

PROFESSIONAL TRAINING REPORT

at

Sathyabama Institute of Science and Technology

(DEEMED TO BE UNIVERSITY)

Submitted in partial fulfillment of the requirements for the award of

Bachelor of Engineering Degree in

Computer Science and Engineering

By

MOPURI ANNIE SOWMYA

(Reg. No. 38110325)

PANABHAKAM VANDANA

(Reg. No. 38110615)



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade “A” by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI,

CHENNAI – 600119, TAMILNADU

MARCH 2022



SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with Grade “A” by NAAC

(Established under Section 3 of UGC Act,1956)

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI, Chennai -
600119**

www.sathyabama.ac.in



**DEPARTMENT OF COMPUTER SCIENCE AND
ENGINEERING**

BONAFIDE CERTIFICATE

This is to certify that this Professional Training Report is the bonafide work of **Mopuri Annie Sowmya(Reg No.38110325), Panabhakam Vandana (Reg No. 38110615)** who have done the project work as a team who carried out the project entitled “**SOCIAL MEDIA ANALYSIS USING MACHINE LEARNING**” under my supervision from nov 2021 to MAR 2022.

Internal Guide

DR.R.AROUL CANESSANE.,M.E., (PH.D).,

Head of the Department

Submitted for Viva voce Examination held on_____

Internal Examiner

External Examiner

DECLARATION

I, **MOPURI ANNIE SOWMYA (RegNo.38110325), PANABHAKAM VANDANA (Reg No. 38110615)** hereby declare that the professional Training Report on **“SOCIAL MEDIA ANALYSIS USING MACHINE LEARNING”** done by me under the guidance of **DR. R. AROUL CASSANNE, M.E., PH.D.**, at Satyabhama Institute of Science and technology is submitted in partial fulfillment of the requirements of the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

We are pleased to acknowledge our sincere thanks to Board of management of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

We convey our thanks to **Dr. T. SASIKALA M.E., Ph.D., Dean of the Department, Department of Computer Science and Engineering** for providing us the necessary support and details at the right time during the progressive reviews.

We convey our thanks to **Dr. L. LAKSHMANAN., M.E., Ph.D., Head of the Department, Department of Computer Science and Engineering** for providing us the necessary support and details at the right time during the progressive reviews.

We would like to express our sincere and deep sense of gratitude to our Project Guide **DR. R. AROUL CASSANNE M.E., PH.D.**, for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

We wish to express our thanks to all Teaching and Non-teaching staff members of the Department of **COMPUTER SCIENCE AND ENGINEERING** who were helpful in many ways for the completion of the project.

ABSTARCT

Recently, social media is playing a vital role in social networking and sharing of data. Social media is favored by many users as it is available to millions of people without any limitations to share their opinions, educational learning experience and concerns via their status. Twitter API is processed to search for the tweets based on the geo-location. Someone posts on social network gives us a better concern to Analyze about the particular systems process of the system. Evaluating such data in social network is quite a challenging process. In the proposed system, there will be a workflow to mine the data which integrates both qualitative analysis and large-scale machine learning technique. Based on the different prominent theme's tweets will be categorized into different groups. Machine learning classifier will be implemented on mined data for qualitative analysis purpose to get the deeper understanding of the data. It uses multi label classification technique as each label falls into different categories and all the attributes are independent to each other. Label based measures will be taken to analyze the results and comparing them with the existing sentiment analysis technique.

LIST OF FIGURES

FIGURE NO	FIGURE NAME	PAGE NO
7.1	SYSTEM ARCHITECTURE	30
5.3.2	ER DIAGRAM	25
5.3.1	WORKFLOW DIAGRAM	25
5.3.3	USECASE DIAGRAM	26
5.3.4	SEQUENCE DIAGRAM	27
5.3.5	COLLABRATION DIAGRAM	27
11	OUTPUTS	37

