Recommendation System: Gaming

Prasang Prajapati
001310428
Department of Computer Science
University at Albany, State University of New York
Albany, NY 12206 USA
pprajapati@albany.edu

Rajat Gupta
001308897
Department of Computer Science
University at Albany, State University of New York
Albany, NY 12206 USA
rkgupta@albany.edu

Abstract—**This paper describes my first experience in making a recommendation system for users using the open source Software PyCharm and using Twitter API. Recommendation systems have become drastically increased in different areas like Movies, Music, Books, Games etc. Eventually, a project itself focuses on recommending a game to users by gathering information from them about the played games in past. A recommendation system tells its user about what decisions will be beneficial for them. Using Twitter Data, this system recommends users that which game they should play next according to their taste.**

Keywords—recommendation; gaming; twitter; PyCharm.

I. INTRODUCTION

To recommend means something that you put forward with approval as being suitable for a role as per the given inputs and interests shown [2]. Here, our system recommends users a game which they should play by analyzing the data available in the system. There are many games in the market as well as there are many new games are arriving, usually people are confused that which game they should play next. Whether people will have a satisfaction after starting a game as to their taste or interests, this a demanding question in the world for us. Thus, here our system recommends users to play this game by entering the games a user likes to play generally. By analyzing user's input and available stored data of different users in the system, we have successfully made this system and our system has accuracy around in late 90's which is comparatively good for us as this is our first experience. Twitter Data collection, Analysis of Twitter Data Collection, Classification of Tweets using SVM and Association Rule are our division of the Methods in our system. Here, we have used Twitter Data because we know that Twitter is one of the most popular websites in the world. It has around 328 million active users in the 1st Quarter of 2017. At present, every second on average 7618 tweets are tweeted by the users. It is very much large amount of data. So, it is very easy to give good recommendation to users about the gaming. Thus, it has been used to provide services to the users.

II. RELATED WORKS

A. A researach conducted by PGDBA Group[4]

This researcher suggests that a recommendation system must be the accurate to find the most used the words count in the Twitter Data. A researcher has recommended their users a movie based on words used in the Twitter Data. Firstly, they have divided their methods into the various parts and worked on each part one by one which is the best approach for any recommendation system nowadays. Through Twitter Data they have first found the movies which are trending into the market and further implemented different algorithms for each method to recommend a movie to their users based on the words used in the tweets.

B. A recommendation engine using Association Rule by fellow from Marmara University, Instanbul, Turkey[3]

This project idea has been suggested by a fellow from the Marmara University in which researcher has made a recommendation engine using Association Rule Mining which is an accurate method to give the best recommendation. Based on the selected products in the cart on E-Commerce Website, researcher suggests users' the other products in which they are highly interested, and to get higher conversion rate.

C. Most popular recommendation system used by Youtube

As all know and every day we are into the system which enforces us to build the system like recommendation. We usually search one song or any video topic on YouTube and based on our search, it automatically gets the idea about our interest and our taste based on entered search words previously.

D. Google Search

Based on our recent and ongoing search words, google suggests us different links to open that websites. Moreover, they always take our words into the consideration and suggests users to go through such website where we can find the solutions.

E. A project on NHL Playoff Prediction by Brendan Hean from University of Victoria[6]

This is very interesting project in the sports area as it predicts the winner of a particular game based on the gathered

Fig. 2. Training the Neural Network

Total Network Error Graph
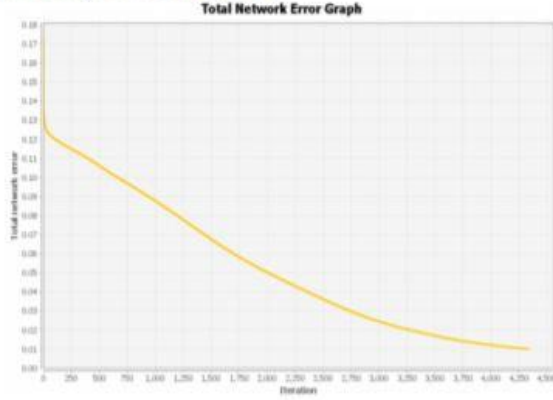
statistics over the whole season. In this project, they have tested their data on a variety of classifiers provided by scikit-learn in python. Researchers also have experimented with a Neural Network using a NetBeans based on development and environmental called Neuroph Studio. Also, tested data using Decision Tree and Support Vector Machine.

