RELAÇÕES INTERNACIONAIS NO MUNDO ATUAL

CENTRO UNIVERSITÁRIO CURITIBA - UNICURITIBA - VOLUME 3 - NÚMERO 36/2022 CURITIBA/PARANÁ/BRASIL - PÁGINAS 671 A 691- ISSN: 2316-2880

LEGAL ANALYSIS OF ENTERTAINMENT APPS IN APPLE APPSTORE USING BIG DATA

Syifa Hanifa Salsabil
School of Economics and Business, Telkom University
https://orcid.org/0000-0002-8872-2651
g.publication13@yahoo.com

Gadang Ramantoko
School of Economics and Business, Telkom University
https://orcid.org/0000-0001-5517-5834
g.publication13@yahoo.com

ABSTRACT

This research aims to analyze the competition and legal issues among entertainment apps in Apple AppStore. The samples of this study were 2400 apps that were included in the entertainment category. Big data analytics is employed in this study to analyze the competition among entertainment apps and measured by user's minimum age, price, in-app purchase option, rate, and a number of raters; these factors are processed using hierarchical clustering. Another factor is the description of the apps extracted to keywords and processed using LDA topic modeling and text network to analyze the competition based on the offering products/contents. Hierarchical clustering resulted 7 clusters. The user's minimum age that the apps' developers commonly state is four years old. The apps are majority priced at IDR.0 or free, and the highest is IDR3.299.000. The price makes it the only app in cluster 4 because there are no other apps that have a price close to it. Cluster 5 contains free apps with good performance measured by good ratings, many raters, and an in-app purchase option. Ten topics resulted from LDA topic modeling that was further visualized by the text network. Text network shows that the words game, app, video, and image are the most common words that occur in most topics; these also have a high degree centrality and betweenness centrality.

Keywords: Big Data Analytics, Competition, Entertainment Apps, Hierarchical Clustering, LDA, legal issues.

[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

ANÁLISE JURÍDICA DE APPS DE ENTRETENIMENTO NA APP STORE DA APPLE USANDO BIG DATA

ABSTRATO

Esta pesquisa tem como objetivo analisar as questões concorrenciais e jurídicas entre aplicativos de entretenimento na Apple AppStore. As amostras deste estudo foram 2.400 aplicativos que foram incluídos na categoria entretenimento. A análise de big data é empregada neste estudo para analisar a concorrência entre aplicativos de entretenimento e medida pela idade mínima do usuário, preço, opção de compra no aplicativo, taxa e número de avaliadores; esses fatores são processados usando agrupamento hierárquico. Outro fator é a descrição dos aplicativos extraídos para palavras-chave e processados usando modelagem de tópicos LDA e rede de texto para analisar a concorrência com base nos produtos/conteúdos oferecidos. O agrupamento hierárquico resultou em 7 agrupamentos. A idade mínima do usuário que os desenvolvedores dos aplicativos geralmente declaram é de quatro anos. Os aplicativos têm preços majoritários em IDR.0 ou gratuitos, e o mais alto é IDR3.299.000. O preço o torna o único aplicativo do cluster 4, pois não há outros aplicativos com preço próximo a ele. O Cluster 5 contém aplicativos gratuitos com bom desempenho medido por boas classificações, muitos avaliadores e uma opção de compra no aplicativo. Dez tópicos resultaram da modelagem de tópicos LDA que foi posteriormente visualizado pela rede de texto. A rede de texto mostra que as palavras jogo, aplicativo, vídeo e imagem são as palavras mais comuns que ocorrem na maioria dos tópicos; estes também têm um alto grau de centralidade e centralidade de intermediação.

Palavras-chave: Big Data Analytics, Competição, Apps de Entretenimento, Hierarchical Clustering, LDA, questões legais.

1. INTRODUCTION

In 2020, online home entertainment became the primary source of entertainment. The statement is supported by a report published by Google Indonesia (2020), which stated an increase in the consumption of online entertainment content. Although the consumption was rising, app developers were facing difficulties in surviving the market, retaining users, and being utilized for a more extended time by the users (Purizaga-Sorroza et al., 2022; Antunes de Souza, & Soares; 2021). Field Modgil (2018) research shows that most entertainment apps stay in users' devices only 7 days after the users downloaded the apps. According to the calculation done by Steiner (2021), 67.8% of apps have been downloaded less than 1000 times, 17.9% of apps have less than 1000 of active users, 1.4% of apps don't get any profit and only 0.5% are successfully run the business.

The apps in the entertainment category are varied in contents. Ramantoko et al. (2018) stated that online contents could be in form of audios, videos, texts, sentences, and images. The various types of content and sub-categories of entertainment apps make the competition

[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

more intense and more challenging to identify the position of an app in the competition, which apps are the direct competitors, and whether there is competition in apps between sub-categories.

Another issue contributing to the failure of an app is the pricing issue. Some apps offer prices that are not suitable for the content given and the target market. This pricing issue becomes the cause of failure for about 15% of the total reasons of apps downfall (CBInsights, 2021). Pricing issues usually occur because the business is not paying attention to the standard price in the market (CBInsight, 2021; Garnov et al., 2022; Teixeira, & Rodrigues; 2021). The problem could be avoided by identifying competitors' prices and other characteristics such as the quality of products to view the overall price available in the market and prevent setting prices too high/low or not suitable for potential users' ability to purchase the products (Faganello, & Muniz Fiuza Neto, 2021).

The competitive blind spot is one of the factors of the unsuccessful business. It causes the company to miss important events that may contribute to its business practices. Six blind spots have been identified by Czepiel & Kerin (2012); one of them is poor identification of competitors. It often happens because the managers only focus on competitors with a large scale of business and have top brands; therefore, this research is conducted to see the competition between apps with leading brands and apps that are not well known. Czepiel & Kerin (2012) agreed that the competition analysis process starts by identifying the competitors; if the company failed in the first task, further analysis would not be very accurate.

2. LITERATURE REVIEW

Uddin et al. (2020) uses two factors to identify competitors: supply and demand. Both factors are taken from research conducted by Clark & Montgomery (1999) which defines supply-based factors as characteristics of companies, who and what they do, and demand-based factors are characteristics of consumers, who and what consumers do. This is in line with Uddin et al. (2020) research statement. Still, another article written by Czepiel and Kerin (2012) revealed that the demand-based is a competitor in how companies meet the demands of the same consumer group. Supply-based is competitors through resources and owned and implemented by competing companies from an operational perspective.

Uddin et al. (2020) wrote eight factors included in supply-based and demand-based elements: service offered, value proposition, resource, pricing, and distribution included in supply-based aspects, and absolute size, success, and user perception in demand-based factors. These factors were also adapted from Clark and Montomery (1999), adapted from other research conducted by Reger & Palmer (1996), which did not classify these factors into supply and demand (Table 1).