LIST OF TABLES

TABLE NO	TABLE NAME	PAGE NO
4.5	.NET FRAMEWORK	18

LIST OF GRAPHS

GRAPH NO	GRAPH NAME	PAGE NO
1.2	SVM	05
1.2.2	LINEAR SVM	07
1.2.3	NON-LINEAR SVM	08

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	V
	LIST OF FIGURES	VI
	LIST OF TABLES	VI
	LIST OF GRAPHS	VI
1.	INTRODUCTION	01
	1.1 OUTLINE OF PROJECT	01
	1.2 PROPOSED ALGORITHM	01
	1.3 DOMAIN INTRODUCTION	02
	1.3.1 PYTHON IMPLEMENTATION	04
2.	LITERATURE REVIEW	10
3.	AIM AND SCOPE	14
	3.1 PURPOSE	
	3.2 SCOPE	14
	3.3 EXISTING SYSTEM	14
	3.4 PROPOSED SYSTEM	14
4.	SYSTEM REQUIRMENTS	15
	4.1 HARDWARE REQUIRIEMENTS	15

	4.2 SOFTWARE REQRIMENTS	15
	4.3 LANGUAGE SPECIFICATION	15
	4.4THE .NET FRAMEWORK	15
	4.5 LANGUAGES SUPPORTED BY.NET	17
	4.6 SQL SERVER	19
	4.7 FEASIBILITY STUDY	22
	4.7.1 ECONOMICAL FEASIBILITY	23
	4.7.2 TECHNICAL FEASIBILITY	23
	4.7.3 SOCIAL FEASIBILITY	23
5.	SYSTEM DESIGN	23
	5.1 INPUT DESIGN	23
	5.2 OUTPUT DESIGN	24
	5.3 DATA FLOW DIAGRAMS	24
6.	MODULES	28
	6.1 COLLECTION OF USER REVIEWS	28
	6.2 PRE-PROCESSING	29
	6.3 FEATURE SELECTION	29
	6.4 SENTIMENT CLASSIFICATION	29
	6.5 ANALYSIS OF REVIEWS	29
7.	ARCHITECTURE DIAGRAM	29
	7.1SYSTEM ARCHITECTURE	30

8.	SYSTEM TESTING	30
	8.1 TEST PLAN	30
	8.2 VERIFICATION	30
	8.3 VALIDATION	30
	8.4 BASICS OF SOFTWARE TESTING	30
	8.5 BLACK BOX TESTING	31
	8.6 WHITE BOX TESTING	31
	8.7 TYPES OF TESTING	31
9.	CONCLUSION	35
10.	REFERENCE	36
11.	APPENDIX	37
	A.OUTPUTS	37

CHAPTER 1

INTRODUCTION

Social media website is defined as “a website that facilitates meeting people, finding like minds, communicating and sharing content, and building community”; this kind of website allows or encourages various types of activities, such as commercial, social, or a combination of the two. Social media categories include digital library, e-commerce, entertainment, forum, geolocation, social bookmark, social review, social game, and social network. Social network is the subcategory of social media, which is the social structure of people who are joined by common interest. Social media are social channels of communication using web-based technologies, desktop computers, and mobile technologies. These technologies create highly interactive platforms through which individuals, communities, and organizations can share information, discuss, rate, comment, and modify user-generated and online contents. These advancements enable communication among businesses, organizations, communities, and individuals. Social media technologies change the way individuals and large organizations communicate, and they are increasingly being developed.

Wide range of applications in business and public policy uses sentiment analysis. Sentimental analysis is now being used from specific product marketing to antisocial behaviour recognition. Businesses and organizations have always been concerned about how they are perceived by the public. This concern results from a variety of motivations, including marketing and public relations. Before the era of Internet, the only way for an organization to track its reputation in the media was to hire someone for the specific task of reading newspapers and manually compiling lists of positive, negative and neutral references to the organization, it could undertake expensive surveys of uncertain validity. Today, many newspapers are published online. Some of them publish dedicated online editions, while others publish the pages of their print edition in PDF. In addition to newspapers, there are a wide range of opinionated articles posted online in blogs and other social media. This opens up the possibility of automatically detecting positive or negative mentions of an organization in articles published online, thereby dramatically reducing the effort required to collect this type of information. To this end, Organizations are becoming increasingly interested in acquiring fine sentiment analysis from news articles. Fine-grained sentiment analysis is an extremely challenging problem because of the variety of ways in which opinions can be expressed. News articles present an even greater challenge, as they usually avoid overt indicators of attitudes. However, despite their apparent neutrality, news articles can still bear polarity if they describe events that are objectively positive or negative. Many techniques used for sentiment analysis involve naïve approaches based on spotting certain keywords which reveal the author or speaker's emotions. We use naïve performs fine-grained sentiment analysis to classify sentences as positive, negative or neutral.

