



**Department of Computer Science and Engineering  
University of Asia Pacific**

# **Design and Simulation of a Novel Classification Framework for Separating Sentiment from Assorted Game Related Tweets**

**A thesis submitted in partial fulfillment of the requirements for the award of the  
degree of Bachelor of Computer Science and Engineering**

**Submitted by  
Syeda Tasmiah Islam  
Registration ID: 12201020  
Ahmad Al-Sajid  
Registration ID: 12201028**

**Supervised By  
Molla Rashied Hussain  
Assistant Professor  
Dept. of CSE, UAP**

**November 2016**

UNIVERSITY OF ASIA PACIFIC

DECLARATION OF UNDERGRADUATE THESIS / PROJECT PAPER AND  
COPYRIGHT

Title : Design and Simulation of a Novel Classification  
Framework for Separating Sentiment from Assorted Game Related Tweets

Academic Session: 2015/2016

We declare that this thesis entitled “*Design and Simulation of a Novel Classification Framework for Separating Sentiment from Assorted Game Related Tweets*” is the result of our own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : .....

Name :

Date :

Signature : .....

Name :

Date :

“I hereby declare that I have read this thesis and in my  
opinion this thesis is sufficient in terms of scope and quality for the  
award of the degree of Bachelor of Computer Science and Engineering”.

Signature :

Name of Supervisor :

Date :

This thesis is dedicated to my family for their endless support and encouragement.

# **ACKNOWLEDGEMENT**

First of all, we would like to thank the almighty ALLAH. Today we are successful in accomplishing our work with such ease because ALLAH gave us the ability, chance, and cooperating supervisor.

We would like to take the chance to reveal the gratefulness to Mr. Molla Rashied Hussein, our respected supervisor. Although he was always loaded with several other activities, he gave us more than enough time in this task. He gave also proper guidance and valuable advice to us. His comments and guidance helped us in preparing our thesis report.

Last but not the least, we are grateful to our family; who are, always supporting us in our every step of life.

# ABSTRACT

Twitter has been exceedingly a ubiquitous phenomenon with a thriving appearance in almost every developed country in the world. In this modern epoch, game is one of the most illustrious themes to tweet. Sentiment analysis is used to analyze the patent opinion of people by mining text units i.e. tweets. In this paper, strategy based categories of games were identified, top most countries according to users were detected, foremost usable language of the users was ascertained and advertisement from assorted game related tweets were filtered to retrieve sentiment tweets exclusively through simulation. In light of these findings, a new gestalt of sentiment analysis of game related tweets is also accomplished.

## TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	<b>DECLARATION.....</b>	<b>ii</b>
	<b>DEDICATION.....</b>	<b>iii</b>
	<b>ACKNOWLEDGMENT .....</b>	<b>iv</b>
	<b>ABSTRACT.....</b>	<b>v</b>
	<b>TABLE OF CONTENTS.....</b>	<b>vi</b>
	<b>LIST OF FIGURES.....</b>	<b>viii</b>
<b>1</b>	<b>INTRODUCTION.....</b>	<b>1</b>
1.1	Overview.....	1
1.2	Motivation.....	2
1.3	Problem Statement.....	3
1.4	Objective.....	3
<b>2</b>	<b>LITERATURE REVIEW</b>	<b>2</b>
2.1	Background Studies.....	5
2.1.1	Big data.....	5
2.1.2	Web 2.0.....	5
2.1.3	Business Analytics.....	6
2.1.4	Social Network Sites.....	7
2.1.4.1	Twitter.....	7
2.1.5	Sentiment Analysis.....	8
2.1.5.1	Virtual Games.....	9
2.1.6	Machine Learning.....	9
2.1.6.1	Natural Language Processing.....	10
2.1.6.2	Naïve Bayes Classifier.....	10
2.2	Related Works.....	13

<b>3</b>	<b>METHODOLOGY AND DESIGN OF THE FRAMEWORK</b>	<b>15</b>
3.1	Overview.....	15
3.2	Structured Data Analysis.....	16
	3.2.1 Game Categories.....	16
	3.2.2 Language Comparison.....	18
	3.2.3 User Authentication.....	18
	3.2.4 Comparison between Countries .....	19
3.3	Unstructured Data Analysis.....	19
	3.3.1 Classification using Naïve Bayes Classifier.....	20
3.4	Software Tools.....	21
	3.4.1 Python.....	21
	3.4.2 Tweepy.....	22
	3.4.3 TextBlob.....	22
	3.4.4 Matplotlib.....	22
	3.5.5 MongoDB.....	23
<b>4</b>	<b>RESULTS AND DISCUSSIONS.....</b>	<b>24</b>
4.1	Results.....	24
4.2	Analysis of Results.....	27
4.3	Limitations.....	28
<b>5</b>	<b>CONCLUSION.....</b>	<b>29</b>
5.1	Discussion.....	29
5.2	Future Works.....	29
	<b>REFERENCES</b>	



## LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
2.1	Social Network Sites.....	7
2.2	Naïve Bayes Classification.....	12
3.1	Analytic Approach to Classify Sentiment and Advertisement .....	15
3.2	Game Categorization.....	17
3.3	Language Comparison.....	18
3.4	User Authentication.....	18
3.5	Comparison Between Countries.....	19
3.6	Sample Training Data.....	20
3.7	Work Process of Naïve Bayes Classifier.....	21
4.1	Strategy Based Popular Game.....	24
4.2	Top 10 Countries According To The Number Of Users Who Tweet Regarding Games.....	25
4.3	Top 5 Languages According To The Number Of Users Who Tweet Regarding Games.....	26
4.4	Advertisement vs. Sentiment Tweets.....	27

# CHAPTER 1

## INTRODUCTION

### 1.1 Overview

Online social media have acquired astonishing global augmentation and become a resounding fashion of the neoteric vogue by drawing attention from miscellaneous types of users. This is predominantly a digital platform that enriches the social norm overhead virtually. Social relations among people are being expanded and prolonged by content-sharing, social interaction, and also proper cooperation. Facebook, Twitter, LinkedIn, Google+ etc. are the most detailed social network sites.

Social media is a fusion of sociology and technology. People can put themselves out in several good ways over there [18]. Social Network Sites have already made different types of scopes for the users. Locating facilitation, seeking new jobs, promoting products or companies hence Social Media Marketing (SMM) are the most common prominences of social media lately.

At the present time, online social media are getting more popularity for innovating and updating information into a dynamic forum incessantly [1]. An innate feature of many social network sites that differentiates them from mainstream news media sites is the information network evolved by the users following their preferred generators of information [2]. For that microblogging sites are now became the first choice of the most of the peoples in the world to share their activities or interests. On an average three billion users are using microblogging sites for daily updates. Cost-effective business promotion method makes these sites more desired and easeful. Twitter is one of the microblogging

websites which was launched back in 2006. And with hundred million users Twitter was gained more popularity in 2012 for the first time. In fact, Twitter is a network of information constructed of one hundred and forty (140)-character messages, namely Tweets. This is a flowing and easy way to express the latest news that matters the most as well as the phenomenon users typically care about. This is the best platform to commence the peregrination for finding out and following several engrossing Twitter accounts to access resources that users prefer frequently [6].

Crafting a post or tweet is also another spontaneous affair. Eventually, Twitter is a swift content provider of real time updates. Users can transmit and receive updates via the Twitter website, SMS, RSS (receive only), emails or a third-party application such as “Tweetie,” “Twitterrific,” and “Feedalizr.” Twitter declared the establishment of the “Twitter Trust & Safety Council on February 9, 2016 [19]. Some researchers confide that Twitter can be suitable for predicting particular and certain outcomes such as: Elections, Business Strategies, Political Protests and Challenges as well as human behavior and also their actions. Researchers apply different kinds of sorted schemes and processes for their investigation. Sentiment Analysis is one of those investigations. Sentiment analysis is a comparatively recent developing sector which covenants with extracting and assenting user opinion automatically [7]. Feature based sentiment analysis comprehend attribute extraction, sentiment prophecy, categorization and volitional summarization modules [20].

Sentiment analysis prefers using the Natural Language Processing tasks at numerous levels of granularity [7,8]. Generally, Natural Language Processing, Machine Learning methods, Statistics are handled to identify, extract the sentiment contents or substances of a text units (blogs, user reviews, comments etc.). This analysis is cited as Opinion Mining as well. For redacting the sentiment analysis or opinion mining, researchers need both structured and unstructured data.

## **1.2 Motivation**

Motivation is having the encouragement to do something. The term motivation is the act of stimulating someone or oneself to get a desired course of action, to push the right button to get desired action. The process of motivation involves needs, drives and goals. Need is behind most of the actions of a man. Better facilities, more pay,

recognition, opportunities for promotion etc. are some of the needs of the people. Motivation is a process of stimulating people to act for the purpose of achieving desired goals.

Several sentiment analytics on different topics has been done like Elections, Soccer Leagues etc. Not much work has been done on virtual games related sentiments. This research will unearth important information by analyzing virtual game related tweets.