[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

Table 1 Supply and Deman Factors

Supply Factors	Products/Services
	Value proposition
	Pricing
	Resource
	Distribution
Demand Factors	Absolute size
	Success
	Users' perception

Adapted from Uddin et al. (2020)

This study uses factors to identify competition adapted from the studies of Uddin et al. (2020) and Clark and Montgomery (1999). As previously mentioned, the factors are supply-based factors, defined as information provided by developers consisting of products, and demand-based factors, which are information provided by users, such as ratings and the number of raters. Table 2 shows the sub-variables that are classified into supply and demand factors. The first list stated products or/and services offered in the apps. This sub variable is chosen in the study because it determines the types/products offered as the primary indicator of what business actors provide. Products/services are the immediate value delivered to users and are the main factors that indicate what area a business is engaged in. The product or service offered by the apps can be identified through the description written by the developer on the AppStore platform. The description is also used in research conducted by Guo et al. (2017) and Uddin et al. (2020) to determine the application's tested features.

The following sub variable is user characteristics. User characteristics become essential in measuring competition because the products will be offered to users. Competition is seen by whether competitors serve the same user group. In this study, user characteristics can be seen from the minimum age limit of the user. Age was also used in a study written by Schumer et al. (2018), where the minimum age limit for users is employed to see the target users of diet and nutrition apps on the Google Play Store platform.

Performance is also used to see competition measured by ratings and the number of raters. Rating as a sub variable is adapted from several studies, namely Guo et al. (2017), Schumer et al. (2018), and Uddin et al. (2020). The rating and the number of raters will be clustered together because the rating is meaningless without knowing how many users have given the rating. A high rating and many raters indicate that the app is in great demand and shows user satisfaction (Uddin et al., 2017).

The last sub variable is the pricing strategy, which is seen from the price listed by the developer on the AppStore. As in the research conducted by Guo et al. (2017), pricing strategy

[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

is about how much the application's price is free or paid and whether the application provides in-app purchases or payments to the apps after the apps are installed on the device. Some apps can be downloaded for free without any additional charge, as these free apps' income will not be analyzed further in this study. Another pricing strategy is a free application with some items to be purchased to access more products/services in the apps. There are also paid apps in the AppStore, where the apps can't be downloaded if the user doesn't make payment via an Apple id account. These pricing strategies are included in this research because of a phenomenon that shows application failures caused by improper pricing.

Table 2 Variables of this research

Variable	Sub Variable	Indicator	Description
Supply-based Factors	Product	App description	Description about the apps provided by the developer contains information about type of products/services offered by the apps
	Users' characteristic	Minimum age limit	Showing the demographic target of the apps
	Pricing Strategy	Price	Competition in price
		In-app purchase option	One of apps' monetization strategies, where the users can purchase access to premium features
Demand-based Factor	Performance	Rating	Describing users' satisfaction toward the apps, the number of raters add more value to the rating
		Number of raters	

Adapted from Guo et al. (2017), Schumer et al. (2018), and Uddin et al. (2020)

3. METHODS

This study is quantitative research with secondary data collection. Some examples of secondary data are statistical reports, publications issued by the government, published or unpublished information about an agency or company either issued by the company itself or other parties, or large-scale data that has been collected or known as big data (Saunders, Lewis, & Thornhill, 2016). According to Goes (2014), using large volumes of data to assist an organization in making decisions is part of the concept of Big Data. While the analytical technique for processing Big Data is called Big Data Analytics (BDA) (Chen et al., 2012). BDA can also support competitive intelligence activities, such as assisting business players in identifying competitors, market structure, and business position in the market, because it will be difficult for companies to survive in a market if they do not know the detailed market

[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

perspective, such as research conducted by Liang Guo, Ruchi Sharma, Lei Yin, Ruodan Lu, Ke Rong, (2017), therefore BDA is employed in this study to analyze the competition of apps. The secondary data collected using scraping tools is the information listed on the AppStore platform for apps in the entertainment category. 2400 apps were selected to become a sample of this study using the convenience sampling method. The data collected through scraping are:

a) App name
b) Minimum age limit of users
c) Rating
d) Number of raters
e) Prices and in-app purchase information
f) Description that is processed into keywords

The data is taken from the platform using a web scraper. Web scraping is a good technique for extracting unstructured data from websites and converting them into structured data for analysis (De & Sirisuriya, 2015). The data that has been taken from the Appstore then goes through two data processing methods—using hierarchical clustering, the minimum age of users, rating, number of raters, prices, and in-app purchase information. This algorithm builds a hierarchy of the clusters and groups together the clusters close to each other by calculating the data matrix (Castillo, 2015) (Figure 1).

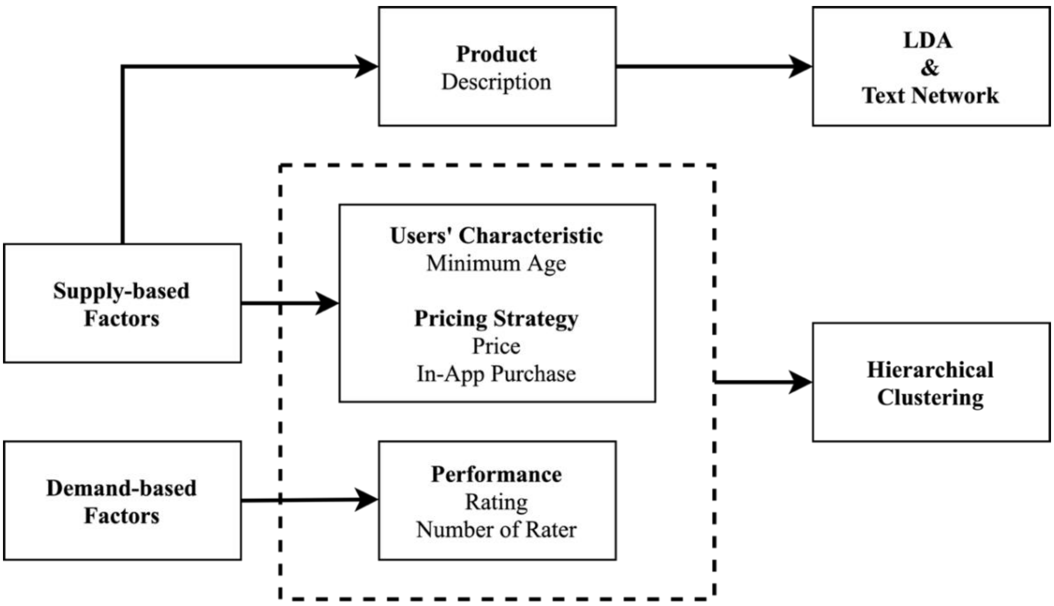

Figure 1 Theoretical framework

Another data collected from the AppStore is a description in the text. The data is prepared to be analyzed correctly and adequately using Text Preprocessing, which converts

[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

text into a list of words (Jo, 2019). This process is carried out to transform the text collected from the web scraping results into tokens that have been filtered or cleansed. A token is a single word or a collection of words calculated based on its frequency and serves as an analytical feature (Anandarajan et al., 2019). In this study, LDA searches for topics in documents taken from the AppStore, especially app descriptions in text data. This model is employed because it has an excellent performance in processing large amounts of text or documents and identifying latent topics (Chiru et al., 2014). After the text goes through the text preprocessing, the token collection is then analyzed with the LDA topic modeling approach to search for topics and see the types of products offered by the developer in the description.

Data that has gone through the topic modeling process is also processed and described with a Text Network. The text network can clearly explain the keywords that play an essential role in topic modeling. This model is used to identify important information from a document in the form of text and how a word or topic relates to each other to find the meaning behind the structure formed (Lambert, 2017).

Words resulting from the extraction of the description that is later used to analyze the topics are shown in Figure 2. The graphic hints at the types of products or features offered by the apps. The graphic states that the tokens or words taken from the raw data are nouns. The occurrence of words is calculated using TFIDF or Term frequency and inverse document frequency, which takes relevant the words by calculating the frequency and adding weight to the words. This method has an offset system that eliminates irrelevant common words and includes unique ones. The most common nouns in the app's description are "game" with the frequency of 3172, "app" with the number of frequencies 1732, "time," and "video" with the number frequency of 1032 and 905, respectively.

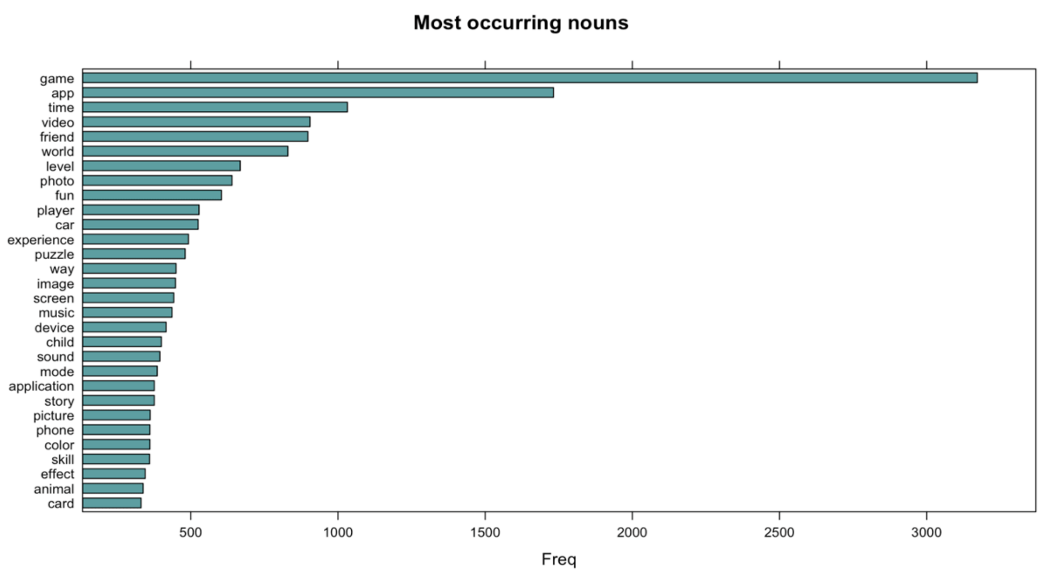

Figure 2 Most occurring nouns

Relações Internacionais do Mundo Atual Unicuritiba.
[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

4. RESULTS

Based on the results of the presentation of data that has been processed using hierarchical clustering, LDA topic modeling, and text networks, several things can be further stated:

Competition cluster based on minimum user age, rating and rating, price/pricing, and monetization

There are 7 clusters formed based on the minimum user age, rating, number of raters, prices, and in-App purchase options. Based on the theory, segmenting, targeting, and positioning are the keys to the digital marketing success (Chaffey & Ellis-Chadwick, 2016); this study discusses user segments targeted for apps grouped by cluster. Based on the results obtained, not all clusters consist of apps with the same age target. Cluster 1 is formed from apps that set a minimum user age of 9-19 years, compared to cluster 2, which is more specific, consisting of apps that offer products to users starting from 4 years. From cluster 1, it can be seen that although the minimum age limit for users ranges from 9 to 19 years, the age range belongs to the same group, namely teenagers to young adults, as well as cluster 6 with a minimum age limit of 9-17 years.

Meanwhile, clusters 5 and 7 are two clusters that do not show a specific group because the minimum age limit for users is from 3 years to 17. However, some things can distinguish the two clusters, which can be seen from the average minimum age of the user. In cluster 5, the average age is 11.25 years old, while the minimum average age in cluster 7 is 4.16; this indicates that in cluster 7, there are more apps with a minimum age limit of users closer to the age of 4 years compared to those in cluster 7 approaching 17 years of age(Table 3).

Table 3 Users' minimum age limit

Users' Minimum Age Limit	Cluster						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Min	9	4	4	4	4	9	3
Max	17	4	4	4	17	17	17
Mean	13,12	4	4	4	11,25	12,51	4,16

The three clusters, namely clusters 2,3, and 4 have a definite minimum age limit of 4 years old; this age can indicate two things, the content offered is content for children, or the content provided is very general and does not contain age-restricted aspects. Using a young

Relações Internacionais do Mundo Atual Unicuritiba.
[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

age limit allows the apps to be downloaded by anyone. On the other hand, the age limit that is too young does not show the actual target. Some apps in the cluster with a minimum age of 4 years target young users or children. This can be proven by examples of the content offered; several apps provide content for children, such as children's stories, children's puzzles, digital coloring books, and videos or songs for children.

The minimum age limit of users describes the type of content and if the content contains elements of violence or aspects that are only accessible to users of a certain age. Besides the type of content and adult elements, the period also indicates whether an app, especially game apps, can be accessed or played by a certain age because of difficulty and complication. If a game has a high level of difficulty, having a minimum age limit for a user that is too young will not be effective, considering most younger players, especially those who are still illiterate, will face difficulties in playing and will not enjoy the play. As has been said by Halbhuber et al. (2019) that the game must be made and adapted to the target user, not too easy or too difficult. If the game is designed to be too difficult, it can cause players to feel frustrated and decide to stop playing, as well as if it is too easy, players will feel bored.

In terms of price, most of the apps in this study can be accessed for free, which means that these apps can be downloaded without making a payment. All apps in cluster 5 can even be downloaded for free. However, these apps do not merely provide all services for free; some apps collect revenue from in-app purchases or using a freemium business model. Freemium is one of the business models usually applied by apps to maximize profit; other business models typically used by mobile apps are advertisements, subscriptions, and free-to-play or F2P (Davidovici-Nora, 2014). While all apps in cluster 5 can be downloaded for free, all four offer in-app purchases. The four apps in cluster 5 use a freemium business model, where users can still use the basic product for free, and if the user wants more features, the user can pay more to the developer (Holm & Günzel-Jensen, 2017). Some basic products offered by the developers in the four apps are not entirely free; users have to watch advertisements which are another source of income for the application; for example, Spotify, where users can listen to music without making a payment but have to listen to ads and can't use some feature (Table 4).

Table 4 In-app purchase

	Cluster						
In-App Purchase	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Yes	127	335	0	1	4	212	260
No	316	0	907	0	0	74	164

Beside clusters 4 and 5, some clusters consist of the app with a range of price that is not too broad. The price range in clusters 1 and 2 is IDR 0 – IDR109,000. Although the price

[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

ranges for the two clusters are the same, the difference is the average app price, with cluster 2 showing a lower average cost of IDR 3,056 and cluster 1, IDR 4,171. Cluster 3 has a higher price range and a higher average than the previous two clusters, IDR.0 - IDR159.000, with an average of IDR6,323. Cluster 6 has apps that offer a lower maximum price than clusters 1,2 and 3, but this cluster has a higher average number than clusters 1 and 2. The last cluster, cluster 7, has an extensive price range, IDR0 – IDR729,000, with a relatively low average of IDR14,605. It can be concluded that from the average price of clusters 1,2,3,6, and 7, which is much lower than the maximum price, more apps offer their apps for free compared to apps with prices close to the maximum price in these clusters(Table 5).

Table 5 Price

Cluster	**Number of Apps**	**Price**		
		Min	**Max**	**Mean**
1	443	IDR0	IDR109.000	IDR4.172
2	335	IDR0	IDR109.000	IDR3.057
3	907	IDR0	IDR159.000	IDR6.323
4	1	IDR3.299.000	IDR3.299.000	IDR3.299.000
5	4	IDR0	IDR0	IDR0
6	286	IDR0	IDR79.000	IDR5.091
7	424	IDR0	IDR729.000	IDR14.604

Cluster 4 is the only cluster with a single application. In terms of price, it is understandable why this application is the only one in cluster 4. The application has a very high price for a mobile application. In evaluating a product, the cost is one factor that determines user satisfaction; advance payment has a more significant influence on the assessment than the freemium pricing method (Wolkenfelt & Situmeang, 2020). The high price makes users hesitate to download the application because users do not know the application's features and do not have experience using the application. This contrasts with the Freemium business model, which offers users an app experience to entice users to use a premium account (Table 6).

Table 6 Cluster 5

App Name	**App Price**	**In App Purchase**	**Minimum Age**	**Rating**	**Number of raters**
TikTok	0	Yes	12	4,9	612.500
YouTube	0	Yes	17	4,8	1.400.000
Spotify	0	Yes	12	4,8	1.100.000
Trivia Crack	0	Yes	4	4,6	620.600

[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

The success of using the freemium business model compared to the paid app is evidenced by user feedback through ratings and raters. In cluster 5, the four apps have a high rating and many raters. Meanwhile, with a very high price, the app in cluster 4 does not get feedback, indicating that there are no users who have downloaded the application. Campbell (2022) says that users are willing to pay for computer software even at $50, while for a mobile app, users will think twice even if the app is priced at $1.99. Therefore, it will be difficult for mobile apps to get downloaded if the price offered is too high.

The average apps that became the research sample received poor rating to no rating. This statement is shown in several clusters with low ratings, including clusters 1,2,3, and 4, with the average ratings obtained, 0,0509; 0,2428; 0,0158; and 0, respectively. Although the means are low, several apps in these clusters, except for cluster 4, received a relatively high number of raters, and even cluster 3 had a maximum number of raters of 971. However, the average rater was still very low, namely 1.4. Among these clusters, cluster 2 has the highest average number of raters, which is 3.96 users, and the number is still very far from the maximum number. A low mean value indicates that many apps in the clusters did not receive a rating. Cluster 4 did not even get a rating at all. It could be because there were no users who downloaded the application due to the very high price (Table 7).

Table 7 Rating and number of raters

Cluster	Rate[1]			Number of Rater		
	Min	**Max**	**Mean**	**Min**	**Max**	**Mean**
1	0	2	0,0509	0	338	1,439
2	0	3	0,2428	0	766	3,964
3	0	1,4	0,0158	0	971	1,56
4	0	0	0	0	0	0
5	4,6	4,9	4,775	612500	1250000	933275
6	2	5	4,3646	1	115300	3780
7	0	5	4,425	0	360500	3677

Clusters 5,6 and 7 are clusters with better performance than the previous four clusters, especially cluster 5. Cluster 5 is a group of apps with a high rating, and the average rating obtained is also high. There are several factors why an application gets a high rating; besides the product/content offered to be attractive, the content of the information shared by the developer affects the rating (Pal Kapoor et al., 2020). In game and entertainment apps, information can be from user manuals and feature update information. Another study states that promotions and features on apps influence the rating (Askalidis, 2018); therefore, the

[1] Out of 5 rating

[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

freemium business model as applied to cluster 5 is a very effective model employed by apps because they let users experience their features and increase the possibility for the apps to get downloaded and feedbacks. Getting a rating is not just showing the application's performance, but with the rating that is obtained it triggers other users to download the application, Nicholas et al. (2015) say that apps with a high number of ratings look more attractive and get a higher number of downloads than many of the apps with low ratings.

The cluster results are shown in the previous discussion; from the seven clusters, several clusters in terms of rating or performance show poor performance. It can be said that apps with poor performance are in the same cluster and vice versa for clusters of apps with low performance. Not only in terms of the products offered, but in the cluster, developers can also see how many are in which group and who the competitors are. The cluster shows that apps with good performance and the right pricing strategy are considered competitors in the same environment.

Competition by description/topic

The approach to determining the topic applied in this research is Latent Dirichlet Allocation (LDA) which produces ten topics consisting of a collection of words that state the main idea in the application description. The collected words that express the topic are listed in table 8.

Table 8 Topics

No.	LDA	Network[2]
1	slot, vega, machine, real, time, home, slot machine, game, make, like	camera, device, account, ad, friend, fun, experience, adventure, battle dan character
2	use, child, play, time, animal, puzzle, game, friend, tap, sound	game, child, app, friend, dan puzzle
3	game, card, play, player, world, time, car, wheel, get, use	animal, app, car, child, world, game, friend, card dan player
4	video, music, make, create, wallpaper, home, friend, design, get, new	favorite, image, music, effect, video, app, friend, application dan game
5	video, subscription, use, watch, screen, create, support, stream, live, photo	app, game, application, friend, video, image, movie, picture, photo
6	robot, drive, new, get, screen, game, bus, car, call, city	player, mode, camera, screen, app, friend, photo, car, game, child, fun, application
7	game, shoot, fishing, play, hunt, kill, target, sniper, make, gun	device, experience, game, phone, fun, friend, zombie, city, level, d, app, control
8	photo, keyboard, use, subscription, color, emoji, game, create, make, sticker	game, friend, color, photo, video, image, app
9	game, play, get, new, challenge, puzzle, win, casino, world, real	application, made, app, video, card, game, car, friend, fun, coin, level, puzzle dan adventure
10	story, world, tap, read, new, disney, favorite, like, make, character	adventure, friend, app, story, character, video, channel, child, world, player, picture, game

