Improving Game Marketing using Classification of Data Mining Technique

Breezem Fernandes
Dept. of Information Technology
St.Francis Institue of Technology,Borivali
Mumbai, India

Dr. Nazneen Ansari
Dept. of Information Technology
St.Francis Institue of Technology,Borivali
Mumbai, India

Pratik Dias
Dept. of Information Technology
St.Francis Institue of Technology,Borivali
Mumbai, India

Harsh Deasai
Dept. of Information Technology
St.Francis Institue of Technology,Borivali
Mumbai, India

Abstract— The gaming industry is an emerging industry in today's world. Games are played worldwide on various platforms like mobiles, computers, consoles. These games generate tremendous amounts of data. This gaming data can be used to perform mining operations and hence generate results that can help in improving the overall gaming industry. Hence this Paper aims to apply data mining technique classification on gaming data. Classification algorithms like J48 and REP-Tree are applied on a gaming dataset for improving the marketing factor of the games. Classified player classes are identified as targeted population for potential marketing. This electronic document is a "live" template and already defines the components of your paper [title, text, heads, etc.] in its style sheet

Keywords— Data Mining; Classification; J48 Algorithm; REPTree Algorithm; game data mining.

Due to the development of technologies like Smartphone's, Graphic enhancing cards, online server gaming, the gaming industry is a trending industry in today's world. The revenue generation of the gaming industry competes with top entertainment industries. According to a survey by newzoo "Consumer spend on games will grow to $196.0 billion by 2022"[1].

As a result, it can be deduced that the gaming industry generates data in massive quantities. The mining of such data can be greatly beneficial for the gaming industry. Different data mining algorithms and techniques can be applied to processed data sets of games and the obtained results can be used to make precise decisions that will improve game factors like game design, game marketing, and game life. Game design refers to the overall game play experience of a user. Targeted marketing can be used for specific players so that they get better game recommendations.

The remainder of this paper is organized as follows, Section II provides an overview about data mining and data mining techniques. Section III consists of methodologies that define various terminologies related to game data mining and the algorithms used. Section IV consists of the system overview. Section V consists of implementation of the actual system followed by results in section VI. Section VII consists of summarization and is followed by references.

I. LITERATURE REVIEW

Data mining is the process in which interesting knowledge and patterns are discovered from large amounts of data stored either in data warehouses, databases, or other information repositories. Data mining refers to the use of different data mining techniques like classification, clustering, and association on data to obtain useful knowledge from data for decision making and other applications. Data mining can be applied to any kind of data as long as the data are meaningful for a target application [2].

The paper 'Better Game Design using Association Analysis' conveys us [3] how Association Analysis could be used to extract knowledge from a gamer data set to create strong rules that can guide the game design process. Using the strong rules in the design phase will enable the designers to make a successful game and tap into a wide pool of gamers, thereby generating commensurate revenue. The standard approaches to solve the problem of finding and using the right ingredients to build up a great game, consist of one (or a mixture) of the following: User polls and surveys Gaming Forums, Market Research and sometimes even wild guesses, with the hope that the game will turn out to be likeable. The rules generated from association analysis would be of tremendous benefit to the Game Design industry, as they can then use them to design profitable games data set of game players habits and preferences is collected from St. Francis Institute of Technology, Mumbai, and subjected to the Apriori algorithm. The database contains various aspects of game design that users would like to experience in the games they play. The rules generated from association analysis would be of tremendous benefit to the Game Design industry, as they can then use them to design profitable games.

According to paper 'Tuning Mobile Game Design Using Data Mining'[4], in the paper the decision to instrument the code to collect as much information as possible about the game play, and performed a rather limited play-test during two public events. Applied data mining both to look for peculiar characteristics of the platforms employed, to discover interesting patterns, and to identify flaws and/or opportunities

in our game design. Also discovered some design opportunities to inherently modify the game difficulty by leveraging upon the way players tend to use the devices. The game play based on the analysis presented in this paper, along with a validation of such changes, and further analyses of the collected data (e.g., applying clustering techniques to identify different types of players).