1.2 PROPOSED ALGORITHM

Naïve Bayes Classifier Algorithm

- Naïve Bayes algorithm is a supervised learning algorithm, which is based on **Bayes theorem** and used for solving classification problems.
- It is mainly used in *text classification* that includes a high-dimensional training dataset.
- Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
- **It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.**
- Some popular examples of Naïve Bayes Algorithm are **spam filtration, Sentimental analysis, and classifying articles.**

Why is it called Naïve Bayes?

The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, Which can be described as:

- **Naïve:** It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.
- **Bayes:** It is called Bayes because it depends on the principle of Bayes' Theorem

Bayes' Theorem:

- Bayes' theorem is also known as **Bayes' Rule** or **Bayes' law**, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.
- The formula for Bayes' theorem is given as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where,

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

P(A) is Prior Probability: Probability of hypothesis before observing the evidence.

P(B) is Marginal Probability: Probability of Evidence.

Working of Naïve Bayes' Classifier:

Working of Naïve Bayes' Classifier can be understood with the help of the below example:

Suppose we have a dataset of **weather conditions** and corresponding target variable "**Play**". So using this dataset we need to decide that whether we should play or not on a particular day according to the weather conditions. So to solve this problem, we need to follow the below steps:

1. Convert the given dataset into frequency tables.
2. Generate Likelihood table by finding the probabilities of given features.
3. Now, use Bayes theorem to calculate the posterior probability.

Advantages of Naïve Bayes Classifier:

- Naïve Bayes is one of the fast and easy ML algorithms to predict a class of datasets.
- It can be used for Binary as well as Multi-class Classifications.
- It performs well in Multi-class predictions as compared to the other Algorithms.
- It is the most popular choice for **text classification problems**.

Disadvantages of Naïve Bayes Classifier:

- Naive Bayes assumes that all features are independent or unrelated, so it cannot learn the relationship between features.

Applications of Naïve Bayes Classifier:

- It is used for **Credit Scoring**.
- It is used in **medical data classification**.
- It can be used in **real-time predictions** because Naïve Bayes Classifier is an eager learner.
- It is used in Text classification such as **Spam filtering** and **Sentiment analysis**.

Types of Naïve Bayes Model:

There are three types of Naive Bayes Model, which are given below:

- **Gaussian:** The Gaussian model assumes that features follow a normal distribution. This means if predictors take continuous values instead of discrete, then the model assumes that these values are sampled from the Gaussian distribution.
- **Multinomial:** The Multinomial Naïve Bayes classifier is used when the data is multinomial distributed. It is primarily used for document

classification problems, it means a particular document belongs to which category such as Sports, Politics, education, etc. The classifier uses the frequency of words for the predictors.

- **Bernoulli:** The Bernoulli classifier works similar to the Multinomial classifier, but the predictor variables are the independent Booleans variables. Such as if a particular word is present or not in a document. This model is also famous for document classification tasks.

Python Implementation of the Naïve Bayes algorithm:

Now we will implement a Naive Bayes Algorithm using Python. So for this, we will use the "**user_data**" **dataset**, which we have used in our other classification model. Therefore we can easily compare the Naive Bayes model with the other models.

Steps to implement:

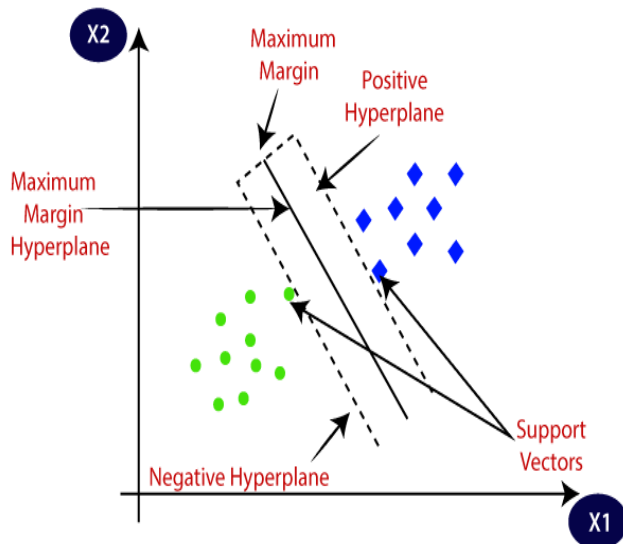
- Data Pre-processing step
- Fitting Naive Bayes to the Training set
- Predicting the test result
- Test accuracy of the result(Creation of Confusion matrix)
- Visualizing the test set result.

SVM

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:



Graph 1.2.1

Example: SVM can be understood with the example that we have used in the KNN classifier. Suppose we see a strange cat that also has some features of dogs, so if we want a model that can accurately identify whether it is a cat or dog, so such a model can be created by using the SVM algorithm. We will first train our model with lots of images of cats and dogs so that it can learn about different features of cats and dogs, and then we test it with this strange creature. So as support vector creates a decision boundary between these two data (cat and dog) and choose extreme cases (support vectors), it will see the extreme case of cat and dog. On the basis of the support vectors, it will classify it as a cat. Consider the below diagram:

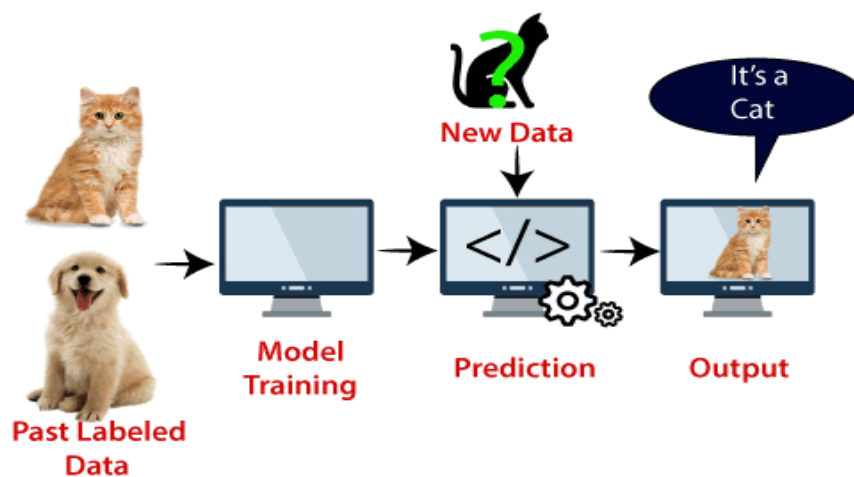


Fig 1.2.1

SVM algorithm can be used for **Face detection, image classification, text categorization**, etc.

Types of SVM

SVM can be of two types:

- **Linear SVM:** Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.
- **Non-linear SVM:** Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

Hyperplane and Support Vectors in the SVM algorithm:

Hyperplane: There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyperplane of SVM.

The dimensions of the hyperplane depend on the features present in the dataset, which means if there are 2 features (as shown in image), then hyperplane will be a straight line. And if there are 3 features, then hyperplane will be a 2-dimension plane.

We always create a hyperplane that has a maximum margin, which means the maximum distance between the data points.

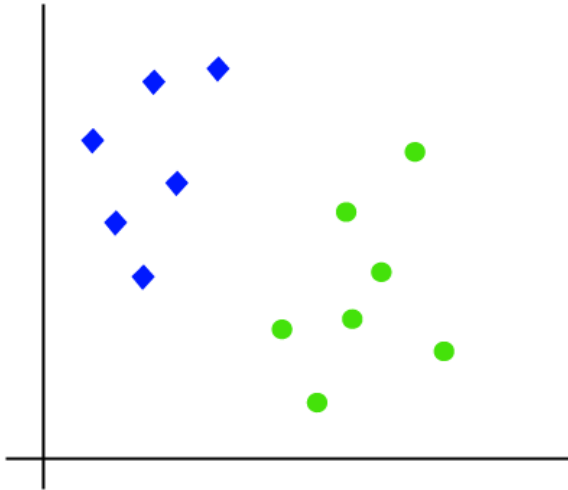
Support Vectors:

The data points or vectors that are the closest to the hyperplane and which affect the position of the hyperplane are termed as Support Vector. Since these vectors support the hyperplane, hence called a Support vector.

