Mobile Games Success and Failure: Mining the Hidden Factors

Abdulrahman Kerim* [ID] · **Burkay Genç** [ID]

Abstract Predicting the success of a mobile game is a prime issue in game industry. Thousands of games are being released each day. However, a few of them succeed while the majority fail. Towards the goal of investigating the potential correlation between the success of a mobile game and its specific attributes, this work was conducted. More than 17 thousand games were considered for that reason. We show that IAPs (In-App Purchases), genre, number of supported languages, developer profile, and release month have a clear effect on the success of a mobile game. We also develop a novel success score reflecting multiple objectives. Furthermore, we show that game icons with certain visual characteristics tend to be associated with more rating counts. We employ different machine learning models to predict a novel success score metric of a mobile game given its attributes. The trained models were able to predict this score, as well as the expected rating average and rating count for a mobile game with 70% accuracy.

*Corresponding author.

Abdulrahman Kerim
Computing and Communications, Lancaster University, UK.
Tel.: (+440) 7341-7077-72
E-mail: a.kerim@lancaster.ac.uk

Burkay Genç
Computer Engineering, Hacettepe University, TR.
Tel.:(+90) 5424-8619-79
E-mail: burkay.genc@hacettepe.edu.tr

1 Introduction

In 1997, the first mobile game, *Snake*, was released by *NOKIA*. This mobile game initiated a new era in the world of game industry. Game industry, today, is a huge market and notably growing by the time. Statistics reveal that the value of the video game market in the United States by 2020 was more than 60 billion dollars [18] which discloses the seriousness and the value of this business. In Apple App Store only there are currently more than 265 thousand games [17]. Free distribution and the low cost of mobile application development allowed both startups and indie developers to enter this flourishing market easily. At the same time, the recent advances in mobile technology made mobile phones more powerful and cheaper. This, in turn, increased memory size, processing power, and graphics quality of mobile devices, which supported more sophisticated and enjoyable games.

Nevertheless, a perfect graphical design, an interesting storyline, appealing animations, and attractive characters are expected to be the main factors behind the success of a game, we can see a lot of examples where simple and non-professional looking games have reached the peak -in terms of popularity and revenue. Apparently, there is more to mobile success than meets the eye. Indeed, studying success and failure stories of mobile games may help us to better comprehend the direct and indirect causes of success.

In this work, we present a thorough study where we analyze more than 17 thousand games to answer the main question of "what makes a game fail or succeed?" An important problem in mobile game industry is the difficulty in defining a general game success measure and another one is predicting the success of a particular game before it is published. We undertook this study to

ask and answer the mentioned two questions following a rigorous scientific approach. We argue that number of ratings or average rating of a mobile game shown in App Store are not enough measures to describe the success of a game. Thus, we introduce a new measure to quantitatively assess the success of a mobile game and use data mining to build accurate models for predicting its value.

In order to predict this novel success metric, we employ data mining methodology on an existing dataset of mobile games. Through data mining, we aim to discover the most important factors determining the value of this measure and we provide means of predicting it based on these factors. A series of steps and procedures are ought to be followed carefully to reach the ultimate goal of understanding the hidden facts and rules that control the phenomenon under consideration [30, 19]. A pre-processing step is usually required to discard unnecessary data (data cleaning), transform the data from one form to another, or perform some data regularization and normalization for later steps. Following this, the feature engineering takes place and with the help of domain knowledge, new features can be introduced. Finally, suitable algorithms can be applied to extract information from the data [39, 16]. Under the scope of this study, we call attention to some unique attributes that were shown to influence the success of a mobile game.

At the same time, we extend our study to the visual features of mobile icons. The visual features of icons are essential for grabbing the attention to the mobile game. This becomes a playing factor once a game is released among thousands of other games. Players cannot appreciate the content of the game if they did not click the icon first and download it. Thus, a great attention should be given to make the game appealing to a large set of users. Unfortunately, the subjective nature of the problem makes it an arduous process for designers to anticipate the attractiveness of an icon. At the same time, the attractiveness may vary from one group of users to others. For example, it is not expected from old and young users to like the same icons. Education, gender, and culture could be other playing factors too. Even, what makes the problem much harder is the time-dependency nature of it. In other words, appealing icons 10 years ago are not expected to be appareling nowadays.

Using a statistical approach instead of the subjective evaluation based studies seems as a resort for the mentioned problems. By looking at the success profile of thousands of games together with the visual features of the icons, we try to spotlight the common visual features observed in top successful mobile games and give recommendations for an appealing icon design. Thus, game developers may consider these suggestions when designing their mobile games icons. Similarly, we highlight the common features observed in the top unsuccessful games icons so that the developers could avoid them. The main contributions of this paper are as follows:

- A novel success score metric reflecting multiple objectives of mobile games development.
- Machine learning models to understand and predict mobile game success.
- Highlighting correlation between mobile game success and certain attributes.
- Finding a correlation between specific visual features of mobile game icons and high rating counts.

2 Related Work

In the literature, we can see a few number of studies which discuss similar questions to the ones asked and answered in this paper. However, what distinguishes each of them is the success definition, the studied attributes, and the scale of the considered mobile games. Thus, generalizability of the drawn conclusions.

Moreira et al. (2014) in [31] considered the top 100 games of Google Play App Store. The number of downloads and revenues were studied with respect to 37 different features associated with the games. They concluded that a feature like allowing IAPs is associated with success. On the other hand, features like inviting friends and customizing the game were shown to be usually linked with the opposite. In parallel to that work, Alomari et al. (2016) in [3] studied the revenue with respect to 31 attributes for 50 iPhone games. The aim of the study was to identify the most 10 important features in game development. They confirmed a strong relation between the total number of people who run or get engaged with a mobile game and its revenue. Similar to these two works, we examined various game's attributes and their effect on average rating and rating count. However, instead of considering a limited number of mobile games, we considered 17 thousand mobile game apps to increase the statistical significance of the findings.

Lee et al. (2014) in [27] concluded that selecting a less competitive genre for an application and keeping the application's quality at higher levels, by updating the application frequently, were shown to be positively correlated with the application staying a longer time among the top applications. In [26], the authors state that the previous history of the application releasing company was shown to be correlated with application

Table 1: The main attributes of the dataset used in this work.

Attribute Name	Type	Attribute Name	Type	Attribute Name	Type
URL	textual	In-app Purchases	numeric	Current Version Release Date	date
ID	numeric	Description	textual	Original Release Date	date
Name	textual	Developer	textual	Price	numeric
Subtitle	textual	Age Rating	categorical	Genres	categorical
Icon URL	textual	Languages	categorical	Primary Genre	categorical
Average User Rating	numeric	Size	numeric	User Rating Count	numeric

success. Similar to these two studies, we examine the success of a mobile game given its genre and the developer maturity. However, we do not study all apps categories but we focus only on mobile application.

Yi et al. (2019) in [45] showed that advertising a mobile game on TV and uploading it on many App Stores increase its chance of success. At the same time, they found that the success of the game decreases over time. In the scope of our work, we do not consider the advertisement parameters. However, that could be done in a future work. The main question could be whether our findings are consistent with games that are released in different App Stores.