### **1.3 Problem Statement**

As problem statement describes the issues that need to be addressed by a problem-solving team and should be created by them before they try to solve the problem. The primary purpose of a problem statement is to focus the attention of the problem-solving team. However, if the focus of the problem is too narrow or the scope of the solution too limited, the creativity and innovativeness of the solution can be stifled.

This research will help to design and run simulation of a unique framework that can Classify advertisements and sentiments.

### **1.4 Objective**

Microblogging websites have evolved to become a source of varied kind of information. This is due to nature of microblogs on which people post real time messages about their opinions on a variety of topics, discuss current issues, complain, and express positive sentiment for products they use in daily life. Many times, these companies study user reactions and reply to users on microblogs. One challenge is to build technology to detect and summarize an overall sentiment.

In this thesis, design of a novel classification framework is constructed for separating sentiment from assorted game related tweets containing advertisements, simulation of the developed framework is run and analysis of the result is presented to validate the performance of the developed framework.

## CHAPTER 2

### LITERATURE REVIEW

#### 2.1 Background Studies

The research interest based on Social Network Sites are increasing swiftly [7]. Sentiment analysis is a relatively new area, which contracts with extracting user opinion automatically. Microblogging site Twitter is most significant for this mechanism. A combination of large data set of game related tweet collected from Twitter, sentiment analysis is useful for this survey. For this Natural Language Processing, Machine Learning, Naïve Bayes Classifier have committed as tremendous approach.

##### 2.1.1 big data

Extremely large data set that describes a holistic approach of information management strategy [22]. This can be used for disclosing trends, patterns etc. by analyzing them computationally. Big data provides some features like:

*Volume:* The amount of generated and reserved data. Big data elicits enormous volumes of data [29]. Organizations collect data from several sources like business transactions, social media and information from sensor or machine-to-machine data [30].

*Velocity:* Big Data Velocity pacts with the motion at which data flows in from different sources. The sources can be business processes, machines, networks and human interaction with things like social media sites, etc. The flow of data is voluminous and continuous.

*Variety:* The genre and raw characteristic of the data. Data can be both structured and unstructured. Data appears in the form of emails, photos, videos, monitoring devices,

PDFs, audio, etc. This variety of unstructured data makes complications for storage, mining and analyzing data.

*Variability:* There are several potential meanings for Variability. When data holds different extreme values, it presents a statistical problem to deduce what to do with outlier values and whether they contain a new and important signal or are just noisy data [32].

*Veracity:* Big Data Veracity refers to the biases, noise and abnormality in data. This is so common that there are inherent discrepancies in all the data collected.

Big data analysis is a challenge. Big data analytics is the method of gathering, organizing and analyzing big data to detect patterns and other information. Big data craves high performance analytics. Analyzing big data allows analysts, researchers, and business users for various types of benefits. Such as: Big data technologies like Hadoop and cloud-based analytics bring notable cost reduction advantages. This is also useful for time reduction. So, this is faster and better to make smart decisions. With the ability to measure vendee needs and amusement through analytics comes the power to give customers what they want. Consequently, new product development and optimized offerings are also possible. The business area getting the most attention relates to increasing efficiency and optimizing operations.

### **2.1.2 web 2.0**

World Wide Web (WWW), is actually a scheme of Internet servers that support specially formatted documents. Web 2.0 is simply an improvement of the existing internet and has been born out of the increased use of the World Wide Web. Web 2.0 is simple, light, social, flexible [22]. Some of the features of web 2.0 are:

*Rich User Experience:* dynamic volume that is reactive to user input. Traditional web is built with HTML and CSS, CGI and had been offered as a static page. On the other hand, Web 2.0 uses Ajax is nothing but Asynchronous JavaScript and XML that is a method of building interactive applications for the user requests immediately.

*Folksonomy:* free categorization of information. This is also known as Social tagging. Such as, the photo sharing site Flickr etc. [22].

*User as Contributor:* Web 2.0 user contributes to the content by means of Evaluation, Review & Commenting.

*Long Tail:* The traditional web was like a retail business. But in web 2.0 the product is not sold directly but offered as a service on demand basis and income is generated as monthly fee and pay per consumption.

*Basic Trust:* web 2.0 the contents are made attainable to share, reuse, redistribute and edit where the contents are protected under Intellectual Property Rights in traditional web.

*Dispersion:* In web 2.0, the content delivery uses multiple channel include file sharing.

Search, authoring, tag, signals etc.- these are the concepts that upgrades the web 2.0 from the traditional web. Users are getting huge welfare from the web 2.0 swiftly [37, 39]. The prime advantages of web 2.0 are:

- a) Web 2.0 is modern and efficient.
- b) Web 2.0 is the ease of communication.
- c) Assists in technological development
- d) Web 2.0 is a better way for users to collaborate with one another to use the ideas and resources of many individuals in the educational environment.
- e) Several learning styles.
- f) Usage of podcasting, blogging, tagging, RSS, Social networking makes modern life more spontaneous.

### **2.1.3 business analytics**

Business Analytics (BA) is used to obtain insights that apprise business decisions and may be used to automate and optimize business processes. Business analytics makes comprehensive use of statistical analysis and predictive modeling [22]. When the business goal is determined then an analysis methodology is chosen and data is attained to support the analysis. It depends on sufficient volume of high quality of data. There are different types of business analytics.

- a) Decisive Analysis: Support human decisions with visual analytics.
- b) Descriptive Analytics: Achieve insight from historical data with clustering, reporting etc.
- c) Predictive Analytics: Use machine learning for predictive modelling.
- d) Perspective Analytics: Suggest decision using simulation, optimization etc.

*Benefits of Business Analytics:* this is a faster process. Can be got insight into customer behavior. Remove the presumptions terms [38]. Business Analytics includes statistical analysis, data mining, multivariate testing. Almost every managers and executive can use BA systems. Different types of users in different sectors can be found such as: students, organizations, companies, small films etc.

#### **2.1.4 Social Network Sites**

Social network sites (SNSs) are progressively drawing the observation of academic and industry researchers [24]. SNSs have redacted a wide variety of technical properties. This is a substance of the modern world that can change minds, catch more business, increase sales or build business. Social networks have common themes of fact to share. Two-third of the world's internet citizens visit SNSs for almost 25% of the internet time [4,17]. Facebook, Twitter, LinkedIn, Google+, YouTube, Instagram, Flickr, Reddit etc. are top most Social Network Sites of the current time. These sites can be used as a law enforcement aid. This may progress the social improvement of a person. Students can explore topics that they are interested in by using online social network sites. This is one of the great advantages. Many companies use social networks like Twitter and Facebook to connect with customers and prospective clients [17]. Currently, people get more updates about the world through social networking sites than televisions and newspapers. And now one of the newest feature is video chat on almost all SNSs.



Figure 2.1: Social Network Sites

##### **2.1.4.1 Twitter**

User-provided information is a sturdy source or origin of opinion and it can be crucial for a diversity of applications that crave understanding the public viewpoint about a conviction [1]. As users post real time reactions almost about everything on micro-blog



website twitter, this poses modern and different types of challenges. Because of easy access of published posts twitter is deliberated as one of the largest datasets of user generated facts. A survey of Twitter discovered, a more distinct social induction of passing beside messages that encircle a hyperlink is a prospect of mutual conjoining by the disciples or followers [10]. Twitter characterized by some features it has conceded. Such as:

*Tweet:* Crafting a post on Twitter. Each tweet allows 140-charaters for sharing their personal information, interests, news, photos, blogs etc.

*Mention:* Mentions in a tweet imply that the post refers another user. For this user, must be referenced to a username. Mentioning on twitter is much easier than the other social network sites like Facebook.

*User:* A user has to be registered him/herself first for twitting with a valid identification. For this user, must be appointed a username.

*Hashtag:* A hashtag is a metadata tag used on micro-blogging services or social network sites. Hashtags are used to indicate the concern of a tweet to a definite object. For example, #game, #ClashClans etc.

*Retweet:* This is nothing just forward or repost of a specific earlier post. The retweeting is regarded a strong gadget for publicizing content.

*Follower:* To be up to date about the latest news, information, interests etc. it refers to the users to follow other users or interesting substances.

*Privacy:* Twitter gives the local option to the users to fix if his or her posts will be apparent to everyone or only to the selected Twitter followers [1]

### **2.1.5 sentiment analysis**

Fact-finding or tracking out sentiment in Twitter is a significant assignment and varies notably from detecting sentiment in operable text like forums and blogs. Sentiment analysis can be arranged into three main categories: knowledge-based techniques, statistical methods, and hybrid approaches [11]. Knowledge bases both list obvious affect words and random words to particular emotions. Statistical methods support the constituents from machine learning. The sentiment uses classifiers trained on both twitter sentiment as well as reviews from the data sets. The larger part of Sentiment analysis and Twitter Sentiment analysis methods unearth sentiment based on a feature set [1]. The outcome will be more accurate and mathematical on text that is similar to original training

data. The most repeatedly used semantic features are sentiment words, concepts, negation and opinion words. Opinion words mentions to phrases that are characterized as suggestive of thought(s) and sentiment words are introductory of both positive and negative sentiment. Twitter sentiment analysis is authentic for sentiment analysis is negation [14, 15].