II. METHODOLOGY

Classification technique here for better game marketing. Game marketing allows advertisers to pay to have their name or products featured in digital games. Which has been a great part of revenue generation in the perspective of the gaming industry. Below in III.B. classification technique is explained while various techniques to do classification w.r.t. Gaming like are explained Estimation, Prediction, Game Marketing, Targeted Marketing are explained from III.B.I to III.B.IV respectively.

A. Classification

It is a data analysis task, i.e. the process of finding a model that describes and distinguishes data classes and concepts. Classification is the problem of identifying to which of a set of categories (subpopulations), a new observation belongs to, based on a training set of data containing observations and whose categories membership is known. A player can be classified into predefined classes and then actions can be performed such can game recommendations can be sent to players. It is used to organize data into classes, which is hugely useful to game development. For example, classifying players based on their potential to become paying users vs. non-paying, or classifying player behavior in a shooter game to test if the players play the game as intended by the games' design. It uses class labels to order objects in a data collection, normally using a training data set where all objects are already associated with known class labels (e.g. playtime per level associated with character class). [5] The classification algorithm here by the write suggests that uses- leans from the test data and builds a model that can be applied to other or future data.

- Estimation :This is similar to classification, but the target variable is numerical, not categorical. In statistics, methods such as regression and correlation are estimation methods. For example, we are interested in knowing a value, not obtaining information about how our data groups into distinct classes. For example, estimating how much money a player will spend on in-game items, or how long a player will continue playing a specific game. In estimation, models are built using training data of complete records, which provide the value of the target variable as well as the predictors (causal variables).

For example in paper[5], using simple regression to find the relationship between two variables, such as playtime and money spent on in-game items.

- Prediction :It is reminiscent of classification and estimation, but with prediction, we want to know about the future. The core idea is to use a large number of known values to predict possible future values. For example, how many players an MMORPG will have 3 months into the future, or when there will only be 1,000 active players left in a social casual game or how many players are needed to reach the critical threshold when player communities become self-sustaining. There are many approaches to prediction, from traditional statistical methods to more specialized knowledge discovery methods, such as decision tree analysis, and k-nearest neighbor. [5]Prediction is one of the most widely applied data mining methods in the analysis of data from multi-player and massively multiplayer persistent games, were predicting the effect of design changes or the behavior of the player community, is important for revenue. Prediction can be used to forecast in many contexts around game development and -publishing.

- Game Marketing:Game marketing is defined as the sale of games to different types of gamers all around the world. There are various traditional methods of marketing games such as pre-purchase, using platforms like ps4, steam, epic games, Xbox, etc. these methods are effective but not very efficient. Targeted game marketing strategies helps to identify players who have a higher probability of purchase.

- Targeted Marketing :Targeted marketing is the process of identifying customers and promoting products and services via mediums that are likely to reach those potential customers. Targeted marketing classifies potential customers, discovers their preferred content delivery mode and digital hangouts and then builds a marketing strategy aimed at that specific group. [6]In the paper, Targeted marketing is generally limited in scope but is often more productive than broader types of marketing because it is designed around specific customer preferences.

B. Algorithms used

- J48 Algorithm: J48 algorithm takes input as factors and the using entropy and gain creates a splitting attribute using which a decision tree can be created. J48 is an extension of ID3. The additional features of J48 are accounting for missing values, decision trees pruning, continuous attribute value ranges, derivation of rules, etc. The decision trees generated by C4.5 can be used for classification, and this reason, C4.5 is often referred to as a statistical classifier. In 2011, authors of the WEKA machine learning software described the C4.5 algorithm as "a landmark decision tree program that is probably the machine learning workhorse most widely used in practice to date". KA data mining tool, J48 is an open-source Java implementation of the C4.5 algorithm. C4.5 builds decision trees from a set of training data in the same way as ID3, using the concept of information entropy At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the

normalized information gain (difference in entropy). The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recurses on the partitioned sub lists.[7]This algorithm has a few base cases.