[2] keywords with highest number of degree centrality and betweenness centrality

Relações Internacionais do Mundo Atual Unicuritiba.
[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

The results of topic modeling and networking for topic 1 are very different compared to other topics with some similar keywords between results from LDA and text network. From the results of the LDA on topic 1, it produces keywords about casino games, while the results from the network show keywords related to battle, adventure, and experience. But if we look closely, the network in figure 3 has words that support modeling topics such as gambling, money, machines, and slots that support the results of the modeling topic. Of course, gambling or gambling-based games are intended for specific user groups. This user group is also the target of apps that provide dating services which are also the keywords listed on the network in the image. The keyword character is also connected to the words criminal and crime, which are not commonly used for underage users. Although there are differences in the main keywords, topic 1 shows a pattern that leads to specific users. The content offered may differ, but the target user group could be the same.

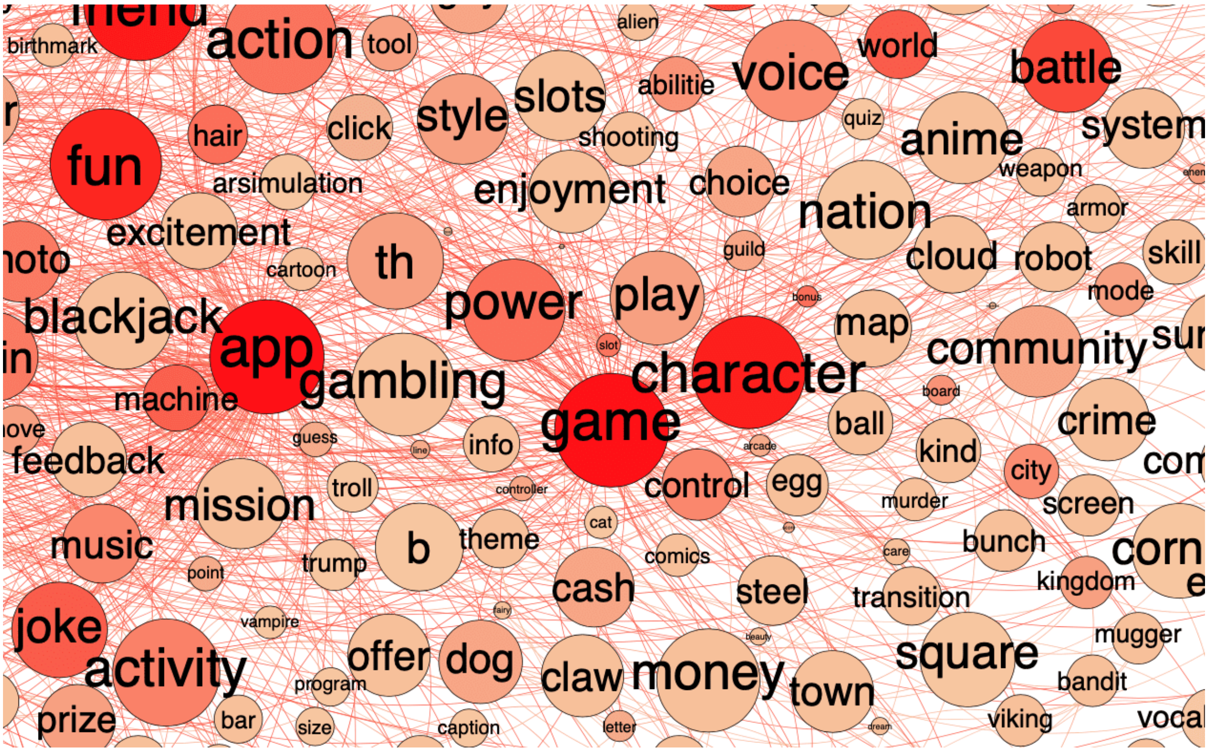

Figure 3 Part of text network of topic 1

Topic 2 is about children's puzzle games, which target specific age groups because the keywords that appear as the keywords with the highest occurrences are child, game, and puzzle. Similar to topic 2, topic three also offers games. The types of games that are most highlighted through modeling and networking topics are cards and cars. In the network and the results from the LDA, it can also be seen that the keyword “child” indicates that the group of apps containing topic 3 is an app that also provides games for children. But unlike the case

[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

with topic 2, which offers simpler games such as talking tom (talking animals and shape puzzles for toddlers and preschoolers), topic 3 offers apps with a broader range of children's ages, from younger to older children, toddlers, and quite advanced children's games for older ones.

Topic 4 is more about competition between apps with content types in the form of music and videos; besides games, music and video are types of content that are mostly enjoyed by many users (We Are Social, 2021). Topic 4 shows that videos can also compete with other content such as music and radio because one app can provide both contents simultaneously. Examples are TikTok, YouTube, and Spotify. Topic 5 has almost the exact keywords as topic 4, but this topic is more about video and TV, for instance, HBO GO, Nimo TV, and Netflix. It can be seen in both topics that there is overlap between the two topics because the keywords are almost the same. In this case, the text network plays a crucial role in explaining the products offered through words connected to the word “Video”. It can show what types of content/features the application provides in clusters 4 and 5 (Figure 4).