How does SVM works?

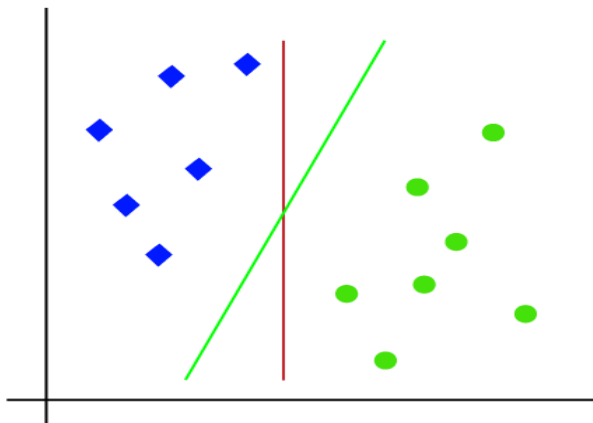
Linear SVM:

The working of the SVM algorithm can be understood by using an example. Suppose we have a dataset that has two tags (green and blue), and the dataset has two features x_1 and x_2 . We want a classifier that can classify the pair(x_1 , x_2) of coordinates in either green or blue. Consider the below image:



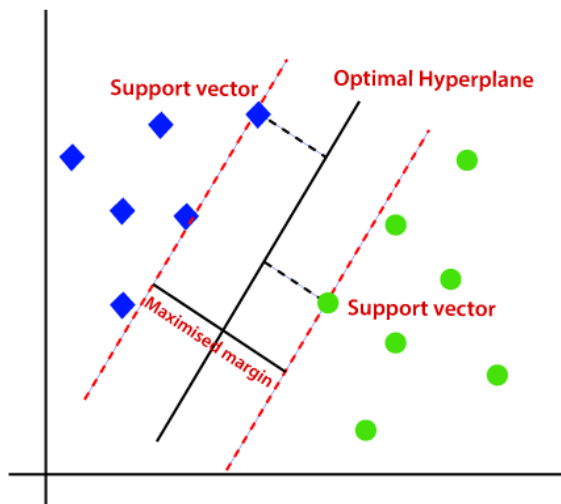
Graph 1.2.2

So as it is 2-d space so by just using a straight line, we can easily separate these two classes. But there can be multiple lines that can separate these classes. Consider the below image:



Graph 1.2.3

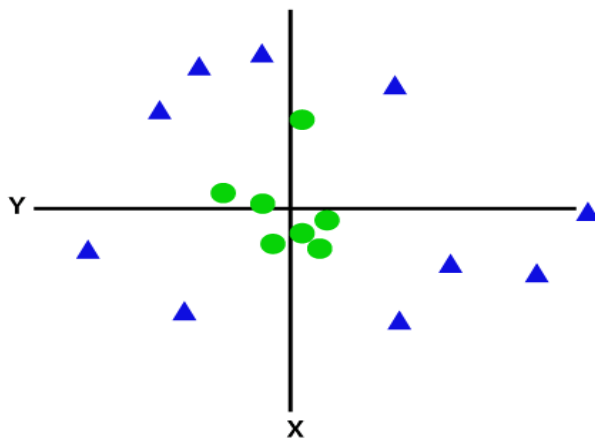
Hence, the SVM algorithm helps to find the best line or decision boundary; this best boundary or region is called as a **hyperplane**. SVM algorithm finds the closest point of the lines from both the classes. These points are called support vectors. The distance between the vectors and the hyperplane is called as **margin**. And the goal of SVM is to maximize this margin. The **hyperplane** with maximum margin is called the **optimal hyperplane**.



Graph 1.2.4

Non-Linear SVM:

If data is linearly arranged, then we can separate it by using a straight line, but for non-linear data, we cannot draw a single straight line. Consider the below image:

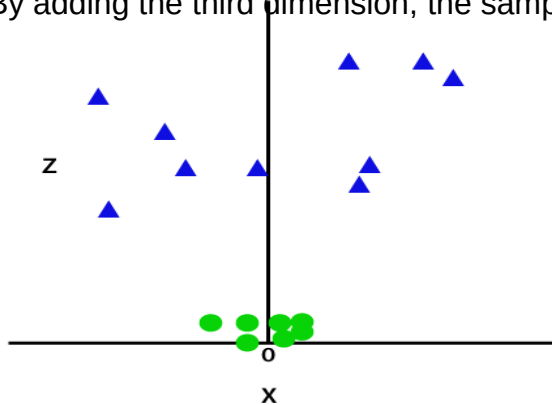


Graph 1.2.5

So to separate these data points, we need to add one more dimension. For linear data, we have used two dimensions x and y , so for non-linear data, we will add a third dimension z . It can be calculated as:

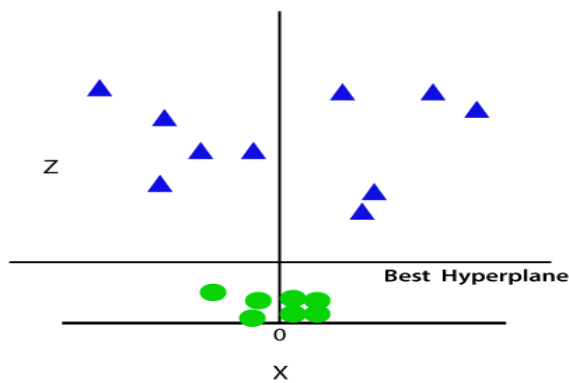
$$z = x^2 + y^2$$

By adding the third dimension, the sample space will become as below image:



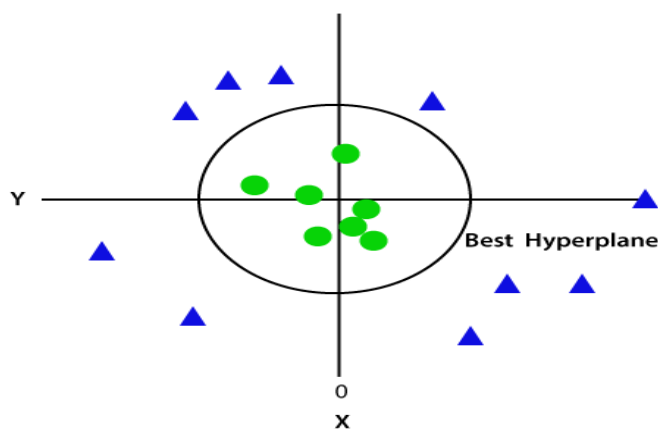
Graph 1.2.6

So now, SVM will divide the datasets into classes in the following way. Consider the below image:



Graph 1.2.7

Since we are in 3-d Space, hence it is looking like a plane parallel to the x-axis. If we convert it in 2d space with $z=1$, then it will become as:



Graph 1.2.8

Hence we get a circumference of radius 1 in case of non-linear data.

CHAPTER 2

LITERATURE REVIEW

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations.

A google wave-based fuzzy recommender system to disseminate information in University Digital Libraries 2.0

Nowadays [Digital Libraries](#) 2.0 are mainly based on the interaction between users through collaborative applications such as wikis, blogs, etc. or new possible paradigms like the waves proposed by Google. This new concept, the wave, represents a [common space](#) where resources and users can work together. The problem arises when the number of resources and users is high, then tools for assisting the users in their information needs are necessary. In this case a fuzzy linguistic [recommender system](#) based on the Google Wave capabilities is proposed as tool for communicating researchers interested in common research lines. The system allows the creation of a common space by means a wave as a way of collaborating and exchanging ideas between several researchers interested in the same topic. In addition, the system suggests, in an automatic way, several researchers and useful resources for each wave. These recommendations are computed following several previously defined preferences and characteristics by means of fuzzy linguistic labels. Thus the system facilitates the possible collaborations between multi-disciplinar researchers and recommends complementary resources useful for the interaction. In order to test the effectiveness of the proposed system, a prototype of the system has been developed and tested with several research groups from the same university achieving successful results.

A hybrid fuzzy-based personalized recommender system for telecom products/services

The Internet creates [excellent opportunities](#) for businesses to provide personalized [online services](#) to their customers. [Recommender systems](#) are designed to automatically generate personalized suggestions of [products/services](#) to customers. Because various uncertainties exist within both product and customer data, it is a challenge to achieve high recommendation accuracy. This study develops a hybrid recommendation approach which combines user-based and item-based collaborative filtering techniques with [fuzzy set](#) techniques and applies it to mobile product and service recommendation. It particularly implements the proposed approach in an intelligent recommender system software called Fuzzy-based Telecom Product Recommender System (FTCP-RS). [Experimental results](#) demonstrate the effectiveness of the proposed approach and the initial application shows that the FTCP-RS can effectively help customers to select the most suitable mobile products or services.

Recommender systems based on social networks

The traditional recommender systems, especially the collaborative filtering recommender systems, have been studied by many researchers in the past decade. However, they ignore the social relationships among users. In fact, these relationships can improve the accuracy of recommendation. In recent years, the study of social-based recommender systems has become an active research topic. In this paper, we propose a social regularization approach that incorporates social network information to benefit recommender systems. Both users' friendships and rating records (tags) are employed to predict the missing values (tags) in the user-item matrix. Especially, we use a biclustering algorithm to identify the most suitable group of friends for generating different final recommendations. Empirical analyses on real datasets show that the proposed approach achieves superior performance to existing approaches.

A Hybrid Trust-Based Recommender System for Online Communities of Practice

The needs for life-long learning and the rapid development of information technologies promote the development of various types of online Community of Practices. In online CoPs, bounded rationality and metacognition are two major issues, especially when learners face information overload and there is no knowledge authority within the learning environment. This study proposes a hybrid, trust-based recommender system to mitigate above learning issues in online CoPs. A case study was conducted using Stack Overflow data to test the recommender system. Important findings include: (1) comparing with other social community platforms, learners in online CoPs have stronger social relations and tend to interact with a smaller group of people only; (2) the hybrid algorithm can provide more accurate recommendations than celebrity-based and content-based algorithm and; (3) the proposed recommender system can facilitate the formation of personalized learning communities.

RecomMetz: A context-aware knowledge-based mobile recommender system for movie showtimes

Recommender systems are used to provide filtered information from a large amount of elements. They provide personalized recommendations on products or services to users. The recommendations are intended to provide

interesting elements to users. Recommender systems can be developed using different techniques and algorithms where the selection of these techniques depends on the area in which they will be applied. This paper proposes a recommender system in the leisure domain, specifically in the movie showtimes domain. The system proposed is called RecomMetz, and it is a context-aware mobile recommender system based on Semantic Web technologies. In detail, a domain ontology primarily serving a semantic similarity metric adjusted to the concept of “packages of single items” was developed in this research. In addition, location, crowd and time were considered as three different kinds of contextual information in RecomMetz. In a nutshell, RecomMetz has unique features: (1) the items to be recommended have a composite structure (movie theater + movie + showtime), (2) the integration of the time and crowd factors into a context-aware model, (3) the implementation of an ontology-based context modeling approach and (4) the development of a multi-platform native mobile user interface intended to leverage the hardware capabilities (sensors) of mobile devices. The evaluation results show the efficiency and effectiveness of the recommendation mechanism implemented by RecomMetz in both a cold-start scenario and a no cold-start scenario.

A novel hybrid approach improving effectiveness of recommender systems

Recommender systems support users by generating potentially interesting suggestions about relevant products and information. The increasing attention towards such tools is witnessed by both the great number of powerful and sophisticated recommender algorithms developed in recent years and their adoption in many popular Web platforms. However, performances of recommender systems can be affected by many critical issues as for instance, over-specialization, attribute selection and scalability. To mitigate some of such negative effects, a hybrid recommender system, called Relevance Based Recommender, is proposed in this paper. It exploits individual measures of perceived relevance computed by each user for each instance of interest and, to obtain a better precision, also by considering the analogous measures computed by the other users for the same instances. Some experiments show the advantