we have tried to map all human value categories to the playability value categories. Though this mapping was done intuitively (after understanding each of the values from both the models) by me individually and later validated by the my supervisor for conformance and disagreements, we may still have missed a few while mapping.

We also found that the categories of human values are not only related to one of the playability categories; there are overlaps between them. We can, for example, associate stimulation with emotion (i.e., reaction, behavior, sensory appeal). Stimulation denotes excitement; if there is excitement, there will be reaction outbursts. Figure 1.2 shows the mapping of playability with human value categories. We found that the two human value categories, *Conformity* and *Tradition*, are unrelated to any of the playability values and are not included in the mapping list. Table 1.1 shows a detailed view of the values (or properties) related to playability and human values.

1.4 Thesis Contribution

In this thesis work, we aim to analyze the playability of the games from the user's perspective by analyzing the user reviews (positive or negative) publicly available in the app stores. We have combined Schwartz's human values model [47] with Sánche's playability model [44] to understand the user issues, concerns, and feedback with the mobile game apps. Additionally, we identified a set of violations the game has based on the user reviews available on the Google Play Store. Furthermore, we calculated the fine-grained features of the reviews showing the violations and used these features to determine the functionalities, issues, and concerns affecting the game's playability. Our analysis of user reviews showed that the mobile game app violated human values and subsequently affected game playability. At last, we develop a machine learning approach to classify a user/player automatically and review whether a review violates a specific playability category.

1.5 Organisation of the Thesis

The organisation of the thesis is as follows:

- Chapter 2 discusses the existing literature, for example, Mining app reviews, Human values in Software Engineering and the Playability of games related to our research work.
- Chapter 3 briefly describes the approach we used for our research work.
- Chapter 4 gives an overview of the results we achieved from our work. At last, we also discuss the Limitations of the work and how we mitigated them.
- Chapter 5 discusses the automatic detection of playability based on user reviews. First, we discuss existing literature regarding playability and machine learning. Then we discuss our Proposed approach, related experiments and results, and a web tool.
- Chapter 6 concludes the thesis and future work.

CHAPTER 2

Literature Review

2.1 Mining App Reviews

App reviews are an excellent resource for understanding user input and providing app developers with this data [6]. Guzman et al. [15] investigated app reviews and culled fine-grained features that programmers found helpful in requirements evolution activities. A related study used Latent Dirichlet Allocation (LDA) techniques and specified linguistic rules to identify and retrieve feature requests from app evaluations, demonstrating, among other things, that customers frequently ask for greater assistance and more regular updates to mobile apps [18]. Sorbo et al. [9] presented SURF, a technique for condensing many evaluations into coherent summaries and recommending informative software improvements, to reduce the effort necessary in analyzing app reviews.

Vu et al. [41] proposed MARK, a keyword-based tool for detecting trends and changes that relate to occurrences of severe issues in reviews. Panichella et al. [39] proposed a taxonomy for classifying reviews and introduced a hybrid approach of NLP, text analysis, and sentiment analysis to classify app reviews. Li et al. [29] propose an approach for analyzing the playability of video games based on reviews. To understand the overall playability of a particular video game, they have provided an effective solution by using the collective opinions of a large number of players. Lin et al. [30] conducted an empirical study on the reviews of 6224 games. They have analyzed the quantity and complexity of reviews, the type of information they contain, and how many hours the user played the game before posting a review.

Obie et al. [35] analyzed app reviews using a dictionary-based approach and identified the most violated and least violated categories of human values in the app reviews. The authors analyzed and investigated human values using the Schwartz theory [47][48]. The authors have also presented an approach for automated identification of honesty human value violations from an end-user perspective [36]. Obie et al. [25] has also conducted a study on Stack Overflow data and identified potential human value violations. Qiu et al. [43] presented a new large-scale dataset on human values called VALUENET. It contains human attitudes in various text scenarios.

2.2 Human Values in SE

Human values are the foundation for what people believe to be significant in life [7]. Although these concepts are frequently not expressed using formal terminology, technologists and non-technical individuals rely on them when making decisions. Thus, from the selection of end-user apps to the technical design choices made by developers in software engineering projects [59], the influence of human values may be seen in people's preferences [37]. According to the fundamental theory of values, values serve as a road map for behaviour and a means of expressing needs [14]. Although practice and research in software engineering have incorporated well-known values like privacy, security, and accessibility, little focus is placed on broader human values like conformity and self-direction in software engineering, particularly in the development of mobile apps [40].

The Schwartz theory of fundamental human values is the theory of human values [47][48]. The theory divided 58 human values into ten categories. These ten categories are *Stimulation, Hedonism, Self-Direction, Achievement, Power, Security, Benevolence, Universalism, Conformity*, and *Tradition*. Hedonism's value category comprises Pleasure, Self-indulgent, and Enjoying life. Beyond the social sciences, the Schwartz theory is widely applied in computer science and software engineering research.

Recent research on human values in SE highlights the need for software companies to explicitly address concerns about human values in their software development processes. The resulting software artefacts impact end users and society at large both directly and indirectly [57]. Shams et al. [49] used Schwartz's model to manually analyze app reviews to understand better the desired and missing human values in existing Bangladeshi agriculture apps. Shams et al. [50] also applied the portrait values questionnaire (PVQ) to 193 Bangladeshi female farmers as the end-users of agriculture mobile apps to evaluate missing and present values. According to their analysis, the two most crucial value categories for Bangladeshi female farmers are security and conformity. The study conducted by Friedman et al. [11] shows how technical tools and ethics, morals, and values are related. He calls for a values-sensitive design, a moral way for technology to consider values during the design process.

2.3 Playability of games

Playability is an essential term for games. Playability is a group of characteristics that describe how a user or player feels while playing a specific gaming system or video game with the primary goals of providing enjoyment, amusement, and educational tactics, for example. The researchers use playability to determine the quality and usability of video games. Sánchez et al. [44] provides a framework for studying and assessing video games' user experiences. Recently, Paavilainen [38] suggested a playability definition based on a game's gameplay, usability, and reliability. The author also discusses the relationship between playability and player experience.

Despite the video game industry's rapid growth, there are still few approaches for assessing game quality and player experience. Usability evaluation is a prevalent task for game playability. Many studies have mapped usability to the heuristic evaluation. It is a non-formal analytical technique in which numerous evaluators are asked to provide feedback on the target design in accordance with pre-established guidelines/heuristics/principles. Desurvire et al. [8] present Heuristic Evaluation for Playability (HEP), a complete collection of playability heuristics developed using play-testing heuristics and productivity literature to assess video, computer, and board games. Pinelle et al. [42] proposes a new set of heuristics that can be applied to video game usability tests. The heuristics made identifying usability issues in early and working game prototypes easier.

Many researchers have worked on researching the creation of video games. Ampatzoglou et al. [2] performed a systematic literature review to discover the research activity in software engineering for computer games that have increased during the past few years. Tschang [53] utilizes a qualitative method to develop the grounded theory for creating video games at many levels of analysis, such as the industry level, the organizational level, and the level of the individual creative. Tschang et al. [54] outline how people's creative efforts can help in video game creation. The authors have also suggested that identifying the products' historical origins (like elements from earlier games and other media or products) would be a constructivist approach to game design.

Burger-Helmchen et al. [5] explored the relationship between video game development firms and player communities. The authors have also analyzed that the engagement between businesses and user groups has significantly improved video game quality and the involvement of users in the video game business help in developing quality video games. Kultima et al. [27] analyzed the results of an interview (conducted in the year 2009) study that gathered information on three significant conferences for the video gaming industry. The study shows increased instrumentalist viewpoints alongside more personal, artistic views within the gaming industry.In the field of game usability, Nacke has presented an innovative hierarchical model that encompasses both abstract and concrete components [34]. This model suggests effective evaluation methods. Mello and Perani delve into a thought-provoking discussion by exploring the similarities and distinctions between gameplay and playability [33]. By examining earlier definitions, they shed light on the interconnectedness and unique aspects of these two concepts. In a comprehensive study by Wiemeyer and colleagues, the focus shifts towards the player experience, which is analyzed across three interconnected levels: the (socio-)psychological, behavioral, and physiological realms [58].

Jan Kruse and Ricardo Sosa introduce an innovative methodology for generating urban maps in First Person Shooter (FPS) games through procedural techniques. Their approach utilizes a multi-agent evolutionary system within the Unity3D game engine to strategically place streets, buildings, and various other elements, culminating in the creation of fully playable video game levels [26]. Philip Bontrager and Julian Togelius present an innovative concept called Generative Playing Networks (GPN), where game levels are autonomously designed for the purpose of self-play [4]. The GPN algorithm consists of two fundamental components: an agent that acquires the skill to play game levels, and a generator that comprehends the distribution of playable levels.