#### **2.1.5.1 Virtual games**

Virtual game mainly computer based simulated game. Here a person can experience being in a three-dimensional environment and interact with that environment during a game. Sight and sound are the two sensors that need to play virtual games. Virtual games provide mainly two types of virtual games, such as massively multiplayer online games and offline virtual games. Virtual spaces can serve a variety of research and educational goals. This may be useful for examining human behavior. Some virtual worlds have off-line, real world components and applications. Nowadays virtual games are familiar in our daily life. About all types of people from every profession are connected with this. This may be one of the leisure time activities in the modern world. Computer games are control the technological world. There are different kinds of platforms for the virtual games: PC, console, handheld, arcade, web browsers etc. And according to classification games are like casual games (puzzle), serious games (virtual reality), educational games etc. are available. Virtual games are also important for online business. Various companies earn huge amount of money from this industry.

#### **2.1.6 machine learning**

“Learning denotes changes in a system that enable a system to do the same task more efficiently the next time”- Herbert Simon. To understand and efficiency of human learning, disclose new things, build software agent, reduce an important aspect of intelligent behavior etc. machine learning is essential [8]. Machine learning has several types of categories:

*Supervised learning:* The training data are known of a selected class.

*Unsupervised learning:* The training data are not known of a selected class.

*Semi-supervised learning:* A combination of supervised and unsupervised learning.

*Reinforcement learning:* Allows machine to learn its behavior based on feedback from the environment.

Machine learning techniques are important for feature extraction. Classification, clustering, prediction, input distribution are the methods of Machine learning [15]. Machine learning and statistics are nearly related fields. The computational analysis of machine learning algorithms and their performance is a branch of theoretical computer science. For the best performance in the context of generalization, the complexity of the hypothesis should match the complexity of the function underlying the data. Decision Tree learning, Association Rule learning, Deep learning, Bayes Network, Artificial Neural Network, Learning classifier system etc. [9, 14]. are the different approaches of Machine learning. Some of its applications are: bioinformatics, brain-machine interface, classifying DNA sequence, computer vision, game playing, machine perception, robot locomotion, sequence mining, software engineering, search engine and so on.

#### **2.1.6.1 Natural Language Processing**

Natural Language Processing is a sub-field of Artificial Intelligence. Aimed to build intelligent system that can interact with human being. It refers languages that spoken by people. Applications for processing large amount of texts require NLP expertise [9]. Segmentation, disambiguation, stemming, contextual analysis, machine translation, speech to speech translation, sentiment analysis are the major areas of research and development for Natural Language Processing [14]. Modern NLP algorithms are based on machine learning, especially statistical machine learning [8]. Many different classes of machine learning algorithms have been applied to NLP tasks. Recent research has increasingly focused on unsupervised and semi-supervised learning algorithms.

#### **2.1.6.2 Naïve Bayes Classifier**

The Naïve Bayes Classifier greatly simplify learning by assuming that features are independent given class. Although independence is generally a poor assumption, in practice, Naïve Bayes often competes well with more sophisticated classifiers [41, 42]. Making a probabilistic model of data within each class Naïve Bayes classifier is convenient. This is based on Bayes' theorem with strong (naïve) independence postulations between the features. Bayes theorem:

Given a hypothesis  $h$  and data  $D$  which bears on the hypothesis;

$$P(h|D) = \frac{P(D|h) P(h)}{P(D)}$$

Here,

P(h): prior probability of h

P(D): independent probability of D

P(D|h): likelihood

P(h|D): posterior probability

Based on Bayes theorem, MAP (Maximum A Posterior) hypothesis for the data can be computed. And hypothesis is represented by h. The Bayes classification is:

$$P(C|X) \propto P(X|C) P(C) = P(X_1, \dots, X_n|C) P(C)$$

Using this formula Naïve Bayes classification can be deduced. Different types of parameters of Naïve Bayes classifier are available. Such as:

Gaussian naive Bayes: Dealing with continuous data, a typical assumption is that the continuous values associated with each class are distributed according to a Gaussian distribution.

$$p(x = v|c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(v-\mu_c)^2}{2\sigma_c^2}}$$

Multinomial naive Bayes: With a multinomial event model, samples represent the frequencies with which certain events have been generated by a multinomial ( $P_1$ - $P_n$ ) where  $P_i$  is the probability that event i occurs.

$$p(\mathbf{x}|C_k) = \frac{(\sum_i x_i)!}{\prod_i x_i!} \prod_i p_{ki}^{x_i}$$

Uses of Naïve Bayes classifier:

- Text classification
  - Classify web page by topic
  - Information extraction
  - Internet filter
- Spam filtering
- Hybrid recommender system
- Online application
  - Simple emotion modeling

Naïve Bayes approach:

- Build the vocabulary as the list of all distinct words
- The word in the vocabulary become the attributes, assuming the classification is free of the position of the words
- Every document in the training set is record
- Train the classifier based on the training data set
- Evaluate the result on the test data

Advantages:

- Easy to implement
- Requires small amount of training data set to estimate the parameters
- Good result obtains in the most cases

The Naïve Bayes classifier is appealing because of its simplicity, elegance, robustness. This is widely used for text classification. The large number of modifications are introduced by data mining, statistical, machine learning, pattern recognition. Credit scoring, marketing application, image processing, search engines etc. are the example that required to use Naïve Bayes classifier.

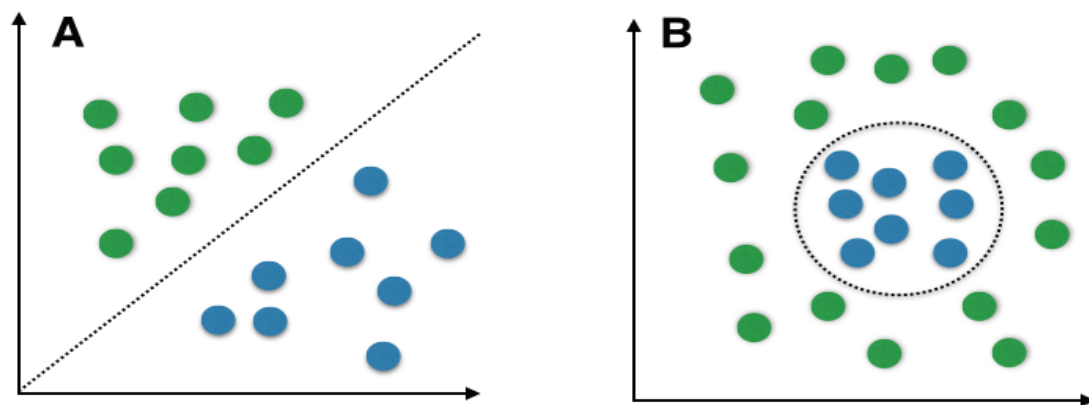


Figure 2.2: Naïve Bayes Classification

## 2.2 Related Works

The increase in social media networks has made sentiment analysis a popular research area, in recent years. In Zimbra et al. [43] review, has conducted Sentiment Analysis in Twitter regarding Brand using Feature Engineering and the Dynamic Architecture. This study presents an approach to brand-related Twitter sentiment analysis using feature engineering and the Dynamic Architecture for Artificial Neural Networks

(DAN2). The approach addresses challenges associated with the unique characteristics of the Twitter language, and the recall of mild sentiment expressions that are of interest to brand management practitioners. They demonstrated the effectiveness of the approach on a Starbucks brand-related Twitter data set. And also, compared the proposed approach to the performances of two state-of-the-art Twitter sentiment analysis systems from the academic and commercial domains.

Yang Yu et al. [44] has collected real-time tweets from U.S. soccer fans during five 2014 FIFA World Cup games (three games between the U.S. team and another opponent and two games between other teams) using Twitter search API. They used sentiment analysis to examine U.S. soccer fans' emotional responses in their tweets, particularly, the emotional changes after goals (either own or the opponent's). they found that during the matches that the U.S. team played, fear and anger were the most common negative emotions and in general, increased when the opponent team scored and decreased when the U.S. team scored. Anticipation and joy were also generally consistent with the goal results and the associated circumstances during the games. Furthermore, they found that during the matches between other teams, U.S. tweets showed more joy and anticipation than negative emotions (e.g., anger and fear) and that the patterns in response to goal or loss were unclear. Finally, this research revealed that sports fans use Twitter for emotional purposes and that the big data approach to analyze sports fans' sentiment showed results generally consistent with the predictions of the disposition theory when the fanship was clear and showed good predictive validity.

H. Wang et al. [45] has analyzed public sentiments towards the 2012 U.S. presidential candidates. This paper described a system for real-time analysis of public sentiment toward presidential candidates in the 2012 U.S. election as expressed on Twitter, a micro-blogging service. Emerging events or news are often followed almost instantly by a burst in Twitter volume, providing a unique opportunity to gauge the relation between expressed public sentiment and electoral events. In addition, sentiment analysis can help explore how these events affect public opinion. While traditional content analysis takes days or weeks to complete, the system demonstrated analyzes sentiment in the entire Twitter traffic about the election, delivering results instantly and continuously. It offered the public, the media, politicians and scholars a new and timely perspective on the dynamics of the electoral process and public opinion.

Research with the combination of Twitter and sentiment analysis is cutting-edge. However, not much research has been carried out on virtual games related sentiments. Therefore, novelty lies in this research. This innovation may be the shared experience of a new cultural phenomenon or the subjective perception of an individual.

## CHAPTER 3

### METHODOLOGY AND DESIGN OF THE FRAMEWORK

#### 3.1 Overview

Social network sites have revolutionized the way in which people interact with each other. Microblogging sites like Twitter is a spectacular and striking tool for

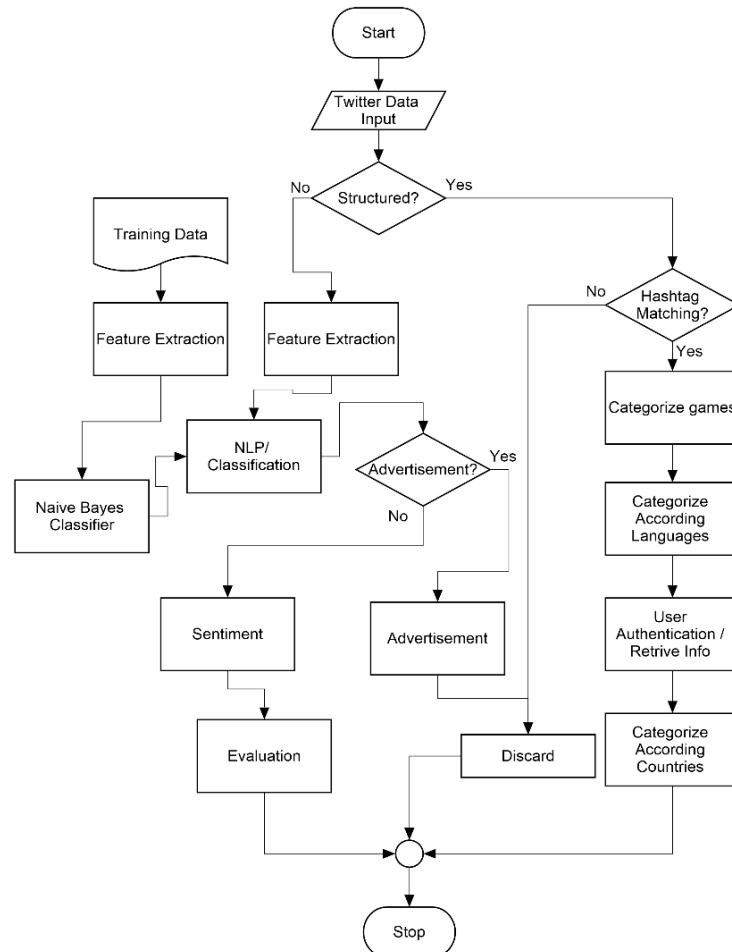


Figure 3.1: Analytic approach of sentiment and advertisement classification

acquiring new thoughts or insight. As random tweets, had been collected from Twitter, different data types were retrieved. Generally, data are labeled (structured) but the text units are unlabeled (unstructured). For this reason, structured and unstructured data were



analyzed separately. The analytical steps of sentiment and advertisement classification are shown in figure-3.1. Here Twitter data are taken as input. Then it will be decided to be in structured or unstructured data.

### **3.2 Structured data analysis**

Structured data is simple to sort. Sentiment analysis provides three types of approaches and one of these three is statistical method. The statistical method is chosen to run the simulation work. Statistical Natural Language Processing uses stochastic, probabilistic, and statistical methods to evolve some of the difficulties including extended sentences which are highly oracular when processed with realistic semantic and conveying numerous possible analyses. Structured data is mined for finding categories of games, countries with most users and major language used in Twitter. In figure-3.1, if the input data are structured then it performs for hashtag matching. And when the hashtags are matched feature extracted as categorize games, categorize according to language, retrieve information, categorize according to language also. And if the hashtags are not matched then these will be discarded.

#### **3.2.1 Game categories**

Games are nothing but a calculated strategy or approach. Rules, competitors and achievements are must criteria for any kind of game. In this modern world, game is one of the best current affairs for posting in various social network sites like Facebook, Twitter etc. [16]. For the research, random tweets regarding games were collected from Twitter. Afterward, those tweets were arranged into different categories. Such as:

*Massively Multiplayer Online (MMO):* This type of game is an online game which is susceptible of supporting large numbers of players simultaneously in the same case. For example, World of Warcraft, Star Wars: The Old Republic, Planet Calypso etc.

*Simulations:* This kind of games attempts to copy different activities from real life in the form of a game for various motives as well as training, analysis, or prediction, e.g. Game Dev Story, Goat Simulator, Infinite Flight Simulator etc.

*Adventure:* For over 30 years, adventure games have been the most story-driven computer game group. The contestant plays a fantasy role in an episodic adventurous feat story. Such as, Crashland, Dead Effect 2 etc.

*Real-Time Strategy (RTS):* Real-time strategy's games already elapsed the popularity of turn-based strategy computer games. The participants position and maneuver units and structures under their control to secure areas of the map and/or destroy their opponents' assets. Like, Total War Battles: Kingdom, Clash of Clans.

*Puzzle:* The Room 2, Monument Valley etc. are puzzle games that are refer to exercise the brain by fitting different pieces together.

*Action:* Adventure games are about exploration and big worlds to look through. Need For Speed Rivals, WARP. Action is probably the spacious and most isolated genre in gaming.

*Browser games:* A browser-based game or online game is played completely within a web browser rather than through a video game or other tackle. Roblox, Animal Jam, Sword Saga etc. are examples of web browser game.

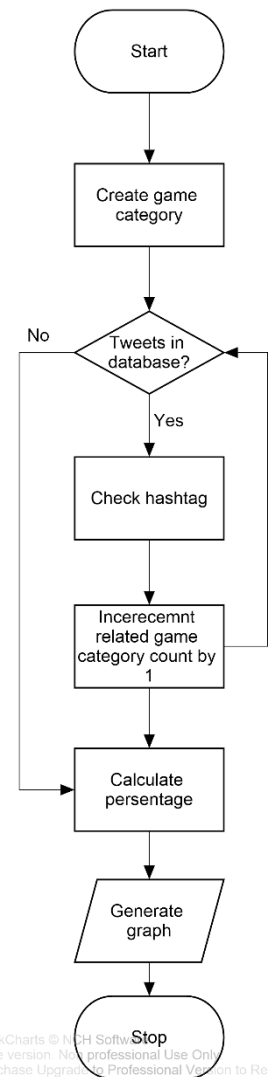


Figure 3.2: Game categorization

*Sports:* An athletic recreation requiring physical strength and skill often of a competitive nature virtually. For example: FIFA 14, PES 2014 etc.

*Shooter:* The main fact of shooter game is that it can be from any perspective. Shooter games are an ample action genre where the player shoots target as a primary gameplay mechanic, such as: Hitman, Sniper, Modern Combat.

*Role-Playing Game (RPG):* Pokémon Go, 7 Mages, Desktop Dungeons etc. are the Role-Playing games. Games of any main genre, except RPG, that incorporate 'light' elements taken from role-playing games, such as customization options for characters or vehicles, leveling and upgrade systems with different perks and effects, abilities and upgrade system that provide room for different gameplay styles and so on.

The games were categorized in above mentioned categories. Then a list of this games categories hashtags was created. Every hashtag was checked to match these hashtags and was categorized according to it. After finishing categorization, percentage was calculated and a graph was plotted to display the percentage of the tweets according to the game categories.

### 3.2.2 Language Comparison

To see which languages are mostly used, first a list of used languages in collected tweets was created. After that, the “lang” entity of every tweet was checked and the related language count was increased by one. When all the tweets were checked,

the percentage was calculated and a graph was created to show the most used top five languages.

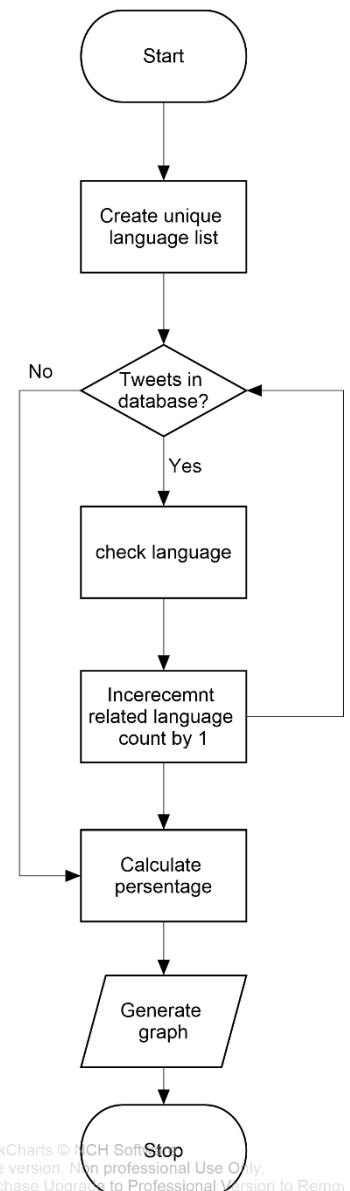


Figure 3.3: Language Comparison

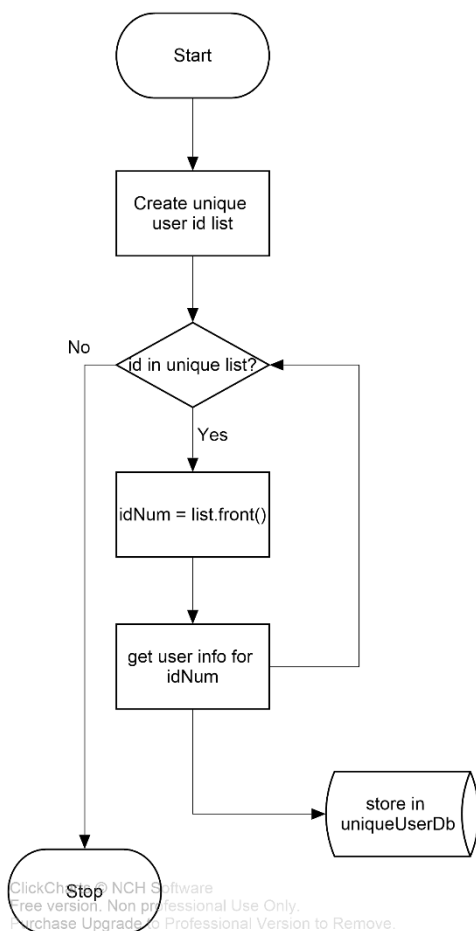


Figure 3.4: User Authentication

### 3.2.3 User Authentication

To get the authentic users data from collected tweets, a unique list of “user.id\_str” collected from the tweets. Then every single user’s “user” entity was searched from the database that contains tweets and stored to a new database.

### 3.2.4 Comparison of countries according to user count

After getting the authentic users from the tweets, a country list was created according to users’ time zone or location. Then every users’ location or time zone was checked to detect the country from the user was.

## 3.3 Unstructured data analysis

For the sentiment analysis of the unstructured data, Natural Language Processing was conducted and Machine Learning process was used to detect advertisement for separating from the personal (sentiment) tweets.

*Natural Language Processing (NLP):* NLP enables computers to derive meaning from human or natural language input. It has some major tasks, namely Automatic condensation, generating a readable and understandable concise of a shred of text.

*Oration Analysis:* This semantic includes a number of connected tasks.

*Machine translation:* Automatically interpret text unit from one human language to another. This is a complex process.

*Natural Language Understanding (NLU):* NLU involves the recognition of the intended semantic from the multiple realistic semantics that can be executed from a natural language expression.

*Information Extraction:* This is apprehensive in general with the release of semantic information from text. It covers the relationship extraction as well.

*Machine Learning:* Machine Learning is about learning structure from data. This is intimately related to computational statistics. This is sometimes mixed up with data mining. It has supervised and unsupervised approaches to learn. It provides some

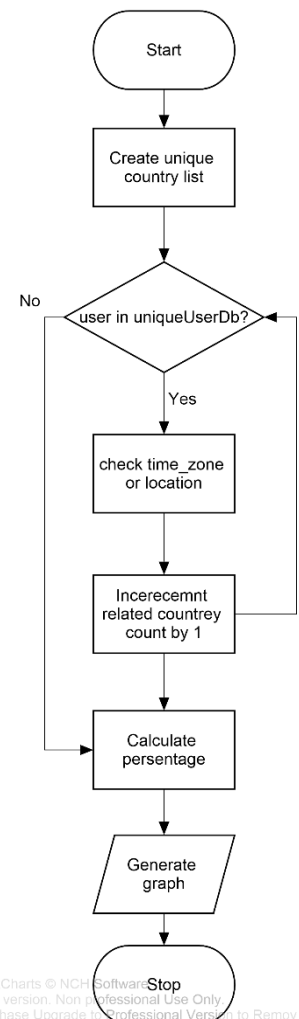


Figure 3.5: comparison of countries

processes to work such as classification, clustering, regression, density estimation etc. Machine learning and Statistics are nearly related fields. This study uses Machine Learning to find out the statistical outcome.

*Naïve Bayes Classifier:* This classifier is one of the simplest probabilistic classifiers. Naïve Bayes Classifier can be trained efficiently and effectively in a supervised learning environment comprising of several kinds of probability models. In this research work, Naïve Bayes Classifier was used for classification.

### 3.3.1 Classification of game related tweets into sentiment or advertisements using Naïve Bayes Classifier

At first, a set of texts from random tweets were collected from the database. Then the tiny urls were removed as they had no impact on the text. After that, they were tagged with two words, either ‘advertisement’, or ‘sentiment’. They were saved in a CSV file. Then the Naïve Bayes classifier was trained to recognize the pattern of advertisements and sentiments with learning using this csv data set. Then, every textual part of the tweets was tested according to this classifier to detect if the text was advertisement or not. After that, a comparison graph was plotted. A manually created testing set of data proved that the accuracy of the classifier was 76.02%.

@DuncanBannatyne @carolvorders Clash of clans th10 war base 2016 anti 2 star:A437 #ClashOfClans	advertisement
I found this image, could be useful for all #PokemonGO trainers! B475	advertisement
#GTAOnline #BossDonz CHOPPER CHOP	sentiment
@ChannelStarWars ESTONIAN ROUGE ONE POSTER!!! #starwars	advertisement
@MountainDew So I saw this across the store, you can say I'm Lazer focused...?? #badhumor #Titanfall2	advertisement
@RonCafc Epic In Every Sense of the Word.	advertisement
@Titanfallgame will split screen be in the game? #Titanfall2	advertisement
and I are adding glitter to the #StarWars universe! #sorrynotsorry @missingwords	sentiment
brand new gta5 glitch how to lower your car past competition suspension #GTAOnline #gta5	advertisement
Clash of clans th10 war base 2016 anti 2 star: B423 #ClashOfClans	advertisement
do give damn good trailer, and this is no exception B428s	sentiment
Estonian Rouge One poster!! #starwars #rougeone B252	sentiment
Even Grandmother Willow gets in on the action. #StarWars #ShipWars	advertisement
Final #starwarsrogueone trailer is out. Could this be the best #StarWars film ever? It looks just incredible!	advertisement
is doing a Flip Knife Blue Steel #Giveaway! Come Join!B20 #csgo #TurboNation	advertisement
is giving out a \$20 Steam/Battlenet Halloween gift! Click this link B326 #halloween #warcraft #steam #giveaway	advertisement

Figure 3.6: Example of training data

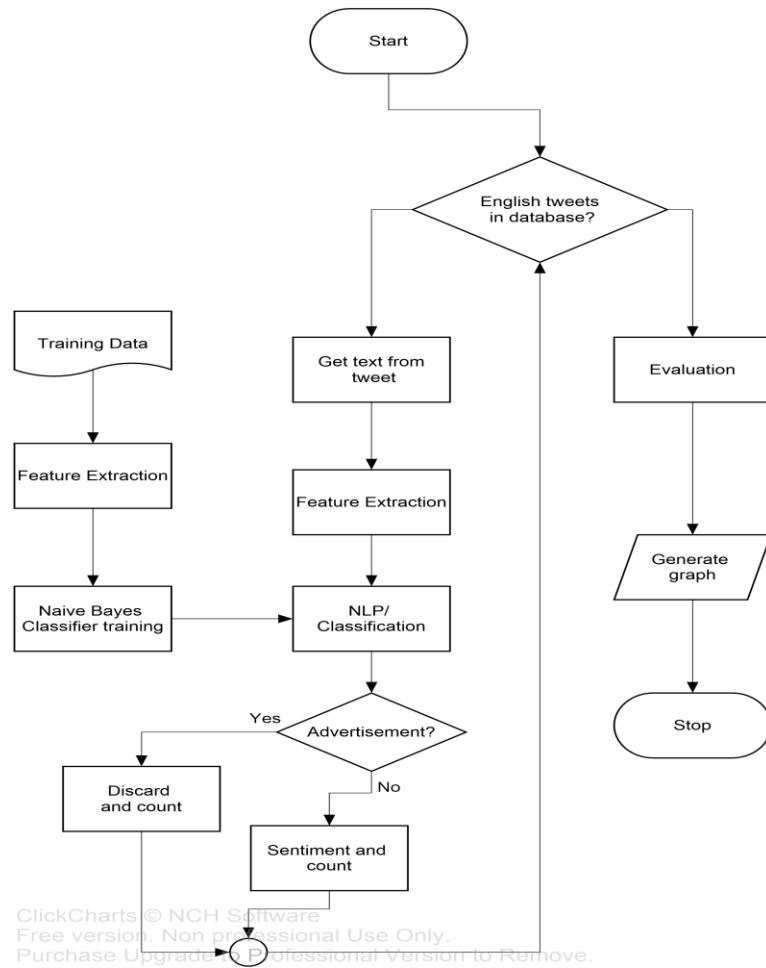


Figure 3.7: Naïve Bayes Classifier work process

### 3.4 Software Tools

Various tools are available for data mining purposes. The selection of choosing tools depends on users' needs and comfortability, availability etc. Some renowned tools are Weka, RapidMiner, SAS Enterprise Miner etc. These tools need less programming knowledge as they are pre-programmed. besides, MATLAB, R, Programming languages like Python, Java are also used for the same purposes. We choose Python and its libraries for our work. We also used MongoDB database to store our data. We need some perquisites to run the programs. Though some libraries have dependencies on other libraries, we only discussed about the libraries we used for our work.

### **3.4.1 Python**

Python is a powerful programming language. Universally, Python has gained a reputation because of it's easy to learn. The syntax of Python is designed to be easy to read. It has significant popularity in Scientific computing. The memory used to process “Big Data” or a bulk amount of data has more importance. In this case, Python works fluently. As many amazing works, has been done to improve Python's performance, it is suitable for high-performance computing.

### **3.4.2 Tweepy**

Twitters OAuth endpoints are used to connect users to Twitter and send secure, authorized requests to the Twitter API.

The Twitter REST APIs provides programmatic access to read and write Twitter data. Author a new Tweet, read author profile and follower data, and more. The REST API identifies Twitter applications and users using OAuth; responses are available in JSON.

The Twitter Streaming APIs continuously deliver new responses to REST API queries over a long-lived HTTP connection. Receive updates on the latest Tweets matching a search query, stay in sync with user profile updates, and more. If your application is rate-limited for over-polling the REST APIs the Streaming APIs may be a good solution for your needs.

To use these APIs from Python programming language, a library “Tweepy” is maintained to obtain the data. There are more similar libraries.

### **3.4.3 TextBlob**

There are many libraries in Python to work with text data like NLTK, pattern etc. but they need logical implementation of algorithms. As we only need to demonstrate the results of using text processing algorithm, not to implement or develop them, we used a library named TextBlob. It is a library for processing textual data. It provides simple API for diving into common natural language processing (NLP) tasks such as parts-of-speech tagging, sentiment analysis, classification and more. TextBlob internally uses NLTK, and pattern to implement the language processing algorithms.

### **3.4.4 Matplotlib**

Matplotlib is a plotting library for the Python programming language. It provides an object-oriented API for embedding plots into applications using general-purpose GUI

toolkits like wxPython, Qt, or GTK+. There is also a procedural ‘pylab’ interface based on a state machine (like OpenGL), designed to closely resemble that of MATLAB.

### **3.5.5 MongoDB**

MongoDB is a NoSQL database program. It is free and open-source cross-platform program. It is a document-oriented database. MongoDB uses JSON-like documents with Schemas. As Twitter APIs returns data in JSON format and MongoDB suits for this format, MongoDB was used in this thesis program.



## CHAPTER 4

### RESULTS AND DISCUSSIONS

#### 4.1 Results

Among these aforementioned categories of total 77 games categorized in ten categories, the most desired strategies of games were ascertained according to statistics which are tweeted most of the time on twitter. Massively Multiplayer Online games are the most exoteric and 43.25% people already tweeted about this on Twitter. As opposed to, Simulation category is in the lowest of the scale of popularity with 0.25% after puzzle category that stands with 0.35%.

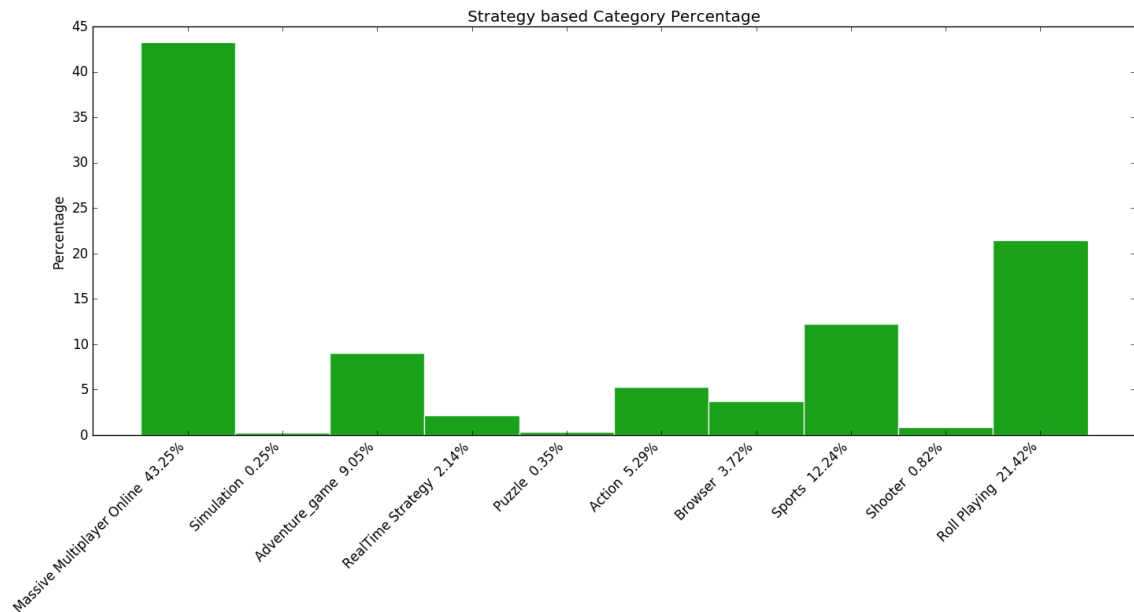


Figure 4.1: Strategy based popular game

Pokémon go, Battle Champ and other Role-Playing games are in second position with 21.42% popularity. Thereafter, Sports (12.24%), Adventure games (9.05%), Action

game (5.29%), Browser games (3.72%) are in the shallower position methodically. Real-Time strategy (2.14%) and Shooter (0.82%) are contiguous to the downcast position. Still these types of games are mostly tweeted.

Considering more games related tweet facts, top countries according to users is one of them. Among 97 countries and 188434 users, United Kingdom (14.78%) is in the top most position. No other countries have that number of users. Moreover, the rest of countries have sundry users according to the statistical approach. Such as, United States have only 7.80% of users who tweet on games. Ecuador (2.89%) and Russia (2.49%) are the lowest in topmost countries. The rest of the countries are The Netherlands (5.96%), Spain (5.92%), France (5.65%), Japan (4.84%), Brazil (4.46%), Greece (3.95%) with the passable number of users.

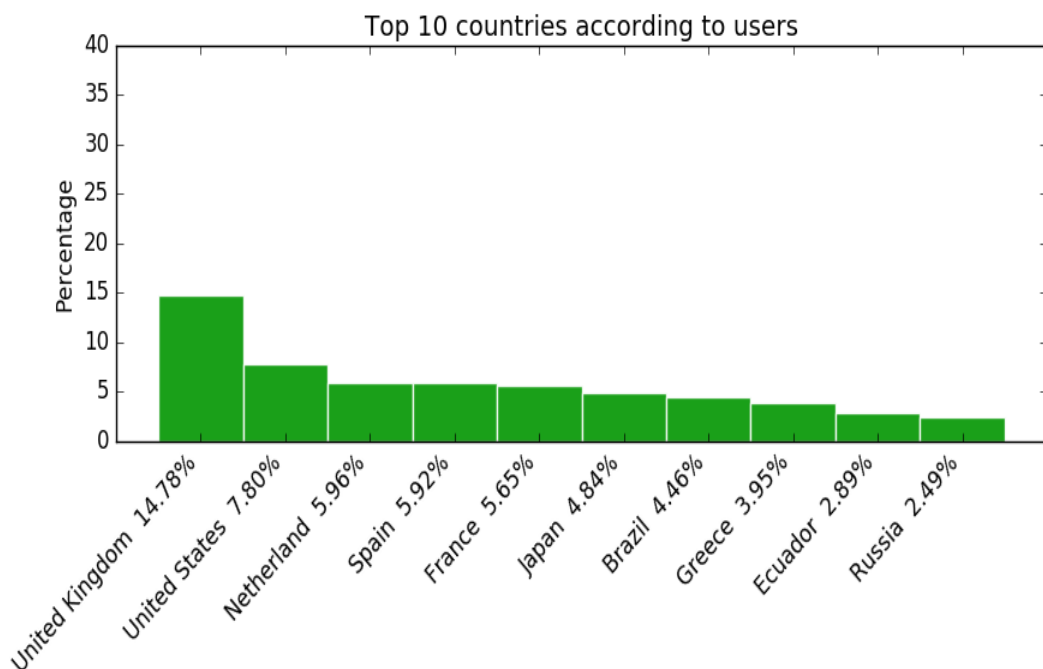


Figure 4.2: Top 10 countries according to the number of users who tweet regarding games

Language is another important segment of game related tweets. Dynamic users and spectators are reflected in the platform's language disruption. And obviously English (73.75%) is on the foremost position. Spanish (7.32%), French (5.17%), Japanese (4.86) and Portuguese (2.33%) follow after English. Total tweets were 353671 of 64 languages.

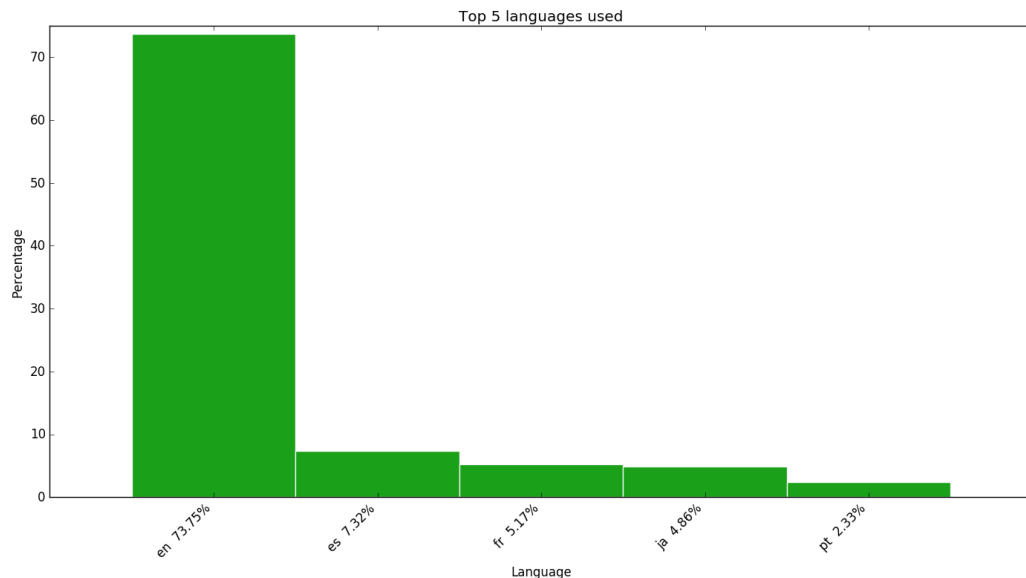


Figure 4.3: Top 5 Languages according to the number of users who tweet regarding games

Everyday about 500 million of tweets are tweeted. Among them all are not same in types. Some tweets are sentiment and rests of them are advertisement. Specially, the retweets are seemed as advertisement. For detecting advertisement and separate them from sentiments, Natural Language Processing task is used, Naïve Bayes classifier is applied into both sentiment and advertisement. As text units of tweets are not labeled (unstructured), they were segregated from the structured data in the beginning. NLP techniques were implemented to automate sentiment analysis on large collections of texts. This task allows movement to a more sophisticated understanding of sentiment. Using Naïve Bayes Classifiers with NLP on a training data set, this statistical survey provides various outcomes of tweets from Twitter, such as: total tweets collected were 353,671 from 188,434 users, among which 329,316 were related to games. In them, 246,281 were in English Language. With the relevant training data, applying Naïve Bayes Classifier, it was detected that 66.33% of the tweets was advertisement and the rest 34.67% were sentiment tweets.

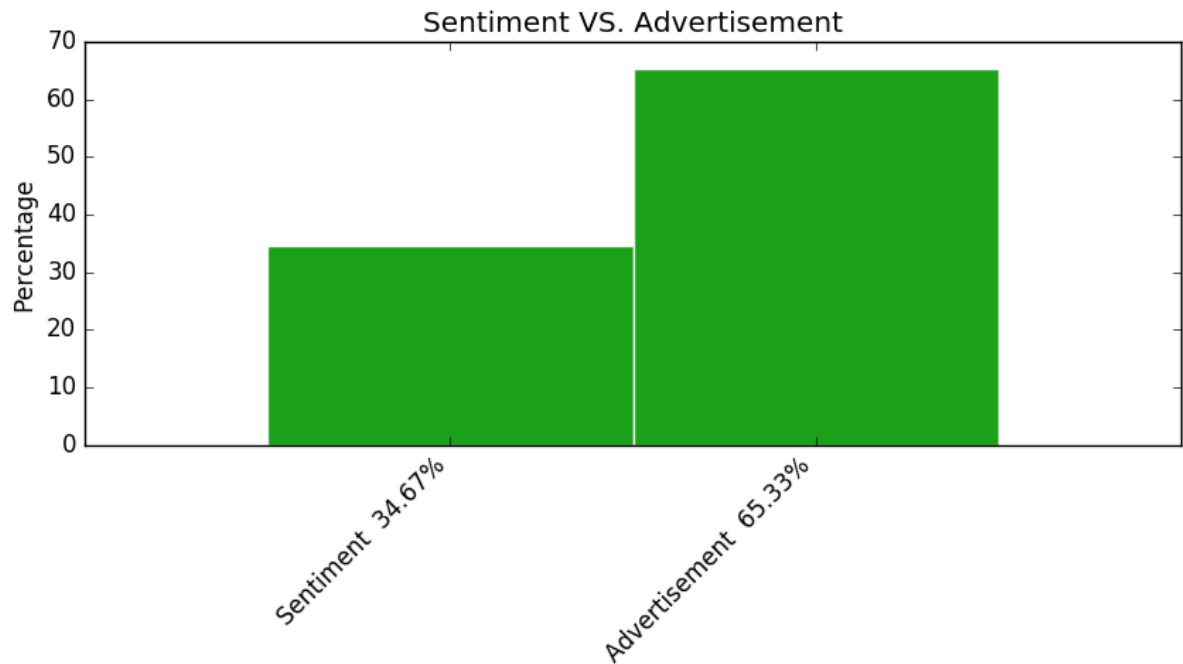


Figure 4.4: Advertisement vs. Sentiment tweets

## 4.2 Analysis of results

Gamers are impatient and eager souls, always keen to look ahead at what's coming down the line rather than be content with what they already have in their hands. As soon as a new console is released, some gamers start thinking about the next generation of hardware. As soon as a new game is released, some gamers start opining on what needs to be fixed in the sequel. Gamers need news. They probably want to know the latest goings-on in the industry. they want rumors, release dates, and trailers. Opinion pieces and round-ups of the best games in a certain genre. Twitter is a great resource for news, and gaming is no exception to that rule.

From this work, it can be evolved that most of the game related tweets are about Massively Multiplayer Online (MMO). Rests of the games concerned tweets are slightly vicinal of MMO.

As English is the most widely used language in the world, most of the social network sites (SNSs) use English as their default Language. Moreover, maximum users don't require any other languages than English to use social network sites (SNSs). English is the most used language for tweeting. In this case, English has no close contender in any way.

Now a day's games are one of the greatest ways of entertainment. On the contrary, most of the games need some expenditure. Citizens of the first world countries are more able to spend money on games than the second or third world countries. As it has been seen, most of the tweets are tweeted from the first world countries. So, it can be decided that they are the prime consumer of the gaming industry.

Naïve Bayes has been studied since the 1950's. Though it is an old algorithm, it has more accuracy than recent algorithms in some cases of textual mining. In this work, it worked greatly as it had accuracy of 76.02%. Naïve Bayes Classifier is confident to identify the advertisement from the sentiment with accuracy. As the training data set was generated manually, this result can be improved providing enhanced training data set.

### **4.3 Limitations**

After observing and classifying all these tweets, the countries or the time zones of some of the users cannot be distinguished simultaneously. As some users located their countries or time zones, several users gave either geolocation or time zones and some particular users did not give any of them during they tweet.

At the present time, some games are made with mixed-categories. Therefore, separating them into specific types of categories gives inconsistent outcomes. The accuracy of entire consequence relies on the training data set. Hence, improving training data set would give better outcome.

As this work is related with big data, it needs fast computing and much more storage capacity. The more data can be collected; the more accuracy will be deduced. But, human behavior is unpredictable, so, the accuracy or outcome may vary time to time.

## **CHAPTER 5**

### **CONCLUSION**

#### **5.1 Contributions of Research**

Twitter is a short broadcast medium of blogging. Recently the research interest for analyzing tweets is increasing rapidly as more people express their sentiment through Twitter. This research reveals an overview on the recent updates of game related tweets. This also elicits that game related advertisements can be detected and separated from sentiment. Research outcome can help shaping up the game development in particular.

#### **5.2 Future Works**

Naïve Bayes classifier was exceedingly appropriate for this simulation work. The accuracy of the outcome is about 76.02%.