1. All the samples in the list belong to the same class. When this happens, it simply creates a leaf node for the decision tree saying to choose that class.
2. None of the features provide any information gain. In this case, C4.5 creates a decision node higher up the tree using the expected value of the class.
3. The instance of previously unseen class encountered. Again, C4.5 creates a decision node higher up the tree using the expected value.

• REPTree Algorithm:RepTree uses the regression tree logic and creates multiple trees in different iterations. After that, it selects the best one from all generated trees. That will be considered as the representative. REP Tree is a fast decision tree learner which builds a decision/regression tree using information gain as the splitting criterion, and prunes it using reduced error pruning. It only sorts of values for numeric attributes once. [8] Reduced Error Pruning Tree ("REPT") is fast decision tree learning and it builds a decision tree based on the information gain or reducing the variance. REP Tree is a fast decision tree learner which builds a decision/regression tree using information gain as the splitting criterion, and prunes it using reduced error pruning. It only sorts of values for numeric attributes once. Missing values are dealt with using C4.5's method of using fractional instances.

III. PROPOSED SYSTEM

The main reasons for using data mining are as follows:

1. Excessive data and insufficient information

2. There is a need to extract useful information from the data and to interpret that data.

"Figure 1.", here shows the system diagram which resembles the data sources from various resources where the data is generated. That data is collected in data sets where in cleaning and transformation with the help of WEKA tool and hence we can perform analysis and right decision making actions

With enormous data volumes, human analysts with no special tools can no longer make sense. However, automating the process of finding relationships and patterns in raw data can be achieved through data mining, and the results can then be either utilized in a decision support system or assessed by a human expert. Similarly, data mining can be applied to computer games to extract information from gaming data and used in taking valuable decisions.

When gamers play games online, a large amount of data is generated. Such data is stored on game servers. Gaming industry can take this data and apply certain data mining techniques to determine relationships and patterns, which can help in better decision making. There are lots of new games launched every year. The success of each game depends upon various factors, one of which is game marketing. If games are marketed to players based upon the players' interest then there is a high probability of purchase.

To Implement this system the gaming industry can use game data and create a prediction model to classify various players based upon his/her interest. The classified players can further be given game recommendations and thus improve the game marketing factor.
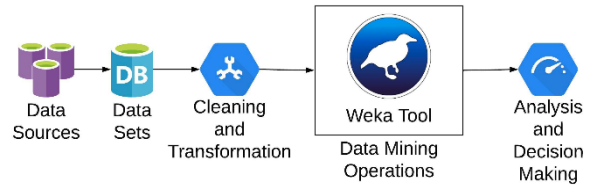

Fig 1: System Diagram

Data mining has been proved advantageous for sectors like agriculture, medical, education. According to the proposed model, it can also be beneficial to the gaming sector. The gaming dataset executed using the WEKA tool can aid the gaming industry in improving marketing. [9]

In "figure 2", where a implementation diagram is shown as the two different algorithms that are J48 and REP Tree algorithm which are used for decision making and the identification of classes i.e. the buyers who have higher potential to buy the game as well as in which a certain advertisements can be made in those games. Game and ads recommendation can be accordingly. and sales as well as ad pichment can be done accordingly.

The are several algorithm categories for both fields, however, the most classical categories are Linear and Logistic for regression algorithms; and Decision Trees, Naive Bayes, Neural Networks and Support Vector Machines for classificationThe algorithms are usually designed to consider this information so feedback in order to generate a good predictive model adapted to the data. There Are two main approaches in these methodologies regression (which tries to estimate the relationship between the variables using statistical models) and classification (which tries to create models predict a target class variable).
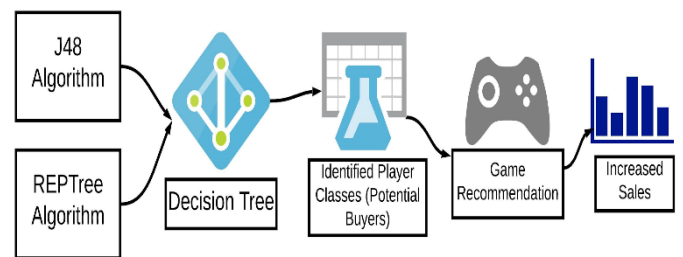

Fig 2: Implementation Diagram

The algorithms J48 and REPTree creates a decision tree as output and the "buying category players" are identified and game recommendations are given to these players increasing the sale of games.

IV PROCEDURE

A. Data Collection

Data gathering tasks are performed using a survey method. The survey form is circulated through online resources. The questionnaire consists of the following questions.

Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NTASU - 2020 Conference Proceedings

1. Age of the Player

2. Salaried or not

3. Plays Online game or not

4. Weekly hours played

5.A platform on which the player games

6.skill level of the player

7.If the player buys a newly launched game or not.

Below is the example of sample dataset.

TABLE 1.   Example of sample data set

Age	Salaried	Online_gaming	Weekly_hours	platform	skill_level	Buys_newGame
20	yes	yes	>8	pc	<5	yes
21	yes	no	>8	console	>5	no
15	no	yes	<8	pc	>5	no
33	yes	yes	>8	pc	>5	yes
21	yes	yes	>8	pc	>5	no
20	yes	yes	>8	pc	>5	yes
20	yes	yes	>8	pc	>5	yes
20	yes	no	>8	console	>5	yes
21	yes	yes	>8	pc	>5	no
21	yes	no	>8	console	>5	yes
22	yes	yes	>8	pc	>5	no
20	yes	yes	<8	pc	>5	yes

B. Tools used.

WEKA is tried and tested open-source machine learning software that can be accessed through a graphical user interface, standard terminal applications, or a Java API. It is widely used for teaching, research, and industrial applications, contains a plethora of built-in tools for standard machine learning tasks, and additionally gives transparent access to well-known tool boxes such as sci-kit-learn, R, and Deep Learning 4.[9]WEKA is an innovative tool in the history of data mining and machine learning research communities. By putting efforts since 1994 this tool was developed by the WEKA team. WEKA contains many inbuilt algorithms for data mining and machine learning. It is an open-source and freely available platform-independent software. The people who are not having much knowledge of data mining can also use this software very easily as it provides flexible facilities for scripting experiments. As new algorithms appear in the research literature, these are updated in the software. WEKA has also become one of the favorite tools for data mining research and helped to progress it by making many powerful features available to all. The steps performed for data mining in WEKA are:

- Data pre-processing

- Attribute selection

- Classification (Decision trees)

- Results Visualization

C. Method of performance

The gaming dataset collected through an online survey consists of 1000 records and 7 attributes. The dataset is in CSV format i.e., an excel file. The following steps depict the execution process of the proposed system in brief:

1. Collect a gaming dataset using the survey method

2. Load the dataset in WEKA for implementation

3. Perform data preprocessing

4. Apply J48 and REP Tree algorithm

5. Observe and analyze results

V. RESULTS

The proposed algorithm uses WEKA.WEKA is a comprehensive open-source Machine Learning toolkit, written in Java. These functions provide a basic MATLAB interface to WEKA to allow the transformation of data back and forth and to use most of the features available in WEKA, such as training classifiers. By doing so the accuracy rate of the J48 algorithm has increased to a large extent as compared to the accuracy of the same algorithm in WEKA. The proposed algorithm works as: The arff data file is loaded from WEKA into MATLAB. Then the refining of the dataset is done. Later the J48 classifier is applied. In the end, the results are obtained that is the accuracy and error rate are calculated. A Decision tree is created using the J48 algorithm. The flow of the tree goes as follows Skill_level > Salaried, Platform > online gaming, Weekly_hours. The "yes"class leaf node identified in the decision tree are the group of players that can be given game recommendations.

Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
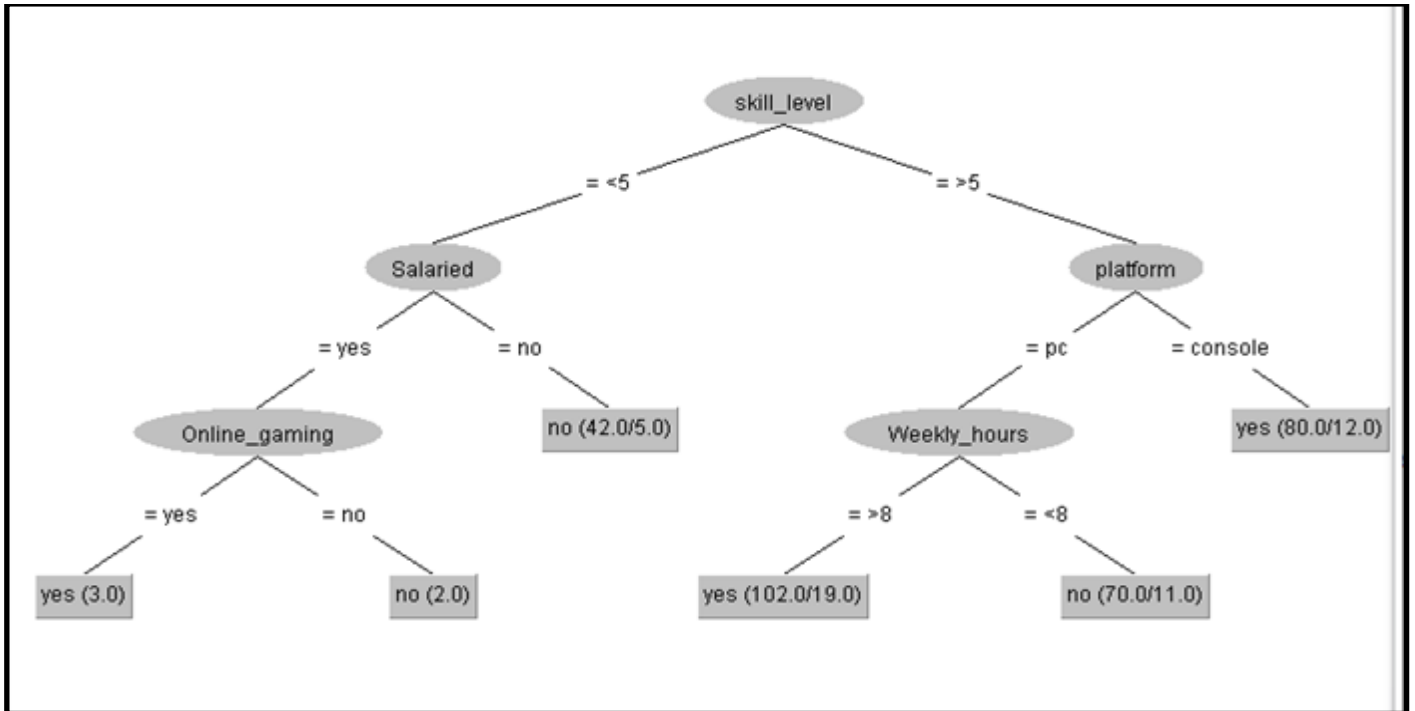**ISSN: 2278-0181**
NTASU - 2020 Conference Proceedings

Fig 3. Decision Tree created by using J48 Algorithm

Using the J48 algorithm decision tree is created. The players classified into yes category can be given game recommendation. The J48 algorithm considers all the attributes specified in the gaming dataset.
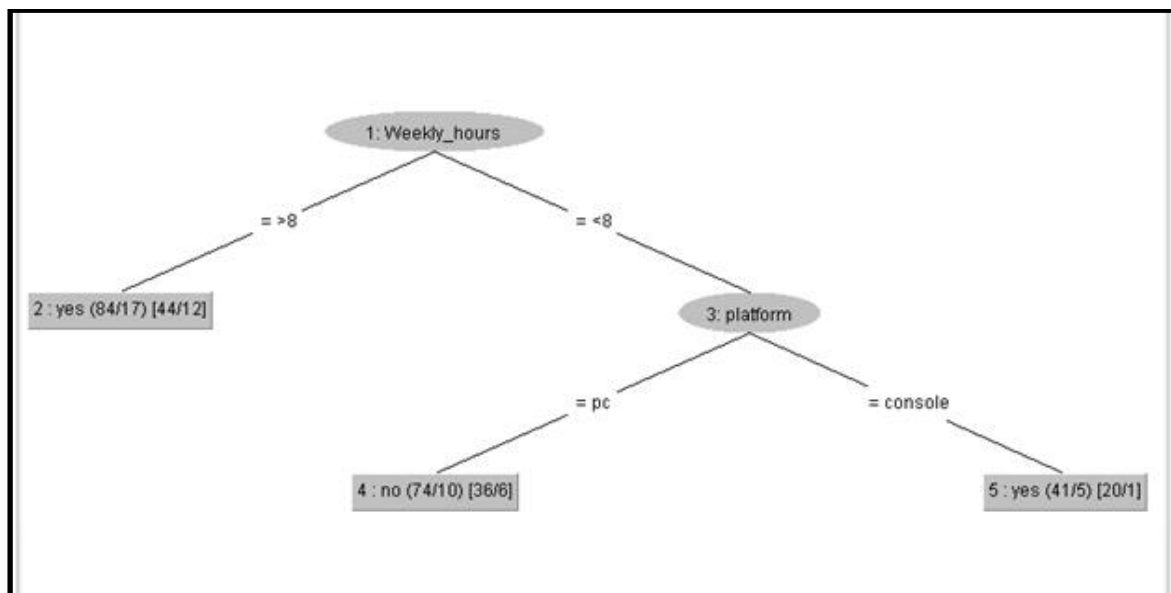

Fig 4. Decision Tree created by using REP Tree Algorithm

RepTree uses the regression tree logic and creates multiple trees in different iterations. After that it selects the best one from all generated trees. That will be considered as the representative. Hence a selected tree is shown in figure 4. Compared to the j48 algorithm the tree in more specified and only selected attributes are identified. The players classified into the yes category can be given game recommendations. This improves the game sales further improving the marketing factor.

Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NTASU - 2020 Conference Proceedings

VI CONCLUSION

Online Computer Gaming Industry is very suitable for Data Mining and Data Warehousing applications. Those applications can provide Management and Technical Team vital information and knowledge about their business and their customers, thus, aid them in making better business decisions of revenue making and cost saving.

REFERENCES

[1] Newzoo. (2020). The Global Games Market Will Generate $152.1 Billion in 2019 as the U.S. Overtakes China as the Biggest Market | Newzoo

[2] J. Han, M. Kamber and J. Pei, ─ Data Mining Concepts and Techniques, ‖ Waltham: Elsevier Inc, 2012.

[3] Rodrigues, Cajetan, and Nazneen Ansari."Better Game Design using Association Analysis",International Journal of Advanced Research in Computer Science 3, no. 3 (2012).

[4] Anagnostou, Kostas, and Manolis Maragoudakis."Data mining for player modeling in video games."In 2009 13th Panhellenic Conference on Informatics, (pp. 30-34). IEEE, (2009)

[5] Drachen, Anders, et al. "Game data mining."Game analytics. Springer, London, 2013. 205-253.

[6] Techopedia.com. (2020). What is Targeted Marketing? - Definition from Techopedia.

[7] Kaur, Gaganjot, and Amit Chhabra. "Improved J48 classification algorithm for the prediction of diabetes." International Journal of Computer Applications 98.22 (2014).

[8] Kalmegh, S. (2015). Analysis of WEKA Data Mining Algorithm REPTree, Simple Cart and Random Tree for Classification of Indian News. Available at: http://ijiset.com/vol2/v2s2/IJISET_V2_I2_63.pdf [Accessed 27 Jan. 2020].

[9] Cs.waikato.ac.nz. (2020). Weka 3 - Data Mining with Open Source Machine Learning Software in Java. [online] Available at: https://www.cs.waikato.ac.nz/ml/weka

[10] Vindel, Rafael, Héctor D. Menéndez, and David Camacho. "A survey in Convergence Technologies for Video Games using Data Mining."

[11] GeeksforGeeks. (2020). Basic Concept of Classification (Data Mining) - GeeksforGeeks. [online] Available at: https://www.geeksforgeeks.org/basic-concept-classification-data-mining.