Though various research and studies have been conducted to understand the development of games with quality, the verification of games in terms of usability, effectiveness perspective and development of heuristics to evaluate the playability of games, still evaluating the playability of the games from the user's perspective is missing. In this work, we aim to analyse the playability of the games from the user's perspective by analysing the user reviews (positive or negative) posed by them while playing the games. We have combined Schwartz's human values model with Sánche's playability model to understand the user issues, concerns and feedback with the mobile game apps.

Chapter 3
Methodology

The presented approach aims to automatically detect human value violations manifested in game app reviews. After that, these violations will be associated with specific game features to provide a better understanding of the violations of human values affecting the playability of the game. Figure 3.1 provides an overview of the presented approach. Initially, a game app review corpus is created by mining reviews of different games from the Google play store. After that, text processing and sentiment analysis are performed to determine the positive, negative, and neutral reviews. Only the negative and the neutral reviews are considered further for human value violation analysis. A human value dictionary is used to detect violations of human values in the review text. Once the human violations are detected, different fine-grained features are extracted from the review text. These fine-grained features are mapped to the detected human value violations. Finally, the functionalities, issues, and concerns affecting the game's playability are determined, and this information is supplied to the developer for the necessary actions. The details of each step are provided in the following sections.

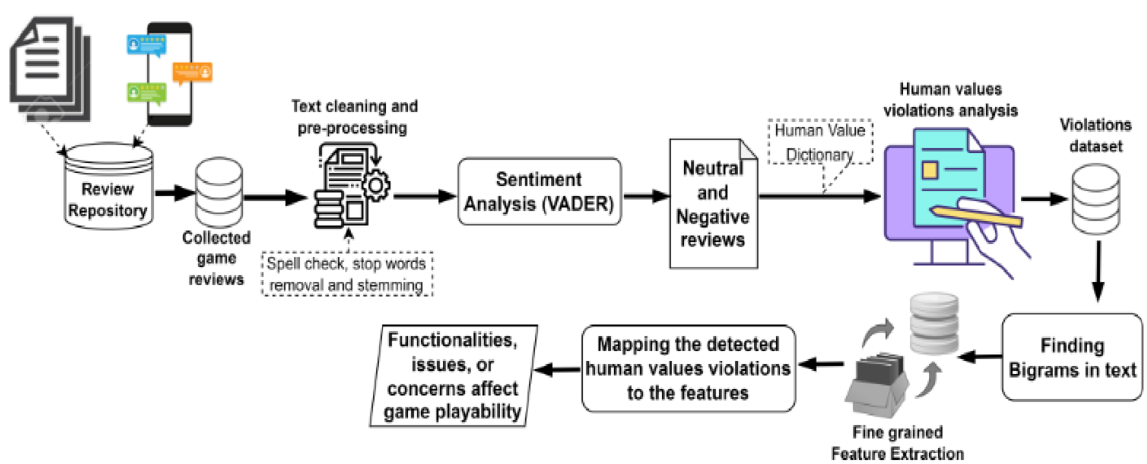

Figure 3.1: Overview of the proposed approach

Table 3.1: Overview of the Dataset

Game Name	#Reviews
Apex Legends Mobile	1700
Call of Duty Mobile	1700
Genshin Impact	1700
GRID Autosport	1700
League of Legends: Wild Rift	1700
Minecraft	1700
MONOPOLY - Classic Board Game	1700
Among Us	1700
Pokémon GO	1700
Sky: Children of the Light	1700
Stardew Valley	1700
The Room: Old Sins	1358
Angry Birds 2	1700
Mini Militia - Doodle Army 2	1700
Candy Crush Saga	1700
Total reviews	**25158**
Reviews after cleaning and pre-processing	**20,346**

3.1 Game review data collection

Initially, we scrutinize the internet and determine famous games which are highly played by the user nowadays. After that, we manually checked in the google play store, verified them, and narrowed it down to 15 top games. Table 3.1 lists the selected games. Next, we fetch the user reviews of these selected games. For this, a python scraper called *google-play-scraper*[1] is used. The information, such as the most recent and most relevant reviews, thumbs-up counts, date, reply content, and star rating, is collected for each game. A total number of 25,158 reviews was collected. We collected the most recent 1700 reviews from each game. However, only 1358 reviews were available for the Room game. After discarding reviews with less than three tokens, duplicate values, and non-informative, 20,346 reviews are used for further analysis.

3.2 Data cleaning and pre-Processing

The collected review data consists of review text, star rating, and the number of likes. This data is further cleaned and pre-processed for the NLP tasks. The following steps are applied for the data pre-processing.

[1]https://github.com/JoMingyu/google-play-scraper

1. **Lowering a case and punctuation removal**: The input review text is first converted to lowercase. Then all punctuations are removed from the textual data. The punctuation removal process will help to treat each text equally. The Python NLTK library[2] is used to lower a case and remove punctuation.

2. **Misspelt Words** Since most people use smartphones to write reviews, it is common to make spelling and typographical mistakes. Therefore, an autocorrect library[3] with a spell checker function is used to correct all the misspelled words in the textual data.

3. **Stopwords Removal and Stemming** To remove common English stopwords, e.g., this, is, a, etc., again NLTK library functions are used. These stop words do not provide useful information and desired human values violations. Therefore, we removed them before proceeding further. Stemming is an NLP process in which the word is reduced to its base form. For this, we use the snowball stemmer[4]. It is a porter stemmer technique that improves a searching capability.

3.3 Sentiment analysis of user reviews

Sentiment analysis is a process of finding the emotional significance of a particular natural language statement. The presented work takes user review text as the input and assigns it one of the three sentiment labels: positive, neutral, and negative [3]. The same process is performed for all the user reviews. For our purposes, we use the VADER sentiment analysis model [17], specifically sensitive to web-based media assumptions. This model performs emotion analysis using rules and a lexicon, called VADER (Valence Aware Dictionary and sEntiment Reasoner). The VADER model uses various lexical highlights, often identified by their semantic direction as either positive or negative. Therefore, VADER informs us of the polarity score and the positive or negative result. By adding the valence scores of each word in the lexicon, adjusting for grammatical and syntactic rules, and then normalizing to fall between -1 (most negative) and +1 (the most positive), the VADER model creates a normalized weighted compound score. After this step, each user review is assigned a positive, negative, or neutral label.

[2] https://www.nltk.org/
[3] https://github.com/filyp/autocorrect
[4] https://snowballstem.org/

3.4 Human value violation detection

We assume only negative and neutral reviews can be the candidates for human value violations. Therefore, the reviews labelled with positive sentiments are discharged, and only reviews with negative and neutral labels are considered for further analysis. We have applied a dictionary-based approach for human-values violation detection similar to the one proposed by Obie et al. [35]. According to the Schwartz theory, human values are categorized into ten main categories. These ten categories were then divided into 58 values (like excitement in life, a varied life, and daring). We have identified synonyms and antonyms for each human value category. The dictionary consists of human values terminology, their corresponding synonyms, and antonyms. We have also stemmed the value dictionary entries using the snowball stemmer, similar to the steps described in section 3.2. This well-curated dictionary creates an NLP classifier to detect human values violation in user reviews.

3.4.1 Truthset creation

We create a truthset to validate the presented approach. This process involved systematic evaluation of reviews and human coders who read each review and evaluated its contents following a rigid coding guide. We randomly picked 1000 reviews from the dataset to create a truthset. We first understand the value terminologies of each of Schwartz's value theories. Subsequently, we analyzed each review, discussed it, and checked whether it exhibited violations. We have resolved any conflict through discussion.

3.4.2 Automated Values-Violation Detection

Similar to the work reported by Obie et al. [35], we created an NLP model that automatically finds value violations in user reviews. The model uses the data from the value dictionary and sentiment analysis from the preceding steps to allocate review text to one or more Schwartz value items. We estimate the probability that a review will find one or more values specified in the value dictionary to be violated. Formally, the NLP model determines the probability that a review R consists of the value V as $\mathrm{P}_{(R,V)}$=$\mathrm{T}_V/\mathrm{T}_R$, where T_V is the number of tokens in R that exist in the *V-related* values dictionary established above and T_V is the total number of tokens in R. A value violation is assigned if $\mathrm{P}_{(R,V)}$>=0.05 and the sentiment analysis output is neutral or negative. We assessed the effectiveness of the presented approach by comparing the values-violations tags supplied by the NLP model to the truth set generated in the earlier step. We used precision, recall, and F-measure measures to evaluate the presented approach. The results showed that the presented approach achieved an F1-score of 0.80, a recall of 0.80, and a precision of 0.82 in detecting values-violation in user reviews.