We generally categorized tweets according to game categories. But it has found that 34.67% tweets contain sentiment, rest are advertisements. To get the real sentiment of people, categorization or other statistical methods can be implemented on this tweets only.

This research will be a supplementary for the start-up companies who will run gaming industry, will help to decide popular games and their categories. This paper will also help to predict the current trends in gaming industry.

This research successfully segregates the advertisement from the sentiment. In future, the sentiment analysis from the human perspective will be accomplished.

## REFERENCES

- [1] A. Giachanou, F. Crestani, “Like It or Not: A Survey of Twitter Sentiment Analysis Methods,” *ACM Computer Survey* 49(2), June 2016, p. 28.
- [2] Mahmoudreza, Przemyslaw, Isabel, Krishna, Manuel, “On the Efficiency of the Information Networks in Social Media,” *WSDM’16*, San Francisco, CA, USA, February 22–25, 2016.
- [3] R. BosaghZadeh, A. Goel, K. Munagala, and A. Sharma, “On the Precision of Social and Information Networks,” In *Proceedings of the first ACM Conference on Online Social Networks*, ACM, 2013, pp. 63–74.
- [4] N. A. Christakis and J. H. Fowler, “Social Network Sensors for Early Detection of Contagious Outbreaks,” *PloS one*, 5(9): e12948, 2010.
- [5] H. Kwak, C. Lee, H. Park, and S. Moon, “What is Twitter, a Social Network or a News Media?” In *Proceedings of the 19th International Conference on World Wide Web*, 2010, pp. 591–600.
- [6] Getting started with Twitter [Online]. Available: <https://support.twitter.com/articles/215585>.
- [7] J. Spencer and G. Uchyigit, “Sentimentor: Sentiment Analysis of Twitter Data,” School of Computing, Engineering and Mathematics University of Brighton, Brighton, BN2 4GJ [Online]. Available: <http://sentimentor.co.uk>.
- [8] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, “Sentiment Analysis of Twitter Data,” Department of Computer Science, Columbia University, New York, NY 10027 USA, 2011.
- [9] A. Bermingham and A. Smeaton, “Classifying Sentiment in Microblogs: is Brevity an Advantage?”, *ACM*, 2010, pp. 1833–1836.
- [10] A. Holton, K. Baek, M. Coddington, C. Yaschur, “Seeking and Sharing: Motivations for Linking on Twitter,” *Communication Research Reports* 31 (1), 2014, pp. 33–40.
- [11] E. Cambria, B. Schuller, Y. Xia, C. Havasi, “New Avenues in Opinion Mining and Sentiment Analysis,” *IEEE Intelligent Systems* 28 (2), 2013, pp. 15–21.
- [12] R. Stevenson, J. Mikels, T. James, “Characterization of the Affective Norms for English Words by Discrete Emotional Categories,” *Behavior Research Methods* 39 (4), 2007, pp. 1020–1024.
- [13] S. M. Kim, E. H. Hovy, “Identifying and Analyzing Judgment Opinions,” *Proceedings of the Human Language Technology / North American Association of Computational Linguistics Conference (HLT-NAACL)*, New York, NY, 2006.
- [14] G. Amati, M. Bianchi, and G. Marcone, “Sentiment Estimation on Twitter,” In *Proceedings of the 5th Italian Information Retrieval Workshop (IIR’14)*, 2014, pp. 39–50.
- [15] N. Aston, J. Liddle, and W. Hu, “Twitter Sentiment in Data Streams with Perceptron,” *Journal of Computer and Communications*, 2014, pp. 11–16.
- [16] The Most Engaging Topics on Social Media in 2015 (so far) [Online]. Available: <http://www.mavrck.co/2015s-most-engaging-topics-on-social-so-far>. [17] X. Hu, L. Tang, J. Tang and H. Liu, “Exploiting Social Relations for Sentiment Analysis in Microblogging,” In *Proceedings of the 6th ACM International Conference on Web Search and Data Mining (WSDM’13)*, ACM, New York, NY, 2013, pp. 537–546.

- [18] Pros & Cons of Social Media <https://sites.google.com/site/bhsdigicitizenship/social-media/pros-cons-of-technology>.
- [19] Twitter's corporate blog, "Announcing the Twitter Trust & Safety Council" <https://blog.twitter.com/2016/announcing-the-twitter-trust-safety-council>.
- [20] M. Z. Asghar, A. Khan, S. Ahmad, F. M.Kundi, "A Review of Feature Extraction in Sentiment Analysis" Journal of Basic and Applied Scientific Research, J. Basic. Appl. Sci. Res., 4(3)181-186, 2014.
- [21] Hu, M., and Liu, B. 2004. "Mining Opinion Features in Customer Reviews.", AAAI'04, 2004.
- [22] H. Chen, R. H. L. Chiang, V. C. Storey, "BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT", MIS Quarterly Vol. 36 No. 4, pp. 1165-1188/December 2012.
- [23] Miller, K. 2012a. "Big Data Analytics in Biomedical Research," Biomedical Computation Review (available at <http://biomedicalcomputationreview.org/content/big-data-analyticsbiomedical-research>; accessed August 2, 2012).
- [24] D. M. boyd, N. B. Ellison, "Social Network Sites: Definition, History, and Scholarship", October 2007, DOI: 10.1111/j.1083-6101.2007.00393.x
- [25] J. Trujillo, "Business Intelligence 2.0: a General Overview", First European Business Intelligence Summer School (eBISS), Ec. Centrale, Paris, 2011.
- [26] BUSINESS INTELLIGENCE: CONCEPTS, COMPONENTS, TECHNIQUES AND BENEFITS - JAYANTHI RANJAN - Journal of Theoretical and Applied Information Technology, 2009.
- [27] Ee P. LIM, H. Chen, G. Chen, "Business Intelligence and Analytics: Research Directions", ACM Transactions on Management Information Systems, Vol. 3, No. 4, Article 17, Publication date: January 2013.
- [28] J. Horkoff, A. Borgida, J. Mylopoulos, D. Barone, L. Jiang, E. Yu, D. Amyot, "Making Data Meaningful: The Business Intelligence Model and its Formal Semantics in Description Logics".
- [29] Barbara W., Thilini A., David D., Michael G., Babita G., "The Current State of Business Intelligence in Academia: The Arival of Big Data", Communications of the Association for Information Systems
- [30] Andreas Et Al, "Comparing and Combining Predictive Business Process Monitoring Techniques", Transactions on Systems, Man, and Cybernetics: Systems
- [31] Rafi Ahmad Khan, "KDD for Business Intelligence", Journal of Knowledge Management Practice,
- [32] Li Z., Ling L., Lian D., "Business intelligence in enterprise computing environment", InfTechnolManag (2012), DOI 10.1007/s10799-012-0123-z
- [33] Two Cross Corporation, "Introduction to Data Mining and Knowledge Discovery", Third Edition
- [34] Jenifer H., et al, "Making Data Meaningful: The Business Intelligence Model and its Formal Semantics in Description Logics"
- [35] Mariam Adedoyin-Olowe et al, "A Survey of Data Mining Techniques for Social Media Analysis"
- [36] Surajit C. et al, "An Overview of Business Intelligence Technology", DOI :10.1145/1978542.1978562



- [37] Nabil Sultan, "Knowledge management in the age of cloud computing and Web 2.0: Experiencing the power of disruptive innovations", *International Journal of Information Management*
- [38] Ee Peng LIM et al, "Business Intelligence and Analytics: Research Directions", [http://ink.library.smu.edu.sg/sis\\_research](http://ink.library.smu.edu.sg/sis_research)
- [39] Julie M. Sykes et al, "Web 2.0, Synthetic Immersive Environments, and Mobile Resources for Language Education", *CALICO Journal*, 25(3), p-p 528-546.
- [40] Jaideep S. et al, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data", *SIGKDD Explorations*, Jan 2000, Vol-1, Issue-2
- [41] K. Ming Leung, "Naïve Bayes Classifier", Available: <https://cis.poly.edu/~mleung/FRE7851/f07/naiveBayesianClassifier.pdf>
- [42] I. Rish, "An empirical study of the naïve Bayes classifier", Available: <https://researchgate.net> [pdf].
- [43] D. Zimbra ; M. Ghiassi ; S. Lee, "Brand-Related Twitter Sentiment Analysis Using Feature Engineering and the Dynamic Architecture for Artificial Neural Networks", DOI: 10.1109/HICSS.2016.244.
- [44] Yang Yu, X. Wang, "World Cup 2014 in the Twitter World: A big data analysis of sentiments in U.S. sports fans' tweets", <http://dx.doi.org/10.1016/j.chb.2015.01.075>
- [45] H. Wang, D. Can, A. Kazemzadeh, F. Bar, S. Narayanan, "A system for real-time Twitter sentiment analysis of 2012 U.S. presidential election cycle", *Proceeding ACL '12 Proceedings of the ACL 2012 System Demonstrations*, pp. 115-120