III. PROPOSED APPROACHES

In our system, we mainly have used many different approaches to get this project done to get the higher accuracy. Number equations consecutively. Our system mainly recommends users that which game they should play by entering the names of games which they have played before. A we know that if we enter the name of any rock music on YouTube then it gives us all the possible and the same category music to listen. Here, in our system we have mainly get the data from the Twitter and then analyzed the Twitter Data and have given the recommendation after applying the SVM to classify Tweets and Association Rule to get the best possible result. Now, we are going to explain all the approaches which we have implemented to get this project done very smoothly.

A. Twitter Data Collection

In our project, we haven't used any dataset but we have collected the Tweets from Twitter and used that tweets as our dataset. A we all know that Twitter is one of the most popular social networking sites in the real world and it has one of the highest users. As comparing with the dataset, Twitter has good accuracy because in twitter we get the user's opinions and real time data from the users. To get the Data from Twitter we have used the REST API. Collected Tweets classification has been mentioned in the Fig. 1.

Total Tweets	Unique	Training	Testing
11033	4869	1151	3718

Fig. 1. Classification of Tweets

Figure 1 concludes that we have total 11033 tweets collected from the Twitter using REST API. We have some difficulties in gathering tweets as we were not getting unique tweets because we have used App Scheduler. An App Scheduler helps us to generate Tweets from Twitter at every one hour automatically. We don't need to run the code and get the tweet at each and every hour. App scheduler is used for not to run the code at every hour. Through this approach, we were getting many tweets which are similar so we were in trouble and further we have approached a new method through which we are removing the same tweets from all the tweets. After doing that we have got total 4869 tweets which are unique. In all the process, we have used result_type in apisearch to get the recent tweets. By using the REST API, we were getting 97-105 tweets at one run. So instead of running the query manually every hour we used App Scheduler to get the tweets automatically at every hour.

B. Twitter Data Analysis

In the next steps, after getting the unique tweets we manually analyzed all the tweets and then we differentiated all the unique tweets into the positive and negative tweets to generate the training dataset. The number of positive and negative tweets are mentioned in the Fig 2 below.

Total Tweets	Unique	Training	Positive	Negative
11033	4869	1151	289	863

Fig. 2. Furter Classification of Tweets

As classified above in the table, after manually differentiated the unique tweets into the positive and negative tweets which are 289 and 863 respectively. After getting the training dataset, we used the SVM Approach to find positive and negative tweets. Therefore, we have designed our code through which it uses our positive and negative tweets and compare the new tweets with the previous tweets. Through SVM now new tweets are automatically differentiate the unique tweets into the Positive and Negative Tweets. Therefore, at each and every run our accuracy of SVM model is increased. The SVM Model Accuracy is increased at each run by exponentially.

C. Linear SVM (Support Vector Machine) [7]

The linear SVM approach is the best approach to get this task done.

$$\vec{w} \cdot \vec{x} - b = 0$$

----------------(1)

In the equation (1) w is the normal vector to the hyper plane. The parameter b divided by w determines the offset of the hyperplane from the origin among the normal vector w. By SVM Model, the model accuracy is increasing highly. By using the linear SVM we have got many advantages to get success in the Project. Firstly, through the regularization

parameter, user has avoided over-fitting. By using the kernel trick, we have built in expert knowledge about the problem via kernel and engineering. SVM is introduced by the convex optimization problem and for that there are so many efficient methods. Support Vector Machines are known as Supervised Learning Models in the machine learning with associated learning algorithms which analyzes the data used for classification and regression analysis, in these, for our project it was used for classification to differentiate the unique tweets into the positive and negative tweets. In our project, firstly the tweets were not labeled. So, in the beginning we have labeled the tweets and then we have applied the supervised learning algorithm for further upcoming new tweets [8].

D. Application of SVM Model on Analyzed Twitter Data

In the next steps, in this part we are describing the application and things which we have implemented using SVM. So, in our project, after classifying the unique tweets in positive and negative tweets, we have generated the top most 10 frequent words from our Twitter Data. In the results, we have got only one word which has highest frequency but we want whole game name. Therefore, we have faced a problem over here. We have got names like NBA, Theft, Grand, Infinite etc. But we need whole game name i.e. Call of Duty, Ghost Recon, NBA 2k17, Grand Theft Auto. So, to overcome this problem, we have approached another method which is explained below.

E. Association Rule Mining

In this innovative approach, we have first generated a Word Cloud of the collected Tweets from the Twitter using REST API.