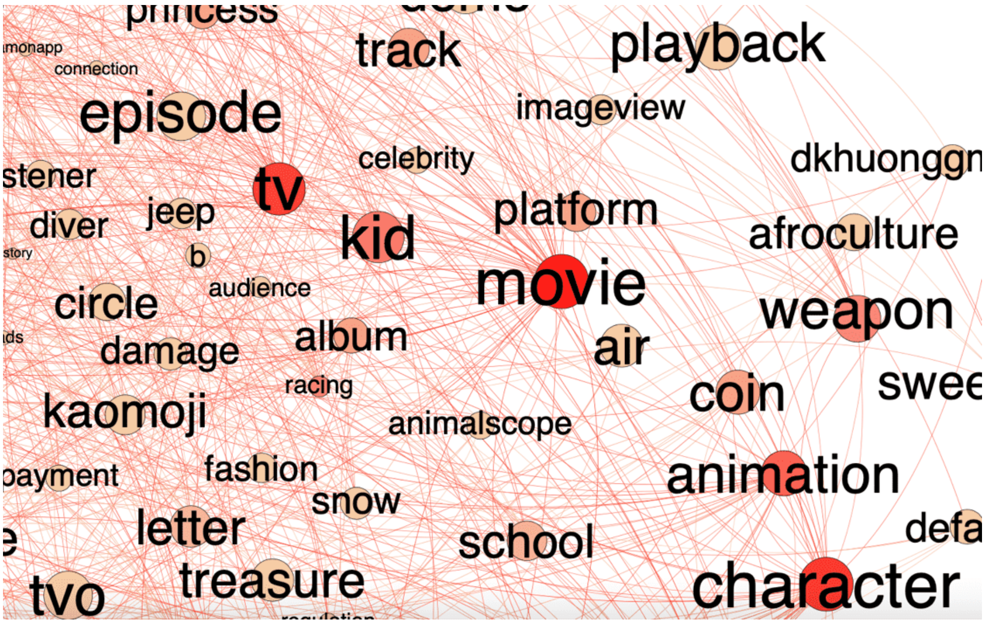

Figure 4 Part of text network of topic 5

On topic 6, keywords generated from the LDA topic modeling show more about driving simulation games and transforming robots into cars. Topic 6 is about games mainly in the form of simulation and reality or virtual reality that uses a camera. In contrast to the previous topics, topic six does not show a prominent pattern in which groups or specific types of apps compete. Still, generally, the apps compete in this group provide games as their main product. In this topic, the keywords resulting from LDA give more apparent hints to what apps are. However,

Relações Internacionais do Mundo Atual Unicuritiba.
[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

the words with high degree centrality and betweenness centrality resulted in the text network being more abstract.

Topic 7 has one common keyword generated from both methods, namely game, but the keywords that appear in both methods are slightly different. In LDA, the keywords are more about the shooting, and the keyword associated with shooting in the text network is “zombie”. The topic in general shows that competing apps on this topic are target shooting apps, the majority of which are zombie shooting as the target or war and shooting games against enemies. These apps, of course, target different users from topics 2 and 3. Games in this application group require skills and age that meet the requirements, and parental guidance is needed when young users play because of violent content(Figure 5) .

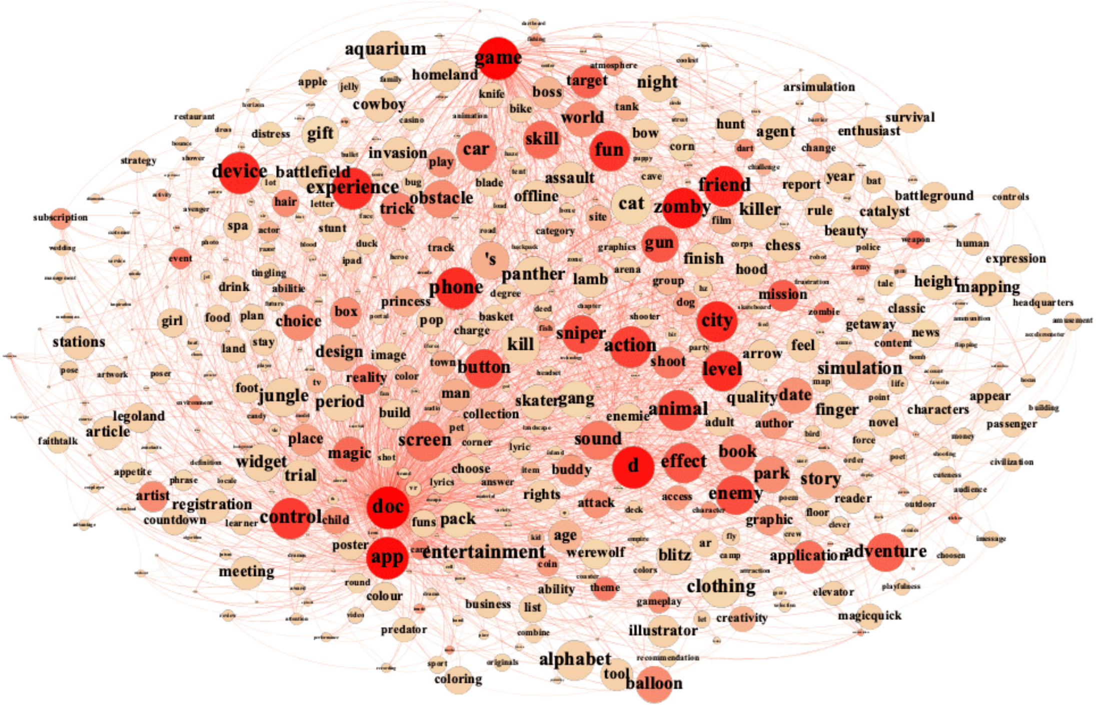

Figure 5 Text network of topic 7

Topic 8 is a collection of competing apps that offer mainly photos; some specific keywords are digital coloring and emojis. An example of an application in topic 8 is Emoji and Facemoji Keyboard: Fonts&Emoji. Besides emoji, another keyword is considered a similar product to emoji; the word is “sticker”. Emojis and stickers are images commonly used in conversation apps and social media to illustrate more emotion and emphasize the messages. As mentioned before about overlapping between topics shown by the exact keywords appearing in both topics, topics 5 and 8 have a common word, "photo.” However, the word photo in topic five is associated with video and movies. The text network visualizes more words connected to the primary keywords mentioned in table 8. As viewed in figure 5, besides emoji,

[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

sticker, and coloring book, other words indicate other features of the apps, such as collage, viewing, and editing, which can be summarized as photo editing apps. The words background and wallpaper show the apps that offer pictures for wallpaper or background pictures as their product. Therefore, it can be concluded that apps provide the main product in the form of an image, emojis, stickers, background pictures, and coloring books. Apps that provide services or features of image editing are considered competing (Figure 6).

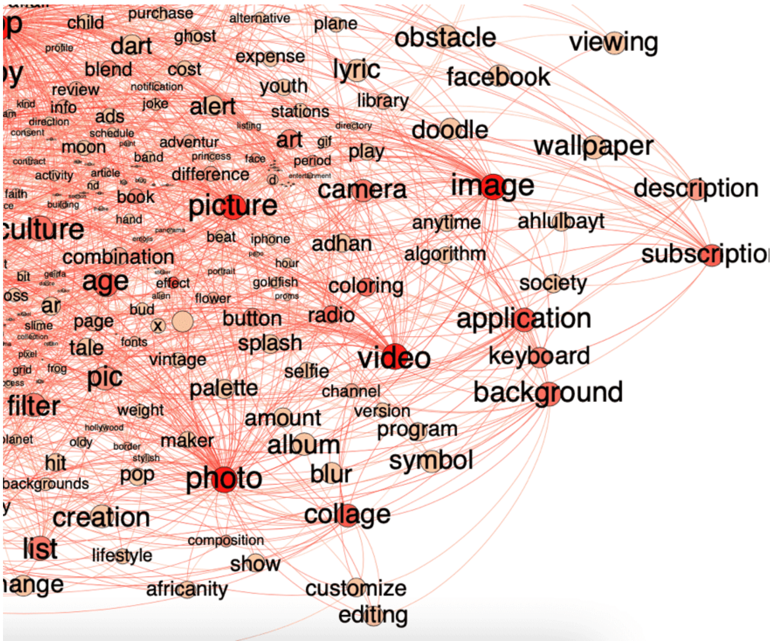

Figure 6 Part of text network of topic 8

Topic 9 has the exact keywords as topic 2, namely games and puzzles. But if you look at other keywords besides puzzles and games, topic nine is not intended for children because several keywords indicate that the application targets adult users. The keywords are casino and coin, which are usually gambling games. Examples of apps included in topic nine are Magic Solitaire: Card Puzzle and Real Casino Vegas Slot Machine, which have a minimum age of 17 years old.

Topic 10 is the last topic, with keywords appearing in two methods: story, world, and character. Topic 10 contains apps that compete to offer content in the form of character stories. Stories can be in the form of images, text, or videos. According to this topic, apps that offer comics and digital novels compete in the same market as apps that provide stories in the form of videos, which can be films or TV series. An example of apps that offer a story in the form of images is Webtoon: Comics, and novels in videos, movies and TV series are Disney+ Hotstar.

From the results of the topic analysis using text network, the words game and app appear the most and have a high degree of centrality and betweenness centrality. Betweenness centrality is when nodes are randomly met with shorter paths and are connected

Relações Internacionais do Mundo Atual Unicuritiba.
[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

or brought together by nodes with high betweenness centrality; words with high betweenness centrality play an essential role in securing and forming a meaning (Paranyushkin, 2019). But in the LDA topic modeling results, the words game and app did not come out as essential words and did not describe much in the topic. However, some keywords appear in the results of both LDA and text network, which means that these words describe important words and provide an overview of the types of content offered to these groups because they have been proven in two different ways, namely from LDA which look for latent topics and text network to find words that have a high relationship with other words.

The keywords with the most occurrences and the highest centrality do not merely describe an application's main types of services. Keywords can be supporting content from an app or just one of many types of content that an app provides. Two apps can have the exact keywords but offer different main products because specific keywords form these apps. By using a network, it can be seen that a type of content provided can have various features or forms. For example, the keyword “story” on the network is connected to other words such as children, religion, or adults, indicating that the story could be a story for children, adults, or religion. Therefore, a text network is needed to see the type of content and distinguish between one kind of app to another. This statement is also in line with research conducted by (Wang et al., 2020); the study used the network to see the product’s features.

In addition to the keywords game and app, another keyword that appears in several topics is "friend.” These keywords show that the application provides features to share content or can be played with friends. The Application Programming Interface (API) allows users to share content and news with friends without leaving the website page to make it easier to share information about the application (Bodle, 2011). Sharing with friends has a positive and negative impact. On the positive side, with this feature, users can share about the application with other potential users; this right can increase the number of application users and become a crucial process in the success of the application (Dubois et al., 2016). On the negative side, Bodle (2011) asserts that this sharing feature makes users more vulnerable in the context of privacy. The word friend is also associated with collaboration, or the use of apps with friends, which is directly related to social factors for users. Yang & Lin (2019) stated that social influence is significant in shaping the behavior of app users. In some cases, this behavior can also lead users to hedonic behavior and higher app usage intensity.

5. CONCLUSIONS

There are 7 clusters generated using hierarchical clustering and ten topics using LDA, which are further analyzed using a text network. The minimum age limit for users most widely used by the application is four years old. The prices offered are very diverse, with a minimum

[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

of IDR 0 or free. The application with the highest cost of IDR3.299.000 is also the only application in cluster 4 because the price offered is much different from other apps in other clusters. Cluster 5 is the application that has the best performance seen from the number of raters and the highest rating, with an application price of IDR0 and offers in-app purchases. Apps with good performance judged by many raters and high ratings are in the same group, with the same pricing strategy, and vice versa, with apps that show poor performance. Although some apps in the same cluster do not have the same price, minimum age of users, and the number of ratings, these apps compete in the same cluster because these characteristics are still similar to each other or still in the same range.

The result from the text network, the keywords app, and the game has a high degree of centrality and betweenness centrality on almost all topics, indicating that many sample apps are games. Some of the exact words appear on different topics. However, the network helps analyze the types of content offered, more specific than the main keywords generated from LDA. In addition to the word game, other keywords included in various topics are video and image.

Limitations

This study is far from perfect. Some aspects can be developed in future research. The variables employed to form the clusters and determine the topics are adapted from previous research; however, some do not entirely reflect the study's true objective because it is limited to the platform chosen for data collection. Future research might investigate more by taking data from various platforms and conducting a study on apps that show a significant role in the results of this study.

Acknowledgments: not applicable.

REFERENCES

Anandarajan, M., Hill, C., & Nolan, T. (2019). *Practical Text Analytics*. *2*. https://doi.org/10.1007/978-3-319-95663-3

Antunes de Souza, M. C. da S., & Soares, J. S. (2021). Natureza Social do Homem, Consumismo e Sustentabilidade. Journal of Law and Sustainable Development, 9(1), e727. https://doi.org/10.37497/sdgs.v9i1.75

Askalidis, G. (2018). *The Impact of Large Scale Promotions on the Sales and Ratings of Mobile Apps: Evidence from Apple's App Store*. http://on.wsj.com/1Haufei

Bodle, R. (2011). REGIMES OF SHARING. *Https://Doi.Org/10.1080/1369118X.2010.542825*,

[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

14(3), 320–337. https://doi.org/10.1080/1369118X.2010.542825

Campbell, P. (2022). Mobile App Pricing: What App Pricing Strategy Works Best? [+tips]. Priceintelligently.com. Retrieved 29 March 2022, from https://www.priceintelligently.com/blog/mobile-apps-can-learn-from-saas-pricing-models. - Google Suche. (n.d.). Retrieved May 8, 2022.

CBInsights. (2021). *Why Startups Fail: Top 12 Reasons l CB Insights*. https://www.cbinsights.com/research/startup-failure-reasons-top/

Chaffey, D., & Ellis-Chadwick, F. (2016). Digital Marketing Strategy, Implementation and Practice Sixth Edition. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9).

Chen, H., Chiang, R.H. and Storey, V.C. (2012). Business intelligence and analytics: from big data to big impact. MIS Quarterly, Vol. 36 No. 4, pp. 1165 1188.

Chiru, C., Rebedea, T., & Ciotec, S. (2014). *Comparison between LSA-LDA-Lexical Chains*. https://doi.org/10.5220/0004798102550262

Clark, B. H., & Montgomery, D. B. (1999). Managerial Identification of Competitors. *Journal of Marketing*, *63*(3), 67–83. https://doi.org/10.1177/002224299906300305

Czepiel, J.A. and Kerin, R.A. (2012), “Competitor analysis”, in Shankar, V. and Carpenter, G.S. (Eds), Handbook of Marketing Strategy, Edward Elgar, Cheltenham and Northampton, MA, pp. 41-57.