3.5 Fine-grained feature extraction

After detecting the human value violations in the reviews, the next step is knowing what kinds of game app issues, functionality, and concerns are affected by these violations. For this analysis, we separated all the reviews found violating human values and created a separate dataset called the "violations dataset". We have pre-processed the violations dataset to extract the nouns, verbs, and adjectives by identifying POS tags using the NLTK toolkit[5], stopwords removal. In the stopwords, we have added a few words in our dataset, such as game, play, fix, problem, please, app, and game app names (e.g., call duty, candy crush, apex etc.), as these words don't represent our features. We have applied the lemmatisation technique using Wordnet lemmatizer[6] from NLTK.

Next, for identifying these things in violation reviews, we use the collocation finding algorithm provided by the NLTK toolkit[7]. A collocation is a group of words that come together unnaturally frequently. However, groups of words are not necessarily consecutive in the text. We use the same approach used in [15] for the fined-grained feature extraction. We find the unique set of words in the whole violation dataset in terms of bi-grams. This produces the list of fine-grained features, each consisting of two keywords with their frequencies and sentiment scores. Finally, we use the technique in each review to look for bi-grams and determine which user review found the bi-grams.

After identifying bi-grams, we map the human value violations to the features. The mapping was done by me individually by randomly selecting the user reviews. We then resolved the disagreements in consultation with my supervisor. Once the human value violations are mapped with features, we identify the corresponding violations of the game playability categories.

[5]https://www.nltk.org/
[6]https://www.nltk.org/howto/wordnet.html
[7]https://www.nltk.org/howto/collocations.html

Chapter 4

Analysis Results

The methodology discussed in chapter 3 has been used to analyze user/Player reviews and to detect violations in them. In this chapter, we present our analysis results and related findings.

4.1 RQ1: What are the most and least violations in mobile game apps affecting game playability?

In this study, after cleaning and pre-processing, we considered 20,346 user reviews from 15 game apps to check which property of the game's playability is most and least affected (as shown in Table 4.1).

Table 4.1: Total number of review violations across 15 gaming apps

Game Name	**# Reviews**
Apex Legends Mobile	514
Call of Duty Mobile	539
Genshin Impact	397
GRID Autosport	401
League of Legends: Wild Rift	549
Minecraft	473
MONOPOLY - Classic Board Game	323
Among Us	413
Pokémon GO	512
Sky: Children of the Light	340
Stardew Valley	376
The Room: Old Sins	151
Angry Birds 2	459
Mini Militia - Doodle Army 2	410
Candy Crush Saga	373
Total	**6230**

We found that 30% of the 20,346 reviews, or 6,230, contained human rights violations. The highest number of violations has been found in *League of the legends:wild Rift* gaming app, and the lowest number of violations have been found in *The Room:old Sins* app. Table 4.2 shows the categories of the value violations and their value items. Table 4.2 also shows the example reviewers highlighting how many reviews have revealed value violations in each gaming app. Table 4.3 reports the total and average violations for each human value category in each gaming app. From the Table, we found that *Universalism* (22%), the most violated human value category, denotes *Socialism* in the playability category. *Stimulation* (7.4%) is the least violated category that directly affects *Emotion* property. Subsequently, Security (15.4%), Hedonism (13.78%), Achievement (11.71%), Benevolence (10.31%), Power (10.14%), Self-Direction (9.6%) are the violated categories. We then mapped these human value violation categories to the game playability values (using Table 1.1 and Figure 1.2). Violation of *Universalism* values directly affects the *Socialism* of the game's playability. Similarly, *Hedonism* violation affected the *Satisfaction* of the game's playability. Similar mappings have been performed for other game's playability categories.

Table 4.2: Categories, Value Items, & Example Reviews (*f= frequency)

Category	Individual Value	f	Example review
Self-Direction	Freedom	2287	When I'm able to play the game it's great but it crashes on me way to often. It literally crashed and closed out 6 times in a row while I was in a middle of a match. I have on the lowest graphics possible and it still does it. I would really to give this more stars but I can't because it consistently keeps crashing. Update: I've still been trying to play but its still very unstable. Crashes a lot.
	Creativity	110	Originality I rated five stars but it's boring now the new players they add are bad
	Independence	185	Invalid Matchmaking. If you are playing with randoms, your teammates will be of low rank whereas enemies are of high rank. There are lots of bugs glitches which makes game unplayable. Report system doesn't work. If you are unable to unlock legend at given time, you cant have it for free. There is no cooldown on commands, so players after getting knocked or taking banner, keeps on continuously spamming. This is really irritating but unfortunately developers don't care about all these things...
	Privacy	298	No privacy even after I deleted the game hoyoverse is still accessing my Google account. I'm going to complain to Google.
	Choosing own goals	105	There's no option to turn it off or change it or more than 1 occasion I have had no vision of the enemy trying to even execute my team member. It's very annoying should be fixed. Otherwise great game

Continuation of table 4.2			
	Curiosity	106	The game is really nice but the graphics is a bit poor. It'll really be nice if you can make the game 3D. The first time I saw the game, about 3 years ago, I wasn't really interested because the graphics wasn't inviting. But ever since I started I've not stopped, it's been the only game on my phone. So please work more on the graphics and make it really inviting
	Self-respect	523	I can't even play without the game constantly crashing every 2 minutes. 0/10, worthless, trash game. FIX THIS DEVS!!! SERIOUSLY! It's been a problem for MONTHS and you haven't touched it since the game dropped. Fix this issue and let give it a better rating.
Stimulation	Excitement in life	1564	Are there even real players? Even your team mates are bots! If you jump up to any weird area they don't know how to act and stay still. Then snap to once you jump down to solid ground. You can NOT lose. No matter how hard you try to get killed. Boring. Challengeless.
	A varied life	1003	Good so far. But the constant lags are bothersome and borderline game breaking. I could play Call of Duty Mobile on maxed out graphics and everything and had little trouble aside from the rare frame spikes when multiple explosives go off in the same spot. Doesn't matter how I change the settings on this game I'm never gonna get good ping. Only usable ping.
	Daring	251	Challenges don't track correctly, just finished a game with 7 kills without being knocked and my 5 knocks Challenge is not finished. Waste of my time.

Continuation of table 4.2

Hedonism	Pleasure	1643	Why graphics is so dark see while play ranked and battle royal match. It is very hard see properly on enemy and there is a problem like glitch happening every single second on match. Please fix this problem like lag, high frame drop and why my character legend is slow like while firing on enemy and moment.
	Self-indulgent	189	Really bad experience... doesn't even have a good server and lots of glitches...very depressing game....it only gets worse as you progress....not a good game to recommend
	Enjoying life	3355	The lagging in this game after every new season is getting worse, especially in BR. I don't mind losing to a better player but to lose because it lags out and puts me back 10 seconds to where they can leisurely kill me, pisses me off! The enjoyment of this game is starting to waiver
Achievement	Ambitious	1486	Changing my rating to 1 star. Nothing in this game works, my characters stay basic, none of the equiped character perks work none of the rank protection items work, game constantly crashes, this phone has 12 gigs of ram and 265 gigs memory, no reason for the crashes but the worst is absolutely no developer support. Contacted them several times, its a total waste of time and effort, no response whatsoever. I'll try a couple of more games but if it doesn't improve I'm uninstalling it.
	Influential	1039	Ranked message is in the way of chat, I have no plans of playing it and all my games after a certain point have been with controller players and it's just painful to play now edit. This is actually so stupid why am I versing high level people when I'm level 13 where is the balance there is not a single low level

Continuation of table 4.2

	Capable	547	Can't even install on my phone. I was able to purchase it, but after downloading it, I get notification, that it can't be installed. I use one plus 8pro
	Successful	425	The game was very enjoyable at first but it kept saying unsuccessful when I try to link my account, I have lost my data four times because of this and the game keeps lagging, I can't even finish a game and it even goes to the extent that it even takes me back to my homescreen when it laggs too much.
	Intelligent	912	This is stupid I accidentally uninstalled it and now I can't reinstall it. It's dumb
Power	Wealth	1709	Game theory is good. Just such a pity about the constant freezing and lagging. Please stop worrying about silly emblems and colours and stupid outfits and focus on getting the platform of the game correct first. Also I am constantly shooting an opponent and he fires a couple of shots and I'm down. That's with gold armour!! Maybe make it more fair!!!!
	Authority	1377	Since last update the game doesn't load my profile and shows me a message saying TIME OUT. Having enough of all this bollocking problems every time you guys force us to update the game.
	Preserving my public image	415	There is a glitch in the system in the game, it does not count the amount of evolution I do through the character (fade) It does not count the number of kills or the amount of damage .. It stopped at the number of 400 the number of kills, And the amount of damage 96114, I played after that 20 times and none of that counted.

Continuation of table 4.2

	Recognition	317	It says player can't accept invitations until they have completed rookie tutorials. But I know the player and he has completed rookie tutorials. PS the player did skip the rookie tutorials the first time, but completed it later on.
Security	National security	1130	Game is good as old ... But hacking players are making the real gamers getting uninterested on this game... You guys should get something out of head to ban this hackers... Let the real gamers experience the war
	Family security	476	My husband and I are writing to Activision to say that the new update sucks and is horrible because for one My husband couldn't find his account on the game and we tried everything else to find and didn't. So we put this New season 10 at 0 no stars. Season 8 and 9 are way better than this.
	Sense of belonging	1285	I love this game but the only thing I hate about this is the new version. Everything is locked and wield two guns you need to go through a lot. Before this was not the case and you could personalise your avatar according to your choice in the previous version.
	Social order	214	15 minutes to find a single ranked match.... Something doesn't seem right there
	Healthy	1267	After the lastest update the mic and sound between team-mates is not working.. there is severe problem in mic and sound.. cannot even activate and deactivate mic.. it used to work perfectly but now absolutely useless button for mic and sound...

Continuation of table 4.2

	Clean	1424	This game is so good. But the fault is in upgrading the weapons, bcz. it requires much money to upgrade .wheareas the earnings is to less and slow with respect to every match.So, it should be change ,all the weapons which get unlocked in crate should be totally upgraded from starting.
Benevolence	Helpful	1182	Already issues with the new update. Reset my zombie points when coming back to the game. Was at 28%, now at 2%. Same goes for all the other progress. This is a new game? I would save my money. Also, no way of telling tech support.
	Responsible	595	Players DO NOT, I repeat DO NOT, have control over whether they are on 4G or 5G when playing on a mobile device, therefore it is NOT OUR CONNECTION that is unstable. UPGRADE YOUR SERVERS TO A MORE STABLE NETWORK!!! STOP PLAYER BLAMING FOR UNSTABLE NETWORKS AND START TAKING RESPONSIBILITY!!! THIS IS COMPLETELY UNACCEPTABLE!!!
	Forgiving	153	This is easily the worst version of Apex. The limitations of mobile drag it down. Everything feels bad, and you're constantly fighting touch controls. If rather just play on my Switch. If controller support was ever added - I think I'd give this another shot. I understand why it hasn't been.

Continuation of table 4.2

	Honest	788	After playing Apex Legends Mobile for a while now, I can say it's a very nice game like others that I've seen, but I have some complaints. First of all, the matchmaking for BR or any game mode takes some time to get going and usually there's mostly bots as teammates and there are never real players to be honest, usually 1 - 10 games, you'll get atleast 1 - 2 real teammates. Second, the amount of hackers make it very unfair to players who actually has skills that are ranking up.
	Loyal	459	I'm going to record my game and sue call of duty for I'll only hackers in a game to still identity and data without any consequences reporting people is pointless
	A spiritual life	0	
	True friendship	704	1 star from me, for microtransactions, pay to win and loads of windows I need to close every time I start a game. No matter gameplay, sounds, music and graphic. Microtransactions and loads of pop-up windows grant this game 1 star. What I could've expect? It's a game from two corpos, that are among the greediest out there. Oh. And also cheaters.
Universalism	Equality	1881	Are there even real players? Even your team mates are bots! If you jump up to any weird area they don't know how to act and stay still. Then snap to once you jump down to solid ground. You can NOT lose. No matter how hard you try to get killed.
	Wisdom	955	Lag is killing this game. The lag is so bad you freeze and when it unfreezes you are dead. By the time you spawn into a match the match is half over. Or you respawn and die instantly.

Continuation of table 4.2

	Inner harmony	729	Matchmaking system and optimatisation are the two huge problems with apex legends mobile. I don’t understand whats the point of keep updating game and nothings improving.if this game was made just for ipad you should change the name into "apex legends ipad". I mean the game is already dead from season 2.
	A world of beauty	1893	This game is amazing my third battle royal game is it. Its graphics also good and play on G70 processor with no lag and without heating this is really good. I got 30fps on Helio G70 this is bad. Plaease give 30fps. This last update make the graphics very bad and dull.black colour is now more dark.kindly fix this.
	Social justice	189	Is more interesting when when you have boosters to play with, and this also comes when you watch some adverts, but lately I don’t get those adverts again. What went wrong? No attention given to my request to watch adverts and get extra boosters. Still no attention... the adverts have started coming. Thanks
	Broadminded	1483	Great game. Would give it 5 stars, but after I got the monthly subscription, it stopped letting me log in with google play. This happened twice. The first time I lost all my game data and lots of wasted money and the developers won’t get back to me. What a joke.
	A world at peace	995	Game lags a lot, too much frame drop specially in battle royale. Even when ping is 94ms the game lags. Very difficult to play as the character just freezes and you lose because of that. There should be an in game report system about lags and glitches. Edit- still no improvement game is TRASH

Table 4.3: Number of Violations in 15 Game apps for each of the human-value violation categories

Mobile Game App	Self-Direction	Stimulation	Hedonism	Achievement	Power	Security	Benevolence	Universalism	#Total
Apex Legends Mobile	285	252	412	312	293	529	344	683	3110
Call of Duty Mobile	297	226	451	339	329	611	364	835	3452
Genshin Impact	242	206	357	292	310	372	253	550	2582
GRID Autosport	168	146	330	371	327	327	199	383	2251
League of Legends: Wild Rift	318	259	405	314	300	571	314	762	3243
Minecraft	217	199	398	355	294	395	309	560	2727
MONOPOLY Game	190	127	255	203	243	334	162	374	1888
Among Us	269	195	343	300	213	348	274	515	2457
Pokémon GO	299	253	424	328	312	470	392	773	3251
Sky: Children of the Light	229	166	295	256	191	247	213	447	2044
Stardew Valley	219	158	335	334	284	340	245	450	2365
The Room: Old Sins	72	57	101	100	76	115	77	169	767
Angry Birds 2	289	235	391	343	232	386	260	574	2710
Mini Militia - Doodle Army 2	295	193	342	206	226	450	236	579	2527
Candy Crush Saga	225	146	348	356	188	301	239	471	2274
Average	**240.9**	**187.8**	**345.8**	**293.9**	**254.5**	**386.4**	**258.7**	**541.6**	**2509.8**

4.2 RQ2: To what extent and how do the game developers respond to the violations?

App stores have a provision that allows app developers to answer the user's questions/reviews and issues with the app. We analyzed the game app developer's response to the user reviews in this RQ. The results of our analysis show that out of 6,230 violations of the values of user reviews, only 1,142 (18%) reviews have received a response from the game app developers. We have analyzed all 1,142 developer responses and found that, except for 4, the remaining 1,138 responses addressed the violations in the reviews. Further, the developers incorporated and solved these violation issues in their app to satisfy the value category of game playability. Table 4.4 shows the distribution of developer responses across all 15 apps (four app developers have not responded to any of the user reviews).

4.3 RQ3: What functionalities, issues, or concerns affect game playability?

The fine-grained feature extraction method discussed in Section 3.5 has been applied to extract features from user app reviews. This analysis has been performed on all 6,230 violated reviews in the "violation dataset". After identifying bi-grams, we found that all the reviews have not resulted in the extraction of the features; out of 6,230 reviews, only 2,225 reviews resulted in features and issues extraction.

Table 4.4: Developer responses against the violations across Gaming Apps

Game Name	#Reviews
Apex Legends Mobile	196
Call of Duty Mobile	0
Genshin Impact	175
GRID Autosport	169
League of Legends: Wild Rift	127
Minecraft	0
MONOPOLY - Classic Board Game	190
Among Us	181
Pokémon GO	0
Sky: Children of the Light	63
Stardew Valley	5
The Room: Old Sins	21
Angry Birds 2	9
Mini Militia - Doodle Army 2	6
Candy Crush Saga	0
Total	**1142**

We have undergone the details of each 2,225 features. The analysis of 2,225 bi-grams was done by me individually. The disagreements on identifying features from 2,225 bi-grams were resolved by consultation with my supervisor. As a result, we found 42 features representing the gaming apps' issues, concerns, and functionalities after removing those features that don't make any sense. From the identified 42 features, we randomly selected the corresponding reviews on violations and mapped them with the playability values. We found the features with their corresponding playability value for the set of violation reviews.

As we can see in Table 4.5 It shows a few example app issues and concerns related to playability violations as reported in the game app reviews. Here, the count column denotes the count of the functionalities, issues, and concerns found out of 6,230 violated game app user reviews. The playability column denotes the subset of playability value violations found in the user review concerning the features, issues, and functionality. Our analysis shows how specific issues, concerns, or functionality are associated with playability violations and human value violations. For example, the battle royal issue affects the Immersion and Socialism property of the playability. Time waste, the last update, hide-and-seek, the old version, keeping crashes, and the daily challenge are associated with learnability violations. Learnability is affected mainly by frequent updates, which may result in issues like app crashes and time wastage and affect game playability. As seen from Table 4.5, the new update, battle royal, time waste, and money waste are the most affected issues and concerns, as documented in the game app violation reviews.

4.4 Limitations and Threats to Validity

In this section, we present our threats to validity considerations.

Construct Validity: The selection of the apps from the app store and user reviews from the apps is one of the threats. We have identified 15 well-known mobile games and fetched user reviews from these games. We fetched the most recent and most relevant reviews across all the star ratings in the game apps. Another threat may be related to creating a truth set for analysis. We (not the actual game developers) have created the truth set by randomly selecting user reviews. I created and labelled the dataset for the truth set, and my supervisor validated them. Although we labelled the dataset by handling all disagreements through discussion between us and calibrating our decisions to refine a common protocol, it is still a subjective process.

Internal Validity: Several threats may affect our analysis results. One of the threats is the dictionary used for identifying the violations. We have used the process followed by Obie et al. [35] to create the dictionary for detecting violations. We have also conducted a pilot study on 150 reviews; We randomly chooses 50 reviews for labelling. The pilot study aims to identify and develop a common framework for mapping the violations with Schwartz value categories. After the pilot study and framing out a common framework, we developed the truth set data. Another major threat is the error in detecting the violations. Based on the analysis of truth set data, we developed the value detector to detect the violations. As the accuracy is close to 81% on the truth set data, it may be possible that the NLP detector may miss a few of the violations on the user reviews. Also, in case of reviews not in the English vocabulary, the detector will probably miss them.

External Validity: The generalisability of the results is a threat here. In this study, we have chosen 15 mobile game apps and the user reviews available on the Google Play Store for analysis. The chosen apps are famous among distinct age groups and communities. We believe these mobile apps may be representative of gaming apps; hence, results may also be applicable in a larger context.

Table 4.5: Set of example app functionalities/issues/concerns and their affect on related playability category

Functionality/Issue/Concern	#Count	Playability Category
fortune tower	14	Motivation, Learnability
time waste	140	Learnability
loading screen	39	Motivation
battle pas	36	Emotion,Satisfaction
dice roll	24	Satisfaction
battle league	41	Immersion,Socialism
quick chat	20	Effectiveness
drop fps	36	Immersion
cut scene	12	Emotion
battle royal	142	Motivation, Satisfaction
last update	105	Learnability
recent update	52	Immersion,Socialism
hide seek	20	Learnability
match ranked	41	Satisfaction,Immersion
stardew valley	13	Socialism
get stick	115	Motivation
match making	29	Emotion,Socialism
get rid	34	Motivation,Emotion
controller support	37	Effectiveness
black screen	94	Motivation
old version	60	Learnability
internet connection	56	Socialism,Motivation
ad watch	39	Effectiveness
reinstall uninstall	52	Immersion,Effectiveness
new update	283	Motivation,Satisfaction
keep crash	88	Learnability,Effectiveness
frame rate	23	Immersion,Socialism
bar gold	18	Emotion
money waste	100	Effectiveness
pc version	49	Motivation
daily challenge	16	Learnability,Satisfaction
money spend	134	Effectiveness
control touch	40	Emotion,Socialism
rift wild	30	Motivation

CHAPTER 5
Automatic Detection of Playability Category

The proposed methodology discussed in Chapter 3 is semi-automated, where human interventions are needed to analyze the results. This chapter presents our automated approach for automatically detecting the playability category, which minimizes human interventions and related bias.

5.1 Related work

Many researchers are actively exploring the concept of game playability through diverse perspectives, utilizing player reviews in conjunction with machine learning techniques. By harnessing the power of these combined approaches, researchers aim to gain deeper insights into the factors that contribute to a game's overall playability. Vasa et al. undertook a preliminary analysis involving a vast dataset comprising 8.7 million reviews of 17,330 mobile apps. Employing statistical methods, they explored the relationship between the character counts of user reviews and their corresponding ratings [55]. Their findings shed light on several intriguing patterns, highlighting the tendency for mobile app reviews to be relatively concise. Furthermore, the researchers observed that both the rating and the category of an app exerted influence over the length of a review, revealing an interesting interplay between these factors. Expanding on the same dataset, Hoon et al. delved into the realm of user sentiment expressed within reviews. Through their investigation, they discovered that the most frequently utilized words in these reviews were often indicative of the underlying sentiment [56]. This insightful analysis underscored the significant role played by sentiment in shaping user feedback.

Harman et al. employed sophisticated algorithms to extract app features and conducted correlation analyses on a sample of 32,108 non-zero-priced apps from the Blackberry app store. By employing these customized algorithms, the researchers were able to uncover meaningful insights regarding the features associated with these apps, facilitating a deeper understanding of their characteristics and user appeal [16]. In a separate empirical study, Lin et al. focused specifically on analyzing the reviews of 6224 games available on the Steam platform [31]. Their investigation encompassed an examination of review content, as well as an exploration of the relationship between players' play hours and the content of their respective reviews. Through their comprehensive analysis, the researchers revealed intriguing insights into the factors influencing players' opinions and the connection between their engagement levels and expressed feedback. Collectively, these studies contribute valuable knowledge to the field of mobile app analysis, shedding light on various aspects such as review length, sentiment expression, app features, and the interplay between play hours and reviews.

Santos et al. conducted a comparative analysis between expert and amateur game reviews sourced from Metacritic. Their study aimed to explore the differences in sentiment and polarization between these two reviewer groups. Intriguingly, the findings revealed that amateur reviews exhibited a higher level of polarization and expressed stronger sentiments when compared to expert reviews [45]. In a similar vein, Lu et al. embarked on an investigation utilizing topic modeling techniques applied to Steam reviews [32]. Their objective was to delve into the temporal dynamics of player review topics and understand how updates to games influenced these dynamics. Through their rigorous analysis, the researchers uncovered valuable insights into the shifting landscape of player discussions over time. Moreover, their findings illuminated the impact of game updates on the topics and themes that players emphasized in their reviews.

Adam Summerville and Sam Snodgrass explore a contemporary approach to game design that involves the utilization of machine learning techniques. In contrast to traditional methods like search-based, solver-based, and constructive approaches, the authors specifically concentrate on the analysis of functional game content [51]. Similarly, Anurag Sarkar and Seth Cooper have coined the term "Game Design via Creative Machine Learning" (GDCML) to describe a novel approach that combines machine learning techniques with game design. GDCML refers to a specific subset of procedural content generation methods that employ models trained on one or more games [46]. This approach enables the development of creative machine learning applications and tools for game design, akin to the ones commonly observed in visual art and music.

Yilei Zeng and Aayush Shah have developed a comprehensive machine learning course specifically tailored for graduate students who are keen on exploring the cutting-edge advancements in deep learning and reinforcement learning within the gaming domain [60]. This course serves as a crucial link, fostering interdisciplinary collaboration among graduate schools. Importantly, it is designed to be accessible to students without any prior experience in game design or development. Cundong Tang and Zhiping Wang conduct a comprehensive analysis of the historical progression and current state of artificial intelligence (AI) in game development. They delve into the intricate dynamics of AI technology, particularly machine learning, and its profound implications for the future of game development [52]. Their work sheds light on the evolving relationship between AI and games, providing valuable insights for both industry professionals and researchers in this dynamic field.

5.2 A Labelled Dataset of Game Playability

In this section, we describe the collection and preparation of the dataset of the playability category for each of the seven categories.

5.2.1 Data Collection

We have used a total number of 20,346 reviews collected with the use of Python scraper called google play scraper. We collected the most recent 1700 reviews as seen in 3.1 from each game.

5.2.2 Data Labelling

We label our data set of all the seven playability categories of the famous Sánchez model of playability values [44]. For labelling a data set of each playability category using 20,346 reviews, we have used two approaches: dictionary-based and manual labelling. In the first step for each seven-playability category, we created a sample dataset of 10,000 reviews from our dataset. Then we applied a dictionary-based approach to the sample datasets of each category. The Human value dictionary is changed before applying a dictionary-based approach because it depends on the playability category. After applying a dictionary-based approach to all seven-playability categories, we have some reviews from the sample datasets of each category that violate the playability category, and some reviews don't possess violations. We then created a separate dataset of the reviews, both violated and non- violated of each playability category.

In the second step, we manually chose 500 reviews from the violated dataset of each category and labelled them as ONE (1), and we chose another 500 reviews from the non-violated dataset of each category and labelled them as ZERO (0) and created a final dataset of 1000 reviews for each category. For two categories *Emotion* and *Motivation*, we created a labelled dataset of 800 reviews because we have fewer violations of these categories in the sample dataset. These datasets of each category are now ready for applying the machine learning classification model.

5.3 Proposed approach

The main aim of this approach is to develop machine learning models to automatically classify whether a particular review violates a specific playability category. The machine learning models are applied on labelled dataset of each playability category.