Fig. 3.   Word Cloud of Collected Tweets

A discussed earlier we have generated the word cloud from the collected tweets which is you can see in the Fig 3. In further steps, we have analyzed the word cloud and we found top 10 most popular Games. From that we have selected total 5 top most popular games for our next recommendation by cross verifying on Steam Community.

In data mining, association rule is useful for predicting and analyzing the user's behavior. Association rule learning is a rule-based machine learning method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness[9].

Therefore, by the association rule mining we created small prototype. We have manually entered the game data the includes 5 most popular game that are played by some users. And with the help of that data, we have recommended a game for inexperienced users as to which game user should play next. Association Rule mining is the approach which has least memory consumption. Also, it has the easy implementation.

Through this approach, we have selected top most 5 popular games which we have manually entered into the system and then we recommend our user a new game as to which game they should play next. And we have 99.93% of accuracy of our Model which is more than enough to get the best result of a recommendation.

IV. SYSTEM DESIGN AND IMPLEMENTATION

In our system, we have used the Python Programming Language. We have used the PyCharm tool for our project. Moreover, we have used the libraries like Pandas, NumPy, Scikit-learn, matplot, pyplot etc. Moreover, we have used the SVM Model and Association Rule Mining Approach for our Project.

Steps of our Project Implementation:

1. Tweet Collection
2. Tweet Analysis (Positive and Negative Tweets)
3. Trained and Tested our Data using SVM
4. Predicted top 10 popular games
5. Divided tweets in Clusters
6. Word Cloud of Clusters
7. Analyzed Word Cloud
8. Created small prototype using Association Rule Mining

Here, above I have mentioned the steps that we have implemented in our project. In these steps, firstly we have gathered the tweets from the Twitter using REST API. We were getting same tweets at a time so we have implemented a python code which remove the same tweets from the total tweets. In addition to that, using REST API we were getting 97-105 tweets each hour so we have used result_type in apisearch to get the recent tweets and by using the app scheduler we are getting tweets automatically at each hour. In

the next steps, we have manually analyzed tweets and have got the positive and negative tweets from the collected unique tweets as shown in Fig 2.

```
Answer 2:-
2k17)
(nba
ops
auto
duty®:
black
theft
infinite
(grand
(injustice
```

Fig 4. Top Most Frequent Words – Output SneakPeek

In addition to that, we have trained our data using SVM Model and we have got accuracy in late 90's later. After training the data, we have tested our data and we have got the top most 10 popular game words as a prediction as shown in Fig 4.

Furthermore, we found that here we have got only one form the game. And we want whole game name i.e. Ghost Recon, Grand Theft Auto etc. Therefore, to overcome that problem we have used another innovative approach for our project. We have made total 10 clusters as shown in Fig 5 and from the biggest cluster we have mode a word cloud which is shown in Fig 3.

```
total number of clusters: 10

cluster 0 has 1680 tweets
cluster 1 has 2355 tweets
cluster 2 has 34 tweets
cluster 3 has 19 tweets
cluster 4 has 58 tweets
cluster 5 has 65 tweets
cluster 6 has 71 tweets
cluster 7 has 42 tweets
cluster 8 has 29 tweets
cluster 9 has 10 tweets
```

Fig 5. 10 Clusters – Output SneakPeek

By generating the word cloud, we have manually analyzed word cloud and then we have compared the most common game names with the names that we have got in the Fig 4. From the top 10 most popular games we have selected the 5 games and then we have used association rule mining for the further process.

In the next steps, we ask user to give input of the following games which are given below in the Fig 6 that are played by the users.

Then users will enter the played games and the game in which they are interested and liked the most. When users enter the games, they have played ten at that time users will be recommended the final game which they should play next. A satisfaction after playing a game is the best feeling I have ever got. Therefore, users should not be get dissatisfied after playing a new game therefore, in our system when users enter the games they have played then that data also be stored in the data and from that data our system is trained again and for the next user the previously entered data will be come into the consideration while recommending a new game to users.

In the Fig 6, we have a look and the user has entered the input as Dota 2 and FIFA for the played games because user has to select the game name from the given names which are as follows:

```
The following is the list of games
Dota 2
Fifa
cs
MW3
ghost reccon
nfs
Please enter from the above mentioned list: Dota 2, Fifa
```

Fig 6. User's input – Output SneakPeek

1. Dota 2
2. FIFA
3. CS
4. MW3
5. Ghost Recon
6. NFS

From these games user has entered the game names which are Dota 2 and FIFA. After entering these input, user has been recommended a Ghost Recon game to play next as per his interests and taste which is show in Fig 7.

Therefore, from the above Fig 7 we have recommended a new user a game which is Ghost Recon. From the given input as FIFA and Dota 2, we have recommended a game to user which is Ghost Recon.

As the final output, here we have built a console input-output to know the user that which game user to should play next for the better satisfaction.


```
The following is the list of games
Dota 2
Fifa
cs
MW3
ghost reccon
nfs
Please enter from the above mentioned list: Dota 2, Fifa


Your Entered Choice are Dota 2, Fifa



The following games is the next best game you should be playing
 ghost reccon
```

Fig 7. Recommendation to User – Output SneakPeek

V. FUTURE IMPLEMENTATION

A recommendation system which must be very accurate as after using the system, customer must be satisfied with the results in future. Therefore, as of our future implementation we will try to improve our results whatever we are getting. As of new we are giving 5 choices of games to the users but we are planning to give more choice to the users to enter as their input to get the good and more accurate recommendation. Moreover, we have implemented all the inputs and outputs in the Console but now we will implement that using Good Graphical User Interface by implementing HTML Code. Furthermore, we have recommended a game to the users from only 11033 tweets but as for future implementation we will try to gather more and more tweets and then will do analysis so that we will get the better and accurate recommendation to the users. In addition to that at present we are giving recommendation of the games which are played in the PlayStation but in next steps we will consider the games of the Xbox, PC games and Mobile games as well. A per the interests and likes of the users we will try to implement better efficient code in the system and will try to give the accurate recommendation to the users.

VI. CONCLUSION

A project report concludes that after entering the selected input in the system, user is getting an (near to 90%) accurate recommendation as per user's interest and taste. A recommendation system means a suggestion to their users by analyzing and considering their inputs of interests as the data. In our project, mainly we have used the Twitter Data to get the large amount of data from the user's recent Tweets. To get the recent tweets from the Twitter we have used the Result Type in API Search. Also, we were getting same tweets at each run. Therefore, we have implemented a code which removes the same tweets from the collected tweets and the number of tweets we have got that are shown in Fig 1. The problem we have faced in data collecting was that we needed to run our query for an hour which is so much time consuming. Therefore, we have implemented a code and used the App Scheduler to generate tweets at every hour automatically. After collecting tweets and from the data we have recommended a user a new game to play next.

I. REFERENCES

[1] Diving into Educational Robotics with Player/Stage Carlos Jaramillo Computer Engineering Department, City College of New York. Available: http://dreuarchive.cra.org/2009/Jaramillo/DREU/files/final_report.pdf

[2] The Application of Data-Mining to Recommender Systems J.
Ben Schafer, Ph.D. University of Northern Iowa. Available: http://faculty.chas.uni.edu/~schafer/publications/dmChapter.pdf

[3] A Recommendation Engine by Using Association Rules by Ozgur Cakir from Marmara University, Istanbul, Turkey. Available: http://www.sciencedirect.com/science/article/pii/S187704281203515X

[4] Movie Recommendation System Using Twitter Data. Available: http://courseprojects.souravsengupta.com/cds2015/movie-recommendation-system-using-twitter-data/

[5] Twitter Statistics. Available: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

[6] Data Mining Project Final Report NHL Playoff Prediction Brendan Heal, Meara Kimball, Renee Fung, and Abdelrahman Alenazi from University of Victoria. Available: http://web.uvic.ca/~afyshe/dm_projs/nhl_final_report.pdf

[7] G. C. Cawley and N. L. C. Talbot, Over-fitting in model selection and subsequent selection bias in performance evaluation, Journal of Machine Learning Research, 2010. Research, vol. 11, pp. 2079-2107, July 2010. Available: http://jmlr.csail.mit.edu/papers/volume11/cawley10a/cawley10a.pdf

[8] Kecman, Vojislav; Learning and Soft Computing — Support Vector Machines, Neural Networks, Fuzzy Logic Systems, The MIT Press, Cambridge, MA, 2001. Available: http://www.support-vector.ws/

[9] Piatetsky-Shapiro, Gregory (1991), Discovery, analysis, and presentation of strong rules, in Piatetsky-Shapiro, Gregory; and Frawley, William J.; eds., Knowledge Discovery in Databases, AAAI/MIT Press, Cambridge, MA.

[10] Post Graduate Diploma in Business Analytics – PGDBA Available: https://pgdbablog.wordpress.com/