Davidovici-Nora, M. (2014). *Paid and Free Digital Business Models Innovations in the Video Game Industry*. Communications & Strategies. https://studylib.net/doc/25265552/paid-and-free-digital-business-models-innovations-in-the-...

De, S., & Sirisuriya, S. (2015). *A Comparative Study on Web Scraping*.

Dubois, D., Bonezzi, A., & De Angelis, M. (2016). Sharing with Friends versus Strangers: How Interpersonal Closeness Influences Word-of-Mouth Valence: *Https://Doi.Org/10.1509/Jmr.13.0312*, *53*(5), 712–727. https://doi.org/10.1509/JMR.13.0312

Faganello, C. R. F., & Muniz Fiuza Neto, O. (2021). Sanitation: Right guaranteed by the Federal Constitution. Journal of Law and Sustainable Development, 9(1), e673. https://doi.org/10.37497/sdgs.v9i1.79

Garnov, A., Ordov, K., Chelukhina, N., Perepelitsa, D., & Asyaeva, E. (2022). Innovative Financial Economic Stimulation Tools For ESG-Transformation of a Company: opportunities for application and specifics of regulation. Journal of Law and Sustainable Development, 10(2), e0219. https://doi.org/10.37497/sdgs.v10i2.219

Goes, P. B. (2014). Big Data and IS research. *MIS Quarterly, 38*, 3–8.

Google Indonesia. (2020). *Year in search 2020: Indonesia search trends - Think with Google APAC*. https://www.thinkwithgoogle.com/intl/en-apac/marketing-strategies/search/2020-year-in-search-indonesia-looking-back-to-move-your-business-forward/

Guo, L., Sharma, R., Yin, L., Lu, R., & Rong, K. (2017). Automated competitor analysis using big data analytics: Evidence from the fitness mobile app business. *Business Process Management Journal*, *23*(3), 735–762. https://doi.org/10.1108/BPMJ-05-2015-0065

[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

Halbhuber, D., Fehle, J., Kalus, A., Seitz, K., Kocur, M., Schmidt, T., & Wolff, C. (2019). The mood game - How to use the player's affective state in a shoot'em up avoiding frustration and boredom. *ACM International Conference Proceeding Series*, 867–870. https://doi.org/10.1145/3340764.3345369

Holm, A. B., & Günzel-Jensen, F. (2017). Succeeding with freemium: strategies for implementation. *Journal of Business Strategy*, *38*(2), 16–24. https://doi.org/10.1108/JBS-09-2016-0096

Jo, V. (2019). Introduction. In *Seminars in Diagnostic Pathology* (Vol. 36, Issue 2). https://doi.org/10.1053/j.semdp.2019.02.002

Lambert, Natalie. (2017). "A Text Mining Tutorial." In *Group Processes: Computational and Data Driven Approaches*, edited by A. Pilny and M. S. Poole. Springer Press.

Modgil, S. (n.d.). *Netflix Sees A Potential Of Adding The Next 100 Mn Customers From India: Reed Hastings, CEO Netflix*. Retrieved May 8, 2022, from https://inc42.com/buzz/netflix-india-reed-hastings/

Nicholas, J., Larsen, M. E., Proudfoot, J., & Christensen, H. (2015). Mobile Apps for Bipolar Disorder: A Systematic Review of Features and Content Quality. *J Med Internet Res 2015;17(8):E198 Https://Www.Jmir.Org/2015/8/E198*, *17*(8), e4581. https://doi.org/10.2196/JMIR.4581

Pal Kapoor, A., Vij, M., Pal Kapoor, A., & Vij, M. (2020). How to Boost your App Store Rating? An Empirical Assessment of Ratings for Mobile Banking Apps. *Journal of Theoretical and Applied Electronic Commerce Research*, *15*(1), 99–115. https://doi.org/10.4067/S0718-18762020000100108

Paranyushkin, D. (2019). InfraNodus: Generating insight using text network analysis. *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, 3584–3589. https://doi.org/10.1145/3308558.3314123

Purizaga-Sorroza, M. A., Campos-Ugaz, W. A. ., Campos-Ugaz, O. ., Hernandez, R. M., & Garcia Flores, L. A. . (2022). Methodological Model for the Research Functionin the Professional Training of Lawyers. Journal of Law and Sustainable Development, 10(2), e0222. https://doi.org/10.37497/sdgs.v10i2.222

Ramantoko, G., Fatimah, L. V., Pratiwi, S. C., & Kinasih, K. (2018). Measuring digital capability maturity: Case of small-medium Kampong-digital companies in Bandung. *Pertanika Journal of Social Sciences and Humanities*, *26*(T), 215–230.

Reger, R. K., & Palmer, T. B. (1996). Managerial Categorization of Competitors: Using Old Maps to Navigate New Environments. *Https://Doi.Org/10.1287/Orsc.7.1.22*, *7*(1), 22–39. https://doi.org/10.1287/ORSC.7.1.22

Saunders, M., Lewis, P., & Thornhill, A. (2012). Research methods for business students. Harlow: Pearson.

Schumer, H., Amadi, C., & Joshi, A. (2018). Evaluating the dietary and nutritional apps in the google play store. *Healthcare Informatics Research*, *24*(1), 38–45. https://doi.org/10.4258/hir.2018.24.1.38

Steiner, R. (2021). *Why 99.5% of consumer apps fail (and how to keep yours alive) - Fyresite*. https://www.fyresite.com/how-many-apps-fail/

[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]

Teixeira, T. C., & Rodrigues, J. de A. (2021). Integrity Programs in the Fight Against Corruption and Their Relevance to the Professional Supervisory Councils. Journal of Law and Sustainable Development, 9(2), e0730. https://doi.org/10.37497/sdgs.v9i2.30

Uddin, M. K., He, Q., Han, J., & Chua, C. (2020). App competition matters: How to identify your competitor apps? *Proceedings - 2020 IEEE 13th International Conference on Services Computing, SCC 2020*, 370–377. https://doi.org/10.1109/SCC49832.2020.00055

Wang, X., Wang, Y., & Liu, A. (2020). Determining Customer-Focused Product Features through Social Network Analysis. *Procedia CIRP*, *91*, 704–709. https://doi.org/10.1016/J.PROCIR.2020.02.227

We Are Social. (2021). Digital Report 2021 : Indonesia. *Global Digital Insights*, 103.

Wolkenfelt, M. R. J., & Situmeang, F. B. I. (2020). Effects of app pricing structures on product evaluations. *Journal of Research in Interactive Marketing*, *14*(1), 89–110. https://doi.org/10.1108/JRIM-11-2018-0141/FULL/XML

Yang, & Lin. (2019). *Why do People Continue to Play Mobile Game Apps? A Perspective of Individual Motivation, Social Factor and Gaming Factor | Yang | Journal of Internet Technology*. https://jit.ndhu.edu.tw/article/view/2179

[Received/Recebido: Março 10, 2022; Accepted/Aceito Maio 16, 2022]