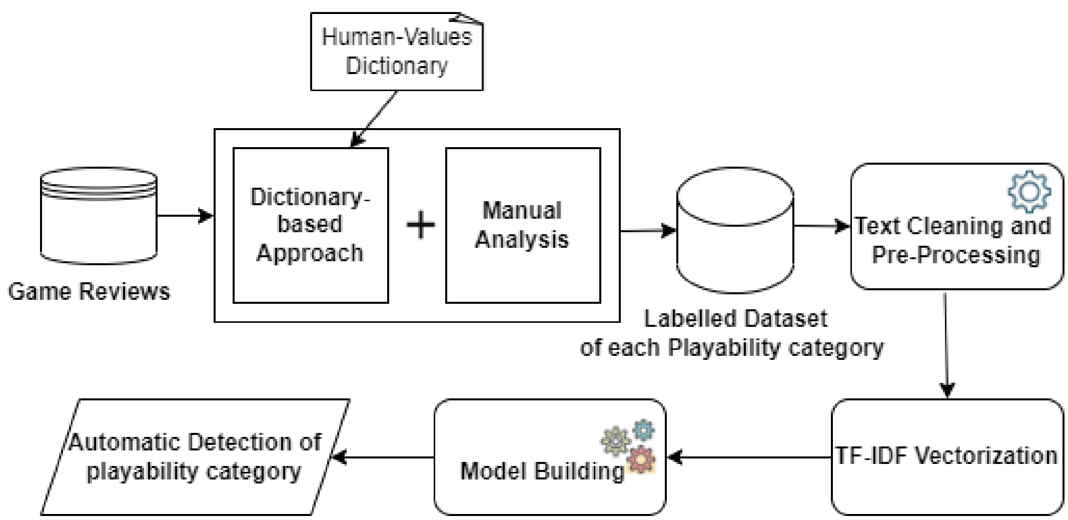

Figure 5.1: Overview of the proposed Machine Learning approach

5.3.1 Data Cleaning and Pre-processing

We applied some common techniques like remove missing and duplicate values from each of our labelled dataset to remove possible noise. Machine learning models cannot understand human language; We applied the following natural language processing technique to pre-process all the datasets. This step was necessary so a learning model could classify reviews correctly.

1. **Lowered case:** Our first pre-processing step includes the lowering of the textual content.

2. **Removal of Punctuation:** We removed all punctuation from the string using the string library. For example, for the string "(There is a tree, near the river!)", after removing all punctuation, we can get "There is a tree near the river".

3. **Removal of StopWords:** We removed all stop-words from the subject of the change using NLTK library. For example, for the same string, after removing stop-words we can get 'there', 'tree, 'near', 'river'.

4. **Stemming:** Stemming is an NLP process in which the word is reduced to its base form. For this, we use the snowball stemmer. It is a porter stemmer technique that improves a searching capability.

5. **Removal of Emoji:** Emojis are symbols or a small number of Unicode characters that people can use to express feelings, concepts, and ideas. Emojis may influence a model's performance in terms of accuracy if they are not adequately pre-processed. As a result, we did away with emojis in the review texts.

5.3.2 TF-IDF Vector Transformation

After cleaning and pre-processing the dataset, we converted the game reviews in the dataset into their vector representation by using TF-IDF vectorization. It transforms a string into numeric vectors. The first step to implement TF-IDF, is tokenization where a string is tokenized into a bag of words. Then we applied a TF-IDF vectorizer to the datasets. We take into account a word's entire document weightage in TF-IDF Vectorizer. The word counts are weighted by a measure of how frequently they appear in the datasets by TF-IDF Vectorizer.

5.3.3 Model Building

In machine learning, choosing a classification algorithm is a very crucial task. We choose 11 Classification models such as Logistic Regression, Support Vector Machine, Multinomial Naïve bayes , Decision Tree, K Neighbors, RandomForest, AdaBoost, Bagging Classifier, ExtraTrees Classifier, GradientBoosting Classifier, xgboost Classifier. The descriptions of some classification model we employed are provided below.

1. **Logistic Regression** is a statistical model called the logistic model uses the log-odds of an event as a linear combination of one or more independent variables to represent the likelihood that the event will take place.

2. **RandomForest** is widely used in Classification and Regression Problems. It builds decision trees on several samples, using the majority vote for categorization and the average for regression.

3. **Bagging Classifier** A bagging classifier is an ensemble meta-estimator that fits basic classifiers one at a time to random subsets of the original dataset, then combines each prediction (either through voting or average) to get the final prediction.

4. **Gradient boosting** classifiers is a class of machine learning techniques. It combines a number of weak learning models to produce a powerful predicting model. Gradient boosting frequently makes use of decision trees.

5. **XGBoost classifier** is one of the machine learning algorithms which is applied for structured and tabular data. It is an extreme gradient boost algorithm. Therefore, it is a large machine learning method with numerous components. Large, intricate datasets are compatible with XGBoost. It is an ensemble modelling method.

5.3.4 K-Fold Cross-Validation

Cross-validation is a re-sampling technique. It is used to assess machine learning models on a small data sample. The process contains a single parameter, k, that denotes how many groups should be created from a given data sample. As a result, the process is frequently referred to as k-fold cross-validation.

We used a 10-fold cross-validation technique to estimate each classification model's performance. Here, we split the labelled dataset into 10 chunks of data that contain an equal number of game reviews. Then, we perform the evaluation process where the training dataset contains 9 chunks of data and another chunk of data is used as the testing dataset. Note that this is repeated until each chunk of data has been used as the testing dataset once. In the end, we choose the average accuracy out of the 10 for evaluation. This method enables us to evaluate the performance of our chosen models using random data.

5.4 Experiments and Results

In this study, we present the findings of our experiment that assessed the performance of various machine learning models.

Table 5.1: Performance Analysis of Classifiers for Playability Categories (Results without Cross-Validation)

Category	Classifier	Precision	Recall	F1 score	Accuracy
Socialism	Logistic Regression	0.71	0.71	0.72	0.77
Effectiveness	Xgboost	0.85	0.70	0.77	0.795
Immersion	Bagging Classifier	0.88	0.88	0.88	0.885
Satisfaction	Xgboost	0.79	0.80	0.80	0.80
Learnability	GradientBoosting	0.80	0.76	0.78	0.79
Motivation	Random forest	0.85	0.67	0.75	0.76
Emotion	Xgboost	0.85	0.86	0.85	0.84

5.4.1 RQ4: Can we automatically detect the violations of the playability categories documented in the user/player reviews?

In this experiment, we utilized machine learning classification models to analyze all seven labelled datasets related to the playability category. To evaluate their effectiveness, we utilized widely recognized metrics such as accuracy, precision, recall, and F1 score. These metrics are commonly used to measure the models' performance in solving a given problem.

Table 5.1 shows the best performing classification model for each playability category. It is evident that the overall accuracy of the models exceeds 0.75. The XGBoost classifier stands out as the top-performing algorithm for the *Effectiveness*, *Satisfaction*, and *Emotion*, categories.

In addition to the previous evaluation, we incorporated the k-fold cross-validation technique for all the labeled datasets. Table 5.2 illustrates the best-performing classification model for each playability category when utilizing the cross-validation technique. The XGBoost classifier performs better for the *Socialism*, *Effectiveness*, *Satisfaction*, *Motivation*, and *Emotion* categories. Moreover, the Bagging classifier emerges as the top-performing algorithm for the *Immersion* and *Learnability* categories.

The primary objective of applying the k-fold cross-validation technique was to enhance the accuracy of our classification models and evaluate their performance on unseen data. Our analysis revealed notable improvements in accuracy for the *Socialism* and *Effectiveness* categories when employing cross-validation. However, in the *Satisfaction* category, the accuracy remained relatively unchanged, suggesting that the models performed consistently without the need for cross-validation. Interestingly, without implementing cross-validation, we observed superior performance in terms of accuracy for the *Immersion*, *Learnability*, *Motivation*, and *Emotion* categories. This indicates that these models demonstrated

Table 5.2: Performance Analysis of Classifiers for Playability Categories (Results with Cross-Validation)

Category	Classifier	Precision	Recall	F1 score	Accuracy
Socialism	Xgboost	0.80	0.78	0.79	0.79
Effectiveness	Xgboost	0.87	0.76	0.81	0.82
Immersion	Bagging Classifier	0.83	0.81	0.82	0.82
Satisfaction	Xgboost	0.81	0.79	0.80	0.80
Learnability	Bagging Classifier	0.75	0.77	0.76	0.76
Motivation	Xgboost	0.86	0.77	0.81	0.82
Emotion	Xgboost	0.82	0.86	0.84	0.83

robustness and generalization capabilities on the available data alone. Furthermore, we employed a voting classifier in an attempt to further enhance accuracy. However, our findings did not indicate any significant changes in accuracy when compared to individual classifiers.

In summary, the application of k-fold cross-validation resulted in accuracy improvements for specific playability categories, while other categories exhibited strong performance without the need for cross-validation. Our results underscore the importance of considering the specific dataset and playability category when determining the effectiveness of cross-validation techniques.

Finally we develop a python based web tool .The primary objective of this tool is to provide automated classification of reviews based on their violations to specific playability categories. We used python-based library and open-source platform called Streamlit[1] to develop this web app tool. It is an exceptional, complimentary framework that empowers you to swiftly construct and distribute visually captivating web applications dedicated to machine learning and data science. Its open-source nature enables seamless collaboration and enables developers to craft stunning interfaces with ease, enhancing the user experience. The Figure 5.2 showcases the initial interface of our web application tool.

During the development of our web application, we employ a technique called pickling to save and store the best performing machine learning model for each playability category. This enables us to effectively determine whether a review violates a specific playability category or not. When a user submits a review and clicks on the "Make Prediction" button, the review text undergoes a series of comprehensive natural language processing (NLP) pre-processing tasks describe in the section . Then it predicts the likelihood of a review violating various playability categories. Subsequently, a curated list of the playability categories that the review potentially violates is displayed to the user. Figure 5.3 shows another

[1] https://streamlit.io/

Figure 5.2: Initial view of the Web tool

snapshot of our tool. Users are invited to compose game reviews within the provided text box. Upon clicking the "Make Prediction" button, our web tool generates a comprehensive list of playability categories, pinpointing any violations present in the review.

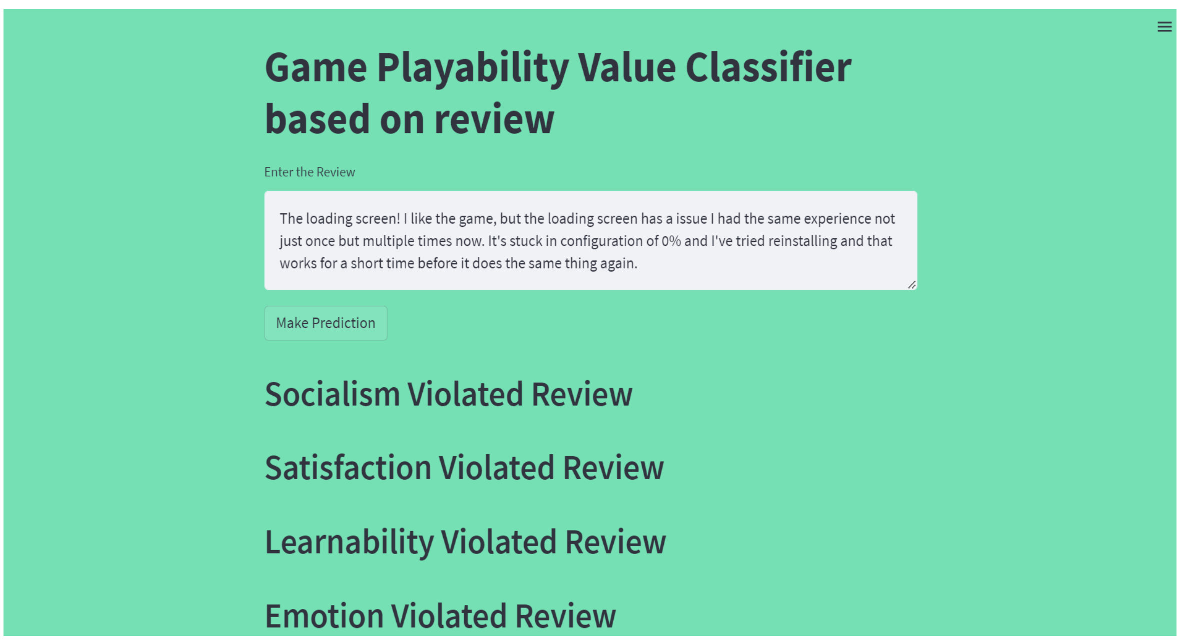

Figure 5.3: A Screenshot of the review violation's results

Chapter 6

Conclusions and Future Work

We presented an approach for analyzing the playability of gaming apps. We considered 15 mobile game apps and evaluated their playability by looking at game-related user app reviews posted by game users on Google Play Stores. In this process, we first identify the human value violations in the user reviews and then subsequently map the so-obtained human value violations with the game playability values. The mapping of violations helps the developers understand the game's playability violations from the user's perspective. Finally, we introduced an automated approach to classify game app reviews based on identifying playability violations from the perspective of end-users. In developing our automated system, we evaluated eleven different and famous classification algorithms using a manually annotated and validated dataset of game app reviews. The results revealed violations in the user reviews; out of 20,346 user/player reviews, 6,230 reviews have human value violations that affect the game's playability. We also discovered that only a few of the reviews (from the ones that were violated) received developer responses that addressed the related concerns and resolved the related violations. The presented analysis also identifies the functionalities, issues, or concerns that affect the game's playability. Also, our evaluation results demonstrated that our top-performing algorithms consistently achieved an overall accuracy exceeding 0.75. Furthermore, our findings highlight how a specific classification algorithm performs best in a particular playability category with and without cross-validation techniques. These analysis results encourage and help app developers to learn about the types of violations and consider such violations while developing mobile game apps.

As a part of future work, collecting feedback from the game app developers on the findings and whether the study results help them improve the app would be interesting.

References

[1] I. 9241-11:2018. Ergonomics of human-system interaction — part 11: Usability: Definitions and concepts. *International Organization of Standardization*, 2018.

[2] A. Ampatzoglou and I. Stamelos. Software engineering research for computer games: A systematic review. *Information and Software Technology*, 52(9):888–901, 2010.

[3] A. Bajaj. Can python understand human feelings through words? – a brief intro to nlp and vader sentiment analysis. https://www.analyticsvidhya.com/blog/2021/06/vader-for-sentiment-analysis/, 2022.

[4] P. Bontrager and J. Togelius. Learning to generate levels from nothing. In *2021 IEEE Conference on Games (CoG)*, pages 1–8, 2021.

[5] T. Burger-Helmchen and P. Cohendet. User communities and social software in the video game industry. *Long Range Planning*, 44(5-6):317–343, 2011.

[6] N. Chen, J. Lin, S. C. Hoi, X. Xiao, and B. Zhang. Ar-miner: mining informative reviews for developers from mobile app marketplace. In *Proceedings of the 36th international conference on software engineering*, pages 767–778, 2014.

[7] A.-S. Cheng and K. R. Fleischmann. Developing a meta-inventory of human values. *Proceedings of the American Society for Information Science and Technology*, 47(1):1–10, 2010.

[8] H. Desurvire, M. Caplan, and J. A. Toth. Using heuristics to evaluate the playability of games. In *CHI'04 extended abstracts on Human factors in computing systems*, pages 1509–1512, 2004.

[9] A. Di Sorbo, S. Panichella, C. V. Alexandru, J. Shimagaki, C. A. Visaggio, G. Canfora, and H. C. Gall. What would users change in my app? summarizing app reviews for recommending software changes. In *Proceedings of the*

2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering, pages 499–510, 2016.

[10] Dictionary.com. Definition for 'playability'. https://www.dictionary.com/browse/playability?s=t, 2023.

[11] B. Friedman, D. G. Hendry, and A. Borning. A survey of value sensitive design methods. *Foundations and Trends® in Human–Computer Interaction*, 11(2):63–125, 2017.

[12] T. Fullerton, C. Swain, and S. Hoffman. *Game design workshop: Designing, prototyping, & playtesting games*. CRC Press, 2004.

[13] W. Geyser. 20 mobile gaming statistics that will blow you away | mobile gaming industry stats. https://influencermarketinghub.com/mobile-gaming-statistics/, 2022.

[14] V. V. Gouveia, T. L. Milfont, and V. M. Guerra. Functional theory of human values: Testing its content and structure hypotheses. *Personality and Individual Differences*, 60:41–47, 2014.

[15] E. Guzman and W. Maalej. How do users like this feature? a fine grained sentiment analysis of app reviews. In *2014 IEEE 22nd international requirements engineering conference (RE)*, pages 153–162. Ieee, 2014.

[16] M. Harman, Y. Jia, and Y. Zhang. App store mining and analysis: Msr for app stores. In *2012 9th IEEE working conference on mining software repositories (MSR)*, pages 108–111. IEEE, 2012.

[17] C. Hutto and E. Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225, 2014.

[18] C. Iacob and R. Harrison. Retrieving and analyzing mobile apps feature requests from online reviews. In *2013 10th working conference on mining software repositories (MSR)*, pages 41–44. IEEE, 2013.

[19] I.Blair. Mobile app download statistics usage statistics (2022). https://buildfire.com/app-statistics/, 2022.

[20] M. Intelligence. Mobile gaming market - growth, trends, covid-19 impact, and forecasts (2023 - 2028). https://www.mordorintelligence.com/industry-reports/mobile-games-market, 2022.

[21] J.Clement. Mobile gaming market in the united states - statistics facts. https://www.statista.com/topics/1906/mobile-gaming/#topicOverview, 2022.

[22] K. Karthikeyan. The history, evolution, and future of mobile gaming. https://www.gameopedia.com/the-history-evolution-and-future-of-mobile-gaming/, 2022.

[23] J. M. Kibbee. Management games and computers. In *Papers presented at the May 9-11, 1961, western joint IRE-AIEE-ACM computer conference*, pages 11–16, 1961.

[24] H. Korhonen. Evaluating playability of mobile games with the expert review method. 2016.

[25] S. Krishtul, M. Shahin, H. O. Obie, H. Khalajzadeh, F. Gai, A. Rezaei Nasab, and J. Grundy. Human values violations in stack overflow: An exploratory study. In *Proceedings of the International Conference on Evaluation and Assessment in Software Engineering 2022*, pages 70–79, 2022.

[26] J. Kruse, R. Sosa, and A. M. Connor. Procedural urban environments for fps games. In *Proceedings of the Australasian computer science week multiconference*, pages 1–5, 2016.

[27] A. Kultima and K. Alha. "hopefully everything i'm doing has to do with innovation": Games industry professionals on innovation in 2009. In *2010 2nd International IEEE Consumer Electronics Society's Games Innovations Conference*, pages 1–8. IEEE, 2010.

[28] G. Leitmann. On some consequences of playability in differential games. In *1974 IEEE Conference on Decision and Control including the 13th Symposium on Adaptive Processes*, pages 346–348. IEEE, 1974.

[29] X. Li, Z. Zhang, and K. Stefanidis. A data-driven approach for video game playability analysis based on players' reviews. *Information*, 12(3):129, 2021.

[30] D. Lin, C.-P. Bezemer, Y. Zou, and A. E. Hassan. An empirical study of game reviews on the steam platform. *Empirical Software Engineering*, 24(1):170–207, 2019.

[31] D. Lin, C.-P. Bezemer, Y. Zou, and A. E. Hassan. An empirical study of game reviews on the steam platform. *Empirical Software Engineering*, 24:170–207, 2019.

[32] C. Lu, X. Li, T. Nummenmaa, Z. Zhang, and J. Peltonen. Patches and player community perceptions: Analysis of no man's sky steam reviews. In *DiGRA'20-Proceedings of the 2020 DiGRA International Conference*. DiGRA, 2020.

[33] V. Mello and L. Perani. Gameplay x playability: defining concepts, tracing differences. *Simpósio Brasileiro de Jogos e Entretenimento Digital*, 11(2012):157–164, 2012.

[34] L. Nacke. From playability to a hierarchical game usability model. In *Proceedings of the 2009 Conference on Future Play on@ GDC Canada*, pages 11–12, 2009.

[35] H. O. Obie, W. Hussain, X. Xia, J. Grundy, L. Li, B. Turhan, J. Whittle, and M. Shahin. A first look at human values-violation in app reviews. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*, pages 29–38. IEEE, 2021.

[36] H. O. Obie, I. Ilekura, H. Du, M. Shahin, J. Grundy, L. Li, J. Whittle, and B. Turhan. On the violation of honesty in mobile apps: Automated detection and categories. *arXiv preprint arXiv:2203.07547*, 2022.

[37] H. O. Obie, M. Shahin, J. Grundy, B. Turhan, L. Li, W. Hussain, and J. Whittle. Does domain change the opinion of individuals on human values? a preliminary investigation on ehealth apps end-users. In *2021 28th Asia-Pacific Software Engineering Conference (APSEC)*, pages 531–535. IEEE, 2021.

[38] J. Paavilainen. Defining playability of games: functionality, usability, and gameplay. In *Proceedings of the 23rd International Conference on Academic Mindtrek*, pages 55–64, 2020.

[39] S. Panichella, A. Di Sorbo, E. Guzman, C. A. Visaggio, G. Canfora, and H. C. Gall. How can i improve my app? classifying user reviews for software maintenance and evolution. In *2015 IEEE international conference on software maintenance and evolution (ICSME)*, pages 281–290. IEEE, 2015.

[40] H. Perera, W. Hussain, J. Whittle, A. Nurwidyantoro, D. Mougouei, R. A. Shams, and G. Oliver. A study on the prevalence of human values in software engineering publications, 2015-2018. In *2020 IEEE/ACM 42nd International Conference on Software Engineering (ICSE)*, pages 409–420. IEEE, 2020.

[41] M. V. Phong, T. T. Nguyen, H. V. Pham, and T. T. Nguyen. Mining user opinions in mobile app reviews: A keyword-based approach (t). In *2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, pages 749–759. IEEE, 2015.

[42] D. Pinelle, N. Wong, and T. Stach. Heuristic evaluation for games: usability principles for video game design. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1453–1462, 2008.

[43] L. Qiu, Y. Zhao, J. Li, P. Lu, B. Peng, J. Gao, and S.-C. Zhu. Valuenet: A new dataset for human value driven dialogue system. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, 44(5):2468–2484, 2022.

[44] J. L. G. Sánchez, F. L. G. Vela, F. M. Simarro, and N. Padilla-Zea. Playability: analysing user experience in video games. *Behaviour & Information Technology*, 31(10):1033–1054, 2012.

[45] T. Santos, F. Lemmerich, M. Strohmaier, and D. Helic. What's in a review: Discrepancies between expert and amateur reviews of video games on metacritic. *Proceedings of the ACM on human-computer interaction*, 3(CSCW):1–22, 2019.

[46] A. Sarkar and S. Cooper. Towards game design via creative machine learning (gdcml). In *2020 IEEE Conference on Games (CoG)*, pages 744–751, 2020.

[47] S. H. Schwartz. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental social psychology*, volume 25, pages 1–65. Elsevier, 1992.

[48] S. H. Schwartz. An overview of the schwartz theory of basic values. *Online readings in Psychology and Culture*, 2(1):2307–0919, 2012.

[49] R. A. Shams, W. Hussain, G. Oliver, A. Nurwidyantoro, H. Perera, and J. Whittle. Society-oriented applications development: Investigating users' values from bangladeshi agriculture mobile applications. In *2020 IEEE/ACM 42nd International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*, pages 53–62. IEEE, 2020.

[50] R. A. Shams, M. Shahin, G. Oliver, J. Whittle, W. Hussain, H. Perera, and A. Nurwidyantoro. Human values in mobile app development: An empirical study on bangladeshi agriculture mobile apps. *arXiv preprint arXiv:2110.05150*, 2021.

[51] A. Summerville, S. Snodgrass, M. Guzdial, C. Holmgård, A. K. Hoover, A. Isaksen, A. Nealen, and J. Togelius. Procedural content generation via machine learning (pcgml). *IEEE Transactions on Games*, 10(3):257–270, 2018.

[52] C. Tang, Z. Wang, X. Sima, and L. Zhang. Research on artificial intelligence algorithm and its application in games. In *2020 2nd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM)*, pages 386–389, 2020.

[53] F. T. Tschang. Balancing the tensions between rationalization and creativity in the video games industry. *Organization science*, 18(6):989–1005, 2007.

[54] F. T. Tschang and J. Szczypula. Idea creation, constructivism and evolution as key characteristics in the videogame artifact design process. *European management journal*, 24(4):270–287, 2006.

[55] R. Vasa, L. Hoon, K. Mouzakis, and A. Noguchi. A preliminary analysis of mobile app user reviews. In *Proceedings of the 24th Australian computer-human interaction conference*, pages 241–244, 2012.

[56] R. Vasa, L. Hoon, K. Mouzakis, and A. Noguchi. A preliminary analysis of vocabulary in mobile app user reviews. In *Proceedings of the 24th Australian Computer-Human Interaction Conference, Sydney, Australia*, pages 245–248, 2012.

[57] J. Whittle, M. A. Ferrario, W. Simm, and W. Hussain. A case for human values in software engineering. *IEEE Software*, 38(1):106–113, 2019.

[58] J. Wiemeyer, L. Nacke, C. Moser, and F. 'Floyd'Mueller. Player experience. *Serious games: Foundations, concepts and practice*, pages 243–271, 2016.

[59] E. Winter, S. Forshaw, L. Hunt, and M. A. Ferrario. Advancing the study of human values in software engineering. In *2019 IEEE/ACM 12th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)*, pages 19–26. IEEE, 2019.

[60] Y. Zeng, A. Shah, J. Thai, and M. Zyda. Applied machine learning for games: A graduate school course. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 15695–15703, 2021.