Understanding the visual attributes of icons and their influence on the click, download, and rating count could make the game more popular. This is essential for attracting first user who does not have any information about the actual content of the game. Few works tried to tackle this problem and give suggestions and recommendations on how to design an icon that could impact and contribute to the success of a game.

The icon design and its relations to mobile game downloads and purchases were examined by Jylha et al. (2019) in [22] for 68 game applications. Uniqueness and realism are found to cause more downloads, clicks, and purchases. In our work, we do not consider the icon design. However, a detailed study was carried on the visual features of the most successful and unsuccessful games. Then, conclusions were drawn on common features observed in each cluster.

Kukka et al. (2013) in [25] showed that color, animation, or textual elements could impact the attractiveness of users to public screens. This seems related to mobile games and icon design. Because the player's first impression about the game could highly be influenced by the icon itself. Similar to this work, we study color information of mobile icons. However, in a more detailed perspective.

Shen et al. (2020) in [37] focused more on the internal characteristics of icons (like the icon meaning) and their relation to the cognitive tasks such as algebraic equation. They gave important suggestions on the icon design. In our work, we investigated, similarly, the ef-

fect of icons' entropy on the decision of downloading the game. At the same time, we provided a set of suggestions for mobile game designers and developers on how the most attractive icon should be, based on the knowledge extracted from the analysed mobile games. Mcdougall et al. (2016) in [29] analysed the determinants of icon appeal in details. Icon familiarity was shown to be correlated positively with icon appealing. Consequently, they found that users who cannot understand an icon, tend to classify it as an unattractive one. Instead of targeting user interfaces, we focus on icons shown in App Store. Ghayas et al. (2013) in [15] showed that recognition rate of icons was related to user's age. They studied 40 icons from two mobile phone brands. What may hint to a possible relation between the age of the person and the tendency to download a mobile game -solely- based on the icon itself. This could be a future research direction to study a possible correlation between some specific icon visual features and tendency of some age groups to download this game.

Kamarulzaman et al. (2020) in [23] carried a comparative study on icon design for mobile applications. They analysed color, semantics, uniqueness, and concreteness elements. They provided general suggestions for designing mobile apps icon as well.

2.1 Machine Learning Methods

Apriori Algorithm Apriori algorithm [1,32] identifies the trends and extracts association rules in a given dataset. In general, it utilizes breadth-first search and a hash tree to find frequent items. It has many applications specially in market basket analysis [2,42] and speech emotion recognition [36].

Support Vector Machines A supervised learning-based algorithm that can be used for both classification and regression problems [10,5]. It works by finding the optimal hyperplane separating the classes while maximizing the margin between them. It is widely applied for various problems like large-scale image classification [28,12], biometrics [33,40], and weather forecasting [34]. Close to our work, Drachen et al. (2016) in [13] utilized

support vector machines to predict the retention of a user in free-to-play mobile games.

Random Forest An ensemble learning algorithm that can be applied for a wide set of classification and regression problems [21,8] such as bioinformatics [35], remote sensing [4], and finance [46]. In general, random forest is utilized to enhance the learning process by combining a number of decision trees. These decision trees are trained on different partitions of the same training set to reduce the overall variance. Thus, to improve generalizability of trained model. Close to our work, Sifa et al. (2015) in [38] deployed two random forest models. The first to predict if a user will purchase a game. While, the other to guess the number of purchases a specific user will make in the future.

Artificial Neural Networks Supervised learning-based models that can be applied for classification and regression problems. They learn the mapping from input space to output space using the training samples. They have been widely applied in different domains such as stock market prediction [20], object detection [43], and semantic segmentation [44]. Similar to our work, Dehkordi et al. (2020) in [11] utilized an artificial neural network to predict the success of Android applications in a new applications repository.

3 Methodology

In this section, we outline our methodology. First, data preprocessing and feature engineering efforts are considered in Sec. 3.1. Second, our novel success score metric is discussed in Sec. 3.2. Third, the predictive models are summarized in Sec. 3.3.

3.1 Data Preprocessing

The used dataset [41] contains more than 17 thousand mobile games taken from Apple App Store. Initially, each game is associated with 18 different attributes. A negligible number of these values are missing. It is relatively a recent dataset since it was collected in August of 2019. That makes it a good proxy of the actual game market. In Table 1, the main attributes of the dataset are shown together with their types.

Attribute Selection Unnecessary attributes such as the *URL*, *ID*, and *ICON.URL* are removed because they are irrelevant to the success of a mobile game. The *ID* column simply represents a unique identifier for each mobile game. Thus, it was removed as the games are randomly given such an identifier. At the same time, we calculated the Relief score [24] for *ID* and *ICON.URL* attributes. We found a negligible correlation between

these attributes and the success scores. Other attributes like the *Subtitle* and *Description* are discarded as well under this scope. However, they can be good candidates for our future work. *Primary.Genre* is neglected since more than 96% of the observations were marked as games. We note that the actual game genre information is presented in the *Genres* attribute which was taken into account.

Handling Missing Data Games with null values for *User Rating Count* means they have less than 5 ratings. We assign them 5 rating counts. At the same time, games with null values for *Average User Rating* means they did not gain more than 5 rating counts. Because of this, we assign them rating average of 0. Following this, we remove games with missing values for *Price* and *Size* attributes. After that, in the *Languages* column, we see some cells are left with no entry, we handle this by assigning the English language for these games. We note that the English language constitutes the majority in the dataset and hence was selected as the default for imputation.

Feature Engineering The attributes of the final dataset are shown in Table 2. Based on the raw dataset, few more attributes were derived to further enhance the level of information associated with each single game. The *Number of Languages* is extracted by counting the number of languages supported by each game. The *Number of Genres* is created similarly by counting the number of genres under which the game is classified. After that, the game's developer maturity (*Developer Category*) is derived by classifying all developers presented in the dataset into two categories: Newbie (the one with less than 4 games in the list) and Professional (the one with 4 or more games in the list). Following this, *Number of IAP*, *Minimum IAP*, *Maximum IAP*, *Sum IAP* and *Mean of IAP*, *Price*, and *Game Size* are extracted as well.

3.2 Success Score Measure

In this work, we propose a new success measure that encodes the three famous traditional success measures, namely, gross revenue, number of downloads, and rating.

The suggested success measure shown in Equation (1) gives importance to the following three objectives: Revenue, Reputation, and Success Speed.

Revenue It represents the expected income by this game. It is calculated by multiplying the price of the game with the expected number of people who will buy the game. Then, the expected income via IAPs is added as well.

Table 2: The main attributes of the categorized dataset.

Attribute	Categories	Attribute	Categories
Num. Languages	Single, Many	Release Month	1,2,3,4,5,6,7,8,9,10,11,12
Num. Genres	Less than 4, Equal or More than 4	Elapsed Months	Less than 17, [17, 31], [32, 45], [46, 64], More than 65
Age Rating	+4, +9	Developer Category	Newbie, Professional
Availability IAP	No, Yes	Game Size	S1, S2, S3, S4, S5
Average IAP	No, Some, High	Game Free	No, Yes
Min IAP	0$, ˜1$, Equal or More than 1.5$	Rating Average	Low, Moderate, High
Max IAP	0$, Less than 2.5$, Equal or More than 2.5$	Success Score	Successful, Undetermined, Unsuccessful
Sum IAP	No, Some, High	Rating Count	Low, Moderate, High

Table 3: The strongest 10 rules associated with high success scores.

| rules	support	confidence	lift	count
Maximum IAPs is less than 2.5$ but not free, game's developer is professional, it is a puzzle game, it is a travel game.	0.001	1.000	3.000	18
Sum of IAPs is between 0.49$ and 3.49$, game's developer is professional, it is a puzzle game, it is a travel game.	0.001	1.000	3.000	20
Average of IAPs is between 0.28$ and 2.1$, game's developer is professional, it is a puzzle game, it is a travel game.	0.001	1.000	3.000	18
The minimum IAPs is more than 1.5$, the game's developer is professional, it is a puzzle game, it is a travel game.	0.001	1.000	3.000	17
There is IAPs, the game's developer is professional, it is a puzzle game, it is a travel game.	0.001	1.000	3.000	20
Game supports many languages, game size is between 164MB and 4GB, game is premium, it is not a board game.	0.042	0.800	2.401	713
Game supports many languages, there is IAPs, game size is between 164MB and 4GB.	0.039	0.811	2.432	655
Game supports many languages, there is IAPs, game size is between 164MB and 4GB, game is premium.	0.039	0.811	2.432	655
Game supports many languages, there is IAPs, game size is between 164MB and 4GB, it is not a utilities game.	0.039	0.814	2.441	655
Game supports many languages, there is IAPs, game size is between 164MB and 4GB, it is not a travel game.	0.039	0.811	2.432	655
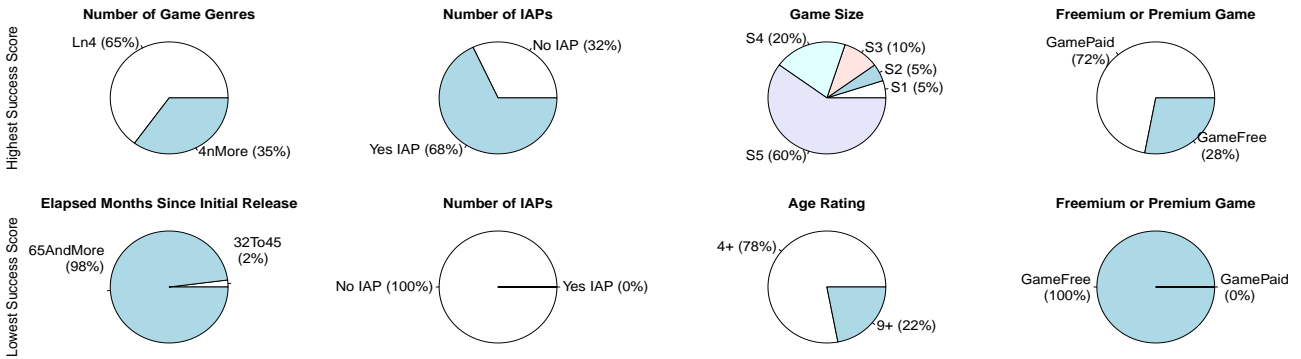
Fig. 1: Attributes distribution in the best and worst 100 games in terms of the success score.

Reputation It is represented by the average rating of the game. It is in the interval 0 to 5 and it was rounded to nearest 0.5 average rating.

Success Speed It is encoded in the success formula as well. A game to keep constant success score should

increase either revenue or rating average (or both) as the game becomes older.

$$S = \frac{(RC * G * P + RC * K * AP)_N + RA_N}{T_N + \epsilon} \qquad (1)$$

Table 4: The strongest 10 rules associated with low success scores.

rules	support	confidence	lift	count
32 to 45 months since release, game is free, not an adventure game, a lifestyle game.	0.001	1.000	3.000	21
32 to 45 months since release, game is free, size is between 20 and 40 MB, it is a trivia game.	0.001	1.000	3.000	25
The game was released in July, it is free, not an education game, a family game.	0.001	1.000	3.000	18
32 to 45 months since release, game is free, its size is less than 20MB, a family game.	0.001	1.000	3.000	23
Less than 17 months since release, number of genres is less than 4, game is free, an adventure game.	0.002	1.000	3.000	35
Less than 17 months since release, game is free.	0.077	0.804	2.412	1308
Less than 17 months since release, game is free, no In-App-Purchases.	0.077	0.804	2.412	1308
Less than 17 months since release, game is free, average In-App-Purchases is 0.	0.077	0.804	2.412	1308
Less than 17 months since release, game is free, sum of In-App-Purchases is 0.	0.077	0.804	2.412	1308
Less than 17 months since release, game is free, max In-App-Purchases is 0$.	0.077	0.804	2.412	1308

where, S is the success score, RC denotes rating count, RA denotes rating average, P is the initial download and installation price, AP stands for average IAPs, T denotes elapsed time since first release, K is the expected number of users who will buy IAP per rating count, G is the expected number of users who will buy the game per rating count, $(.)_N$ stands for the operation of normalizing the operand to [0, 1] interval. ϵ is a very small number to avoid division by zero.

For a game to score high on this measure, it should maximize the revenue and rating average. At the same time, it should do that in a short period of time. It is important to note that K and G are estimated variables that can be changed to match the targeted audience. Parameters K and G were guesstimated as 0.05 and 0.5, respectively.

3.3 Predictive Models

In the process of conducting our statistical analysis, four different models were applied on the processed version of the initial dataset. Under the scope of understating the correlations between success and mobile game's attributes, the Apriori algorithm [1,32] was utilized based on its implementations in [7,6]. At the same time, we ranked the games in terms of success score, rating average, and rating count. Then, the attributes distributions in the best and worst 100 games were analysed. On the other hand, for the predictive part, Support Vector Machines [9], Random Forest [8], and Artificial Neural Network models were used for predicting the success of a mobile game given its metadata.

4 Analysis and Findings

In this section, the rules associated with the success and failure of mobile games are studied in Sec. 4.1. Then, the process of training machine learning models and predicting game success are discussed in Sec. 4.2.

4.1 Wisdom Extraction

One main goal of this work is to show the link between the success and the various game development attributes. For that goal in mind, Apriori algorithm was applied for mining possible associations. The support parameter was set to 0.001 and the confidence one to 0.8. Following this, the highest 5 rules were picked in terms of the confidence and support. Here, our main findings are highlighted in terms of the proposed success measure, the average rating, and the rating count. **Success Score** Most important rules with respect to our novel success score are given in Tables 3 and 4. Some of the interesting findings appear to be being a long-time developer, availability of cheap IAPs, and publishing Puzzle and Travel games help with success. In parallel to that, making your game available in many languages can increase the popularity and revenue. Thus, the probability of success. Surprisingly, releasing a game in July is expected to decrease its chance of achieving a success. The reason could be that the US and Europe start their holidays at that time so people do not prefer playing video games for fun.

In addition to the previous general observations, specifically, we study the 100 games with the highest and lowest success score. As shown in Fig. 1, in a successful game the number of game genres is relatively less, IAPs are found, and the game size is larger.

Table 5: The utilized machine learning models and their corresponding parameters.

| Model	Full Name	Success Score	Rating Average	Rating Count
Apriori	Mining Associations with Apriori	Support = 0.001, Confidence = 0.8	Support = 0.001, Confidence = 0.8	Support = 0.001, Confidence = 0.8
SVM	Support Vector Machine	Kernel = radial, Cost=19.5	Kernel = radial, Cost=83	Kernel = radial, Cost = 83
RF	Random Forest	ntree = 130, importance = TRUE	ntree = 130, mtry = 10	ntree = 130, mtry = 10
ANN	Artificial Neural Network	Hidden Layers = [70, 50, 20, 10, 4] Number of Epochs = 35 L1 = 0.001, L2 = 0.01 Activation Function = Maxout Adaptive Rate = TRUE	Hidden Layers = [50, 100] Number of Epochs = 35 L1 = 0.001, L2 = 0.01 Activation Function = Maxout Adaptive Rate = TRUE	Hidden Layers = [70, 50, 20, 10, 4] Number of Epochs = 35 L1 = 0.001, L2 = 0.01 Activation Function = Maxout Adaptive Rate = TRUE
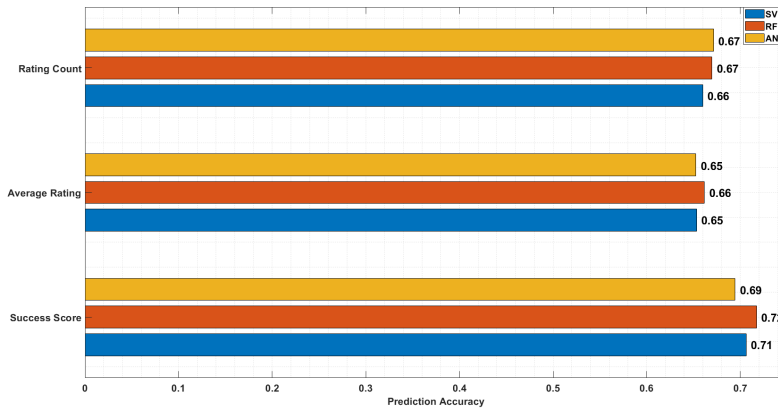
Fig. 2: The prediction accuracy of game success using SVM, RF, and ANN models.

	Prediction		
	Succ.	Undet.	Fail
Succ.	**285**	142	139
Undet.	119	**435**	10
Fail	13	61	**495**

(a) Success Score

	Prediction		
	Low	Medium	High
Low	**778**	35	132
Medium	83	**30**	101
High	194	18	**328**

(b) Rating Average

	Prediction		
	Low	Medium	High
Low	**863**	32	85
Medium	202	**21**	81
High	120	20	**275**

(c) Rating Count

Table 6: Confusion matrices for the best model (RF) for (a) Success Score, (b) Rating Average and (c) Rating Count.
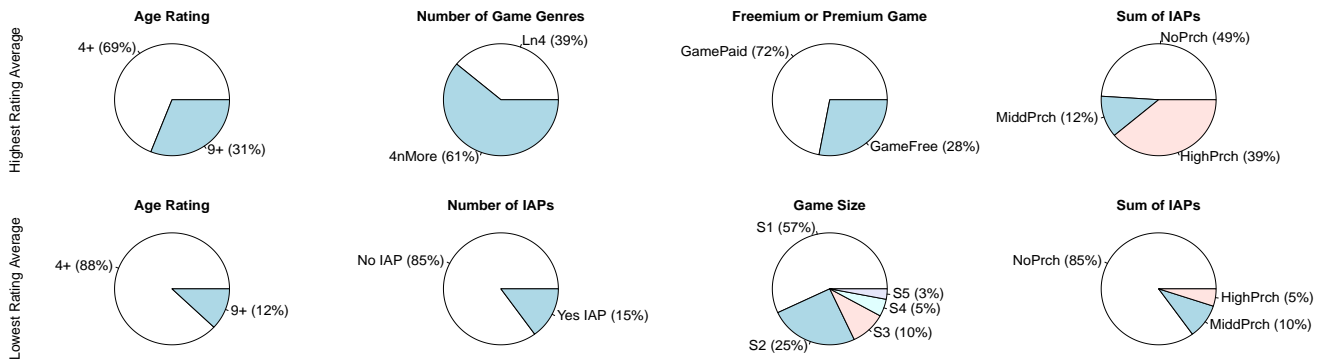

Fig. 3: Attributes distribution in the best and worst 100 games in terms of the rating average.

Average Rating The rules associated to average rating are shown in Tables 7 and 8. At the same time, observations from Fig. 3 suggest game developers who want to achieve higher rating average to think of avoid-

Table 7: The strongest 10 rules associated with high rating average.

rules	support	confidence	lift	count
Maximum IAPs price is less 2.5$ but not free, game developer is professional, it is a puzzle game, it is a travel game.	0.001	1.000	3.047	18
Sum IAPs prices is between 0.49$ and 3.49$, game developer is professional, it is a puzzle game, it is a travel game.	0.001	1.000	3.047	20
Average IAPs prices is between 0.28$ and 2.09$, game developer is professional, it is a puzzle game, it is a travel game.	0.001	1.000	3.047	18
Minimum IAPs price is more than 1.5$, game developer is professional, it is a puzzle game, it is a travel game.	0.001	1.000	3.047	17
There is IAPs, game developer is professional, it is a puzzle game, it is a travel game.	0.001	1.000	3.047	20
Game supports many languages, age rating is +9, there is IAPs, game size is between 164MB and 4GB.	0.025	0.804	2.449	430
Game supports many languages, age rating is +9, sum of IAPs prices is between 3.5$ and 675$, game size is between 164MB and 4GB.	0.025	0.810	2.468	422
Game supports many languages, age rating is +9, maximum of IAPs prices is more than 2.5$, game size is between 164MB and 4GB.	0.025	0.809	2.465	419
Game supports many languages, age rating is +9, average of IAPs prices is between 2.09 and 89.99, game size is between 164MB and 4GB.	0.024	0.808	2.462	412
Game supports many languages, minimum IAPs is 1$, game size is between 164MB and 4GB.	0.024	0.805	2.453	409

Table 8: The strongest 10 rules associated with low rating average.

| rules	support	confidence	lift	count
Age rating is +4, it is a simulation game, it is a travel game.	0.001	1.000	1.802	20
Game size is between 40MB and 81MB, 17 to 31 months since release, it is a travel game.	0.001	1.000	1.802	18
Game size is between 20MB and 40MB, 32 to 45 months since release, it is a trivia game.	0.002	1.000	1.802	38
Game size is less than 20MB, less than 17 months, it is a family game.	0.001	1.000	1.802	23
Game is released in March, 17 to 31 months since release, it is an education game.	0.002	1.000	1.802	27
Minimum IAPs price is 0$, 32 to 45 months since release, it is not a casino game.	0.100	0.803	1.447	1691
Minimum IAPs price is 0$, 32 to 45 months since release, it is neither a casino nor a finance game.	0.100	0.803	1.448	1690
Maximum IAPs price is 0$, 32 to 45 months since release, it is not a casino game.	0.099	0.802	1.446	1688
Sum of IAPs price is 0$, 32 to 45 months since release, it is not a casino game.	0.099	0.802	1.446	1688
Minimum IAPs price is 0$, 32 to 45 months since release, it is not a casino game.	0.099	0.802	1.446	1688
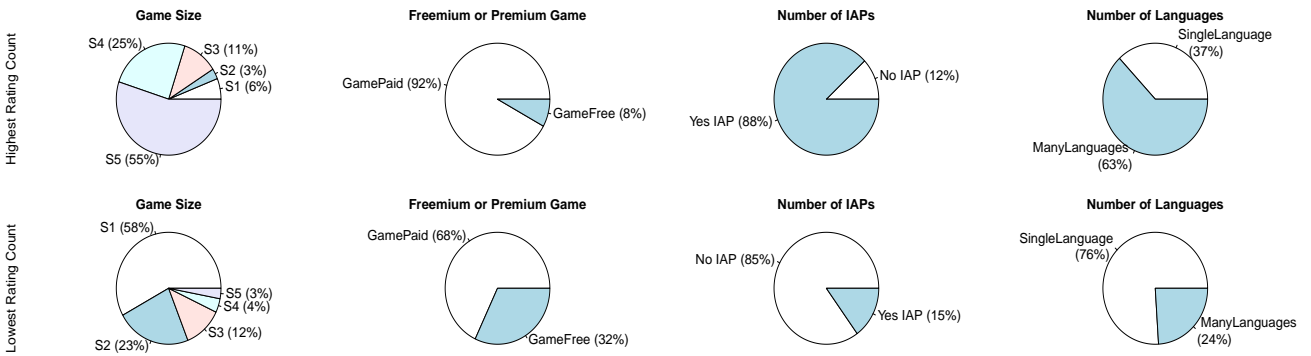
Fig. 4: Attributes distribution in the best and worst 100 games in terms of the rating count.

ing both small size and free games. At the same time, they should aim at including IAPs and focusing on +9 audience.

Rating Count Most important rules with respect to rating count shown in Tables 9 and 10 indicate that if the goal of the game is to attract high audience number, the developer should think of making the game

Table 9: The strongest 10 rules associated with high rating count.

rules	support	confidence	lift	count
Game size is between 164MB and 4GB, more than 65 months since release, it is board game.	0.002	1.000	4.011	30
Game supports many languages, it is released in October, its size between 164MB and 4GB, more than 65 months since release.	0.001	1.000	4.011	17
Average IAPs prices is between 2.09$ and 89.99$, its size is between 164MB and 4GB, more than 65 months since release, it is a board game.	0.001	1.000	4.011	20
Sum of IAPs prices is between 3.49$ and 674.9$, its size is between 164MB and 4GB, more than 65 months since release, it is a board game.	0.001	1.000	4.011	23
Maximum IAPs prices is more than 2.5$, its size is between 164MB and 4GB, more than 65 months since release, it is a board game.	0.001	1.000	4.011	23
Age rating is +9, sum of IAPs prices is between 3.49$ and 674.9$, more than 65 months since release, it is not a music game.	0.023	0.801	3.214	395
Age rating is +9, sum of IAPs prices is between 3.49$ and 674.9$, more than 65 months since release, it is not a puzzle game.	0.022	0.805	3.228	379
Age rating is +9, sum of IAPs prices is between 3.49$ and 674.9$, more than 65 months since release, it is not a casual game.	0.022	0.800	3.211	377
Age rating is +9, maximum IAPs prices is more than 2.5$, more than 65 months since release, it is not a puzzle game.	0.021	0.801	3.214	359
Age rating is +9, average of IAPs prices is between 2.09$ and 89.99$, more than 65 months since release.	0.021	0.803	3.220	350

Table 10: The strongest 10 rules associated with low rating count.

rules	support	confidence	lift	count
Rating average +4, it is a simulation game, it is a travel game.	0.001	1.000	1.731	20
Game size is between 40MB and 81MB, 17 to 31 months since release, it is a travel game.	0.001	1.000	1.731	18
Game size is between 20MB and 40MB, 32 to 45 months since release, it is a trivia game.	0.002	1.000	1.731	38
Game size is less than 20MB, less than 17 months since release, it is a family game.	0.001	1.000	1.731	23
Game released in March, 17 to 31 months since release, it is an education game.	0.002	1.000	1.731	27
Minimum IAPs is 0$, 32 to 45 months since release.	0.102	0.813	1.408	1729
Minimum IAPs is 0$, 32 to 45 months since release, it is not a finance game.	0.102	0.813	1.408	1727
Maximum IAPs is 0$, 32 to 45 months since release.	0.102	0.813	1.407	1726
Sum of IAPs is 0$, 32 to 45 months since release.	0.102	0.813	1.407	1726
Average IAPs is 0$, 32 to 45 months since release.	0.102	0.813	1.407	1726

available in many languages and targeting +9 audience. In parallel to that, Fig. 4 reflects similar observations for other different features. We can see that making a game free does not necessarily make it popular. However, the opposite was shown to happen. Thus, the common sense that making a game free will attract a wider audience does not seem to be valid as our analysis revealed clearly.

4.2 Prediction

The second major aim of this work is to predict the success of a mobile game. Firstly, the dataset is partitioned into train, validation, and test splits by taking 80% for training, 10% for validation, and 10% for testing. Following this, SVM (Support Vector Machines), RF (Random Forest), and ANN models were trained. The parameters of the associated models are shown in

Table 5. All the models obtained around 70% classification accuracy as shown in Fig. 2. Although this level of accuracy may seem to be low, it is in fact a significant achievement considering that we are only using meta data about the game to make this prediction and the model is completely blind to the actual game. In other words, the actual quality and playability of the game is completely ignored in this prediction. From this perspective, 70% accuracy is actually quite significant and shows that certain game meta data is directly associated with the success of a game.

It can be observed that predicting the proposed success score is more accurate as compared to predicting rating average and rating count. That could be the case because our success measure is closely associated with the game development features much more than the other success criteria. In other words, rating average and rating count might be influenced by other factors that are not included in the main studied features. In

parallel to that, it is clear that the best model among the three predictive models is the RF model for all the three success measures. Thus, it is worth studying the true and false positives and negatives of this model. For the success score prediction, it sounds that the model does the best at predicting the unsuccessful games. While the same observations were shown for the low rating and low number of ratings as shown in Table 6. In fact, this observation is critical because it sheds light on two points. First, there are some features and decisions that lead to game failure whether the success was in terms of rating average, rating count, or the general proposed success measure. Thus, the developer should avoid these features and decisions as it sounds that the chance of their success is limited. Oppositely, the successful games are harder to be linked to specific attributes like unsuccessful ones. It implies that there is a clear path for your game to fail. However, the success path is dependent on more sophisticated factors like creativity and novelty which are not directly measured in this study.

To further enrich the analysis, the mean decrease in the Gini index is studied since it highlights the most important features RF employs at predicting the target variable. Fig. 5-a shows the critical features in a decreasing order with respect to their importance for the success score prediction. It is clear that the developers should be very careful as they decide if their game will be premium or freemium. They should consider other factors such as game release month, game size, and the IAPs related decision very wisely. In terms of the rating average, it is shown in Fig. 5-b that original release month, game size, number of genres, and the game developer reputation are the most important features that influence the rating of the game. Consequently, if developers would like to gain a high rating average, they should consider these features with great attention. Similarly, Fig. 5-c indicates that features like game size, IAPs, developer reputation and number of languages are among the most vital features that control number of ratings. Thus, a developer aiming at a high rating count should think of these features carefully.

5 Icon Design Analysis

The icons of the top best and worst 100 games in terms of success score, rating average, and rating count were considered in this analysis. In Sec. 5.1 the icon's entropy was studied for these games. Additionally, in Sec. 5.2 the white and black pixels percentages of icons were studied for the same subset of mobile games.

5.1 Icons Entropy

Entropy in general measures uncertainty or randomness. However, in image processing, it can be applied to give statistical information on the texture of the input image [14]. The used entropy formula is shown in 2.

$$E = -\sum_{i=1}^{255} p_i \times \log_2(p_i) \tag{2}$$

where: E is the icon entropy, p is the normalized histogram counts of the icon image.

We calculated the entropy for the icons in three different color spaces i.e. RGB, HSV, and L*a*b*. However, a clear difference was shown in the HSV one. This could be the reason because HSV color space represents colors in a way that is much more similar to the human eye color perception. The results shown in Fig. 6 reveal that the entropy was higher in all the three success measures. However, it was more explicit in the rating count. Clearly, the complexity of the game icon has a strong correlation with both the attractiveness of the game in the market and the success of the game in general. This means higher entropy icons are more eye-catching and hints at the extra effort spent in preparing and publishing the game.

A qualitative comparison among the icons with the lowest and highest rating count is presented in Fig. 7. The contrast is very clear: the icons with high entropy seem to attract more players, whereas low entropy icons are much less attractive. Thus, game designers are suggested to invest more effort and time in making their icons rich with visual details. On the other hand, they should avoid abstract and simplistic icons.

5.2 Percentage of Black and White Pixels in Icons

To further analyze the common features in successful icons, we study the percentage of white and black pixels in icons. Our analysis shows that icons of mobile games with low rating counts tend to be occupied with white and black pixels much more than the ones with high rating counts as shown in Fig. 8. This confirms our previous findings that game developers should avoid simplistic icons. At the same time, it adds that developers should utilize the space of the icon carefully and should not leave many empty pixels.

Density of the white and the black pixels in icons is not related to rating average as shown in Fig. 8. Thus, it is expected that the icon design does not affect the rating average. As shown in Fig 7, we can find many icons in the lowest rating count games that contain empty spaces (white or black). On the other hand, we can
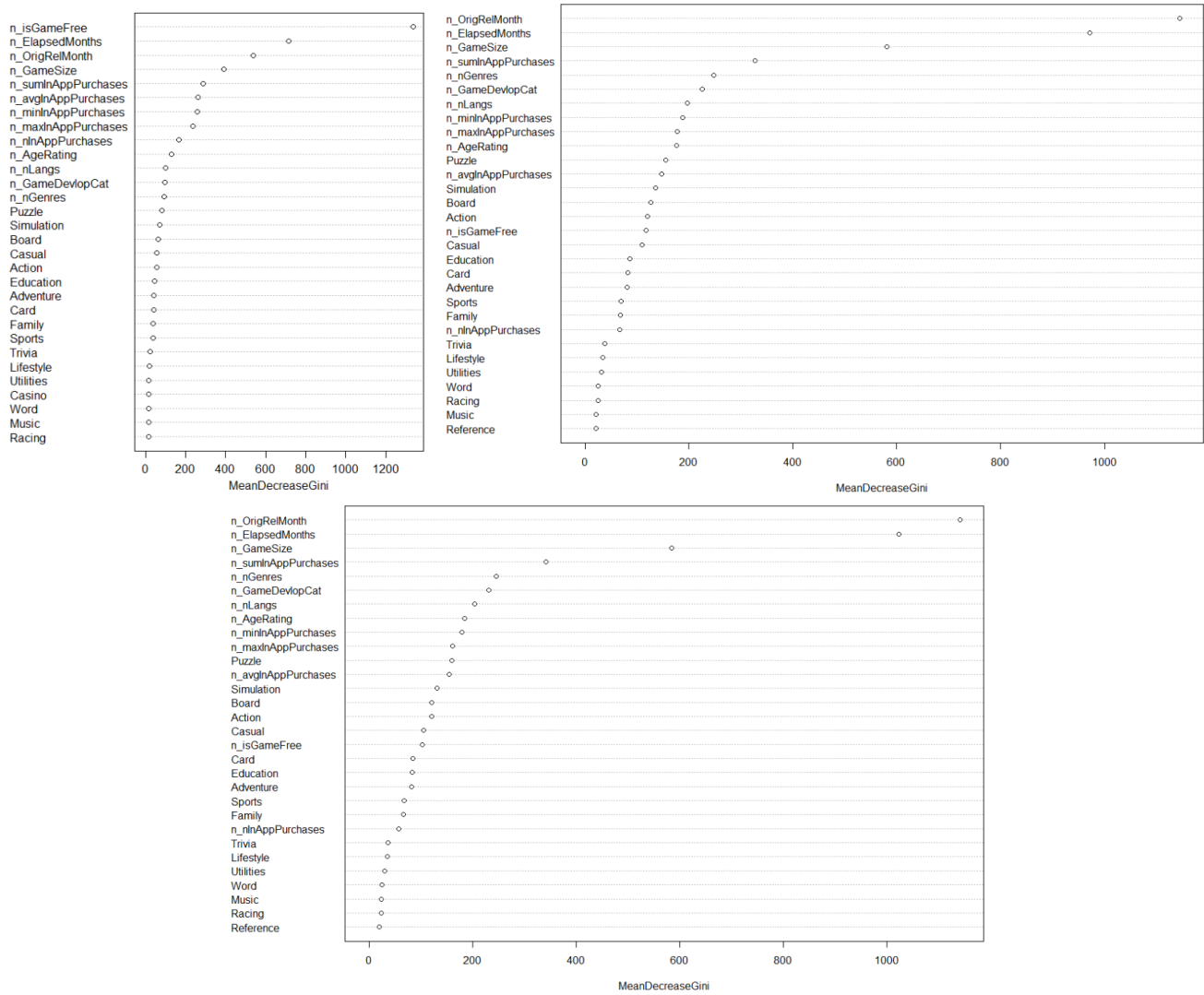
Fig. 5: Mean decrease in Gini index for (a) first row, left: success score, (b) first row, right: rating average and (c) second row: rating count.

6 Discussion

We can see from the results that predicting the success of a game using our new measure gives a better accuracy. We should note that this score depends on rating count and rating average, however, these two attributes are not included in training and prediction phases. At the same time, our measure depends on the time elapsed but that does not only control the success score. Our work has a number of limitations that should be addressed. Despite using a large-scale dataset for analysing the success of mobile games, we should note that all the games were taken from a single mobile App Store (Apple App Store). Thus, the results would be more generalizable if the same conclusions could be reached using mobile games from different application stores. Our proposed success metric was shown to give the models slightly better accuracy at predicting the success of a game. The single most striking result is the correlation between the release month (July) and the high failure expectation. The release month attribute is evenly distributed across all months. Thus, the possibility of unbalance in the attribute distribution causing this association is not expected to be the playing factor. Our analysis shows the correlation. However, it does not investigate the causality. The causality is investigated and shown based on the authors' understanding. Thus, we present a note of caution with regard to interpre-
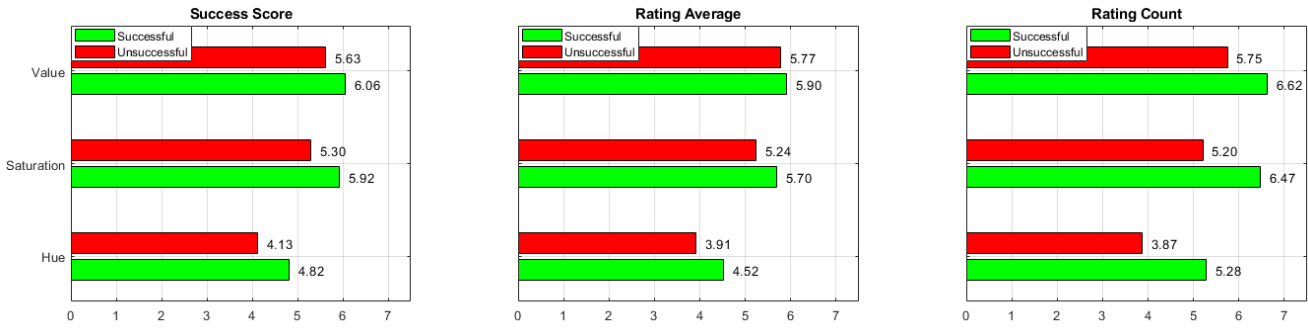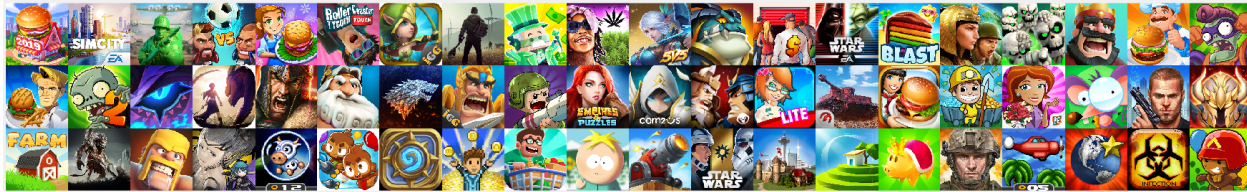
Fig. 6: Entropy of 100 most successful and unsuccessful mobile games icons in HSV color space is shown for Success Score, Rating Average, and Rating Count.

(a) Sample icons of games with lowest rating count.

(b) Sample icons of games with highest rating count.

Fig. 7: Sample of icons from the top 100 games with lowest (a) and highest (b) rating count.
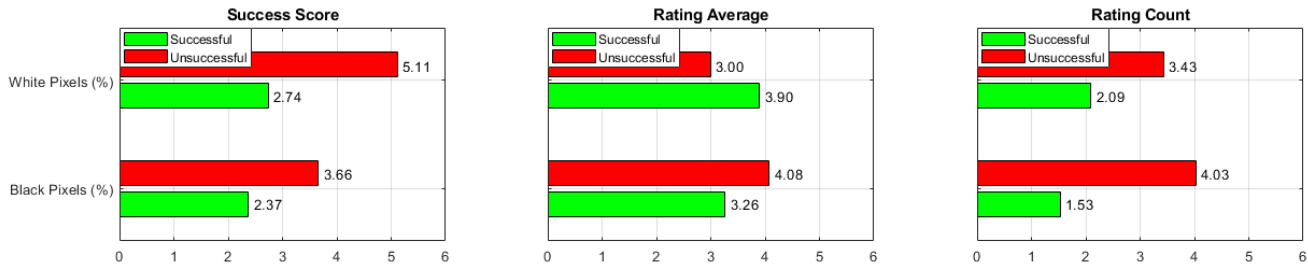

Fig. 8: Percentage of white and black pixels in top 100 successful and unsuccessful games.

tations of the causes of these correlations. They could be biased, wrong, or subjective to the authors of this work.

Predicting the success of a game with around 70% accuracy without considering the actual content of the game itself is an important point to be highlighted. It shows that certain metadata about a game, which are usually mostly neglected during the publishing of the game, can actually be quite effective in the success of the release. Of course, the content of the game, graphics, story, sound effects, etc. are all major factors in the

success of a game. However, we showed that even with an expensive production phase maximizing these features of the game, a release can go wrong simply due to a wrong release month, or a bland icon design.

We would like to note that integrating icon visual information in the prediction process did improve the prediction accuracy, however, slightly. Table 11 shows the three predictive models and their associated prediction accuracy for four different setups. Columns three through six present models prediction accuracy under these four setups namely, without icon visual informa-

Table 11: Effect of icon visual information on game success prediction accuracy.

		w/o Icons Info	w/ Entropy	w/ Black-White Pixels Percentage	w/ Entropy and Black-White Pixels Percentage
SVM	Success Score	0.7057	0.7063	0.7039	**0.7087**
	Rating Average	0.6530	0.6533	**0.6669**	0.6563
	Rating Count	0.6590	0.6598	**0.6733**	0.6586
RF	Success Score	0.7175	**0.7181**	0.7122	0.7092
	Rating Average	0.6616	0.6627	**0.6645**	0.6610
	Rating Count	0.6698	0.6757	0.6669	**0.6775**
ANN	Success Score	0.6945	**0.6951**	0.6910	0.6922
	Rating Average	0.6527	**0.6545**	0.6451	0.6527
	Rating Count	0.6716	0.6669	**0.6763**	0.6651

tion, with icon entropy information, with white-black pixels percentages in icon, and with both entropy and white-black pixels percentages of icon, respectively.

7 Conclusion

In this work, we presented some facts regarding the relation between the success of a mobile game and the various game development process attributes. The success was given by a novel formula that encodes the general goal of game development i.e. revenue, popularity, and reputation. We were able to show that some attributes can impact the success positively or negatively. We show that specific game attributes, such as number of IAPs, belonging to the puzzle genre, supporting different languages, and being produced by a mature developer highly and positively affect the success of the game in the future. Moreover, we show that releasing the game in July and not including any IAPs seems to be highly associated with the game's failure. Furthermore, we show that game icons with high entropy values and less empty spaces tend to be positively correlated with high rating counts. At the same time, we were able to predict the success of a game with a good accuracy given only its external characteristics. Thus, it is critical for game developers to consider the recommendations revealed by our statistical study to avoid failures because of simple decision options.

As a future work, we are planning to conduct a semantic analysis on the game description. At the same time, performing a deeper study on the contents of icons and their correlation with number of downloads is another research direction to be considered as well.

Declarations

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

References

1. Agrawal, R.: I mielinski t and swami a. Mining association rules between sets of items in large databases **993** (1993)
2. Aldino, A.A., Pratiwi, E.D., Sintaro, S., Putra, A.D., et al.: Comparison of market basket analysis to determine consumer purchasing patterns using fp-growth and apriori algorithm. In: 2021 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), pp. 29–34. IEEE (2021)
3. Alomari, K.M., Soomro, T.R., Shaalan, K.: Mobile gaming trends and revenue models. In: International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, pp. 671–683. Springer (2016)
4. Belgiu, M., Drăguţ, L.: Random forest in remote sensing: A review of applications and future directions. ISPRS journal of photogrammetry and remote sensing **114**, 24–31 (2016)
5. Bennett, K.P., Campbell, C.: Support vector machines: hype or hallelujah? ACM SIGKDD explorations newsletter **2**(2), 1–13 (2000)
6. Borgelt, C.: Efficient implementations of apriori and eclat. In: FIMI'03: Proceedings of the IEEE ICDM workshop on frequent itemset mining implementations (2003)
7. Borgelt, C., Kruse, R.: Induction of association rules: Apriori implementation. In: Compstat, pp. 395–400. Springer (2002)
8. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
9. Chang, C.C., Lin, C.J.: Libsvm: A library for support vector machines. ACM transactions on intelligent systems and technology (TIST) **2**(3), 1–27 (2011)
10. Cortes, C., Vapnik, V.: Support-vector networks. Machine learning **20**(3), 273–297 (1995)
11. Dehkordi, M.R., Seifzadeh, H., Beydoun, G., Nadimi-Shahraki, M.H.: Success prediction of android applications in a novel repository using neural networks. Complex & Intelligent Systems **6**, 573–590 (2020)
12. Doan, T.N., Poulet, F.: Large scale image classification: Fast feature extraction, multi-codebook approach and multi-core svm training. In: Advances in Knowledge Discovery and Management, pp. 155–172. Springer (2014)
13. Drachen, A., Lundquist, E.T., Kung, Y., Rao, P., Sifa, R., Runge, J., Klabjan, D.: Rapid prediction of player retention in free-to-play mobile games. In: Twelfth artificial intelligence and interactive digital entertainment conference (2016)
14. Eddins, S.L., Gonzalez, R.C., Woods, R.E.: Digital image processing using matlab. Princeton Hall Pearson Education Inc., New Jersey (2004)

15. Ghayas, S., Sulaiman, S., Khan, M., Jaafar, J.: The effects of icon characteristics on users' perception. In: International Visual Informatics Conference, pp. 652–663. Springer (2013)

16. Gorunescu, F.: Data Mining: Concepts, models and techniques, vol. 12. Springer Science & Business Media (2011)

17. Gough, C.: Apple app store: number of available gaming apps as of q1 2020. `https://www.statista.com/statistics/780238/number-of-available-gaming-apps-in-the-apple-app-store-quarter/` (2020). Online; accessed 06 August 2020

18. Gough, C.: Video game industry - statistics and facts. `https://www.statista.com/topics/868/video-games/` (2020)

19. Gupta, G.K.: Introduction to data mining with case studies. PHI Learning Pvt. Ltd. (2014)

20. Gurjar, M., Naik, P., Mujumdar, G., Vaidya, T.: Stock market prediction using ann. International Research Journal of Engineering and Technology **5**(3), 2758–2761 (2018)

21. Ho, T.K.: Random decision forests. In: Proceedings of 3rd international conference on document analysis and recognition, vol. 1, pp. 278–282. IEEE (1995)

22. Jylhä, H., Hamari, J.: An icon that everyone wants to click: How perceived aesthetic qualities predict app icon successfulness. International Journal of Human-Computer Studies **130**, 73–85 (2019)

23. Kamarulzaman, N.A., Fabil, N., Zaki, Z.M., Ismail, R.: Comparative study of icon design for mobile application. In: Journal of Physics: Conference Series, vol. 1551, p. 012007. IOP Publishing (2020)

24. Kira, K., Rendell, L.A., et al.: The feature selection problem: Traditional methods and a new algorithm. In: Aaai, vol. 2, pp. 129–134 (1992)

25. Kukka, H., Oja, H., Kostakos, V., Gonçalves, J., Ojala, T.: What makes you click: exploring visual signals to entice interaction on public displays. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1699–1708 (2013)

26. Lee, G., Raghu, T.: Product portfolio and mobile apps success: Evidence from app store market. In: AMCIS (2011)

27. Lee, G., Raghu, T.S.: Determinants of mobile apps' success: Evidence from the app store market. Journal of Management Information Systems **31**(2), 133–170 (2014)

28. Lin, Y., Lv, F., Zhu, S., Yang, M., Cour, T., Yu, K., Cao, L., Huang, T.: Large-scale image classification: fast feature extraction and svm training. In: CVPR 2011, pp. 1689–1696. IEEE (2011)

29. McDougall, S., Reppa, I., Kulik, J., Taylor, A.: What makes icons appealing? the role of processing fluency in predicting icon appeal in different task contexts. Applied ergonomics **55**, 156–172 (2016)

30. Mining, D.: Concepts and techniques. Jiawei Han and Micheline Kamber **2** (2001)

31. Moreira, Á., Vicente Filho, V., Ramalho, G.L.: Understanding mobile game success: a study of features related to acquisition, retention and monetization. SBC **5**(2) (2014)

32. Muino, D.P., Borgelt, C.: Frequent item set mining for sequential data: Synchrony in neuronal spike trains. Intelligent Data Analysis **18**(6), 997–1012 (2014)

33. Muthukumar, A., Kavipriya, A.: A biometric system based on gabor feature extraction with svm classifier for finger-knuckle-print. Pattern Recognition Letters **125**, 150–156 (2019)

34. Purwandari, K., Sigalingging, J.W., Cenggoro, T.W., Pardamean, B.: Multi-class weather forecasting from twitter using machine learning aprroaches. Procedia Computer Science **179**, 47–54 (2021)

35. Qi, Y.: Random forest for bioinformatics. In: Ensemble machine learning, pp. 307–323. Springer (2012)

36. Sahu, B., Palo, H.K., Shrotriya, S.: Speech emotion recognition using k-means apriori feature selection algorithm. In: 2021 International Conference in Advances in Power, Signal, and Information Technology (APSIT), pp. 1–4. IEEE (2021)

37. Shen, Z., Zhang, L., Li, R., Liang, R.: The effects of icon internal characteristics on complex cognition. International Journal of Industrial Ergonomics **79**, 102990 (2020)

38. Sifa, R., Hadiji, F., Runge, J., Drachen, A., Kersting, K., Bauckhage, C.: Predicting purchase decisions in mobile free-to-play games. In: Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, vol. 11 (2015)

39. Sumathi, S., Sivanandam, S.: Introduction to data mining and its applications, vol. 29. Springer (2006)

40. Thamaraimanalan, T., RA, L., RM, K., et al.: Multi biometric authentication using svm and ann classifiers. Irish Interdisciplinary Journal of Science & Research (IIJSR) (2021)

41. Tristan: 17k mobile strategy games. `https://www.kaggle.com/tristan581/17k-apple-app-store-strategy-games` (2020)

42. Ünvan, Y.A.: Market basket analysis with association rules. Communications in Statistics-Theory and Methods **50**(7), 1615–1628 (2021)

43. Wang, J., Song, L., Li, Z., Sun, H., Sun, J., Zheng, N.: End-to-end object detection with fully convolutional network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15849–15858 (2021)

44. Yang, R., Yu, Y.: Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis. Frontiers in Oncology **11**, 573 (2021)

45. Yi, J., Lee, Y., Kim, S.H.: Determinants of growth and decline in mobile game diffusion. Journal of Business Research **99**, 363–372 (2019)

46. Zou, Z.B., Peng, H., Luo, L.K.: The application of random forest in finance. In: Applied Mechanics and Materials, vol. 740, pp. 947–951. Trans Tech Publ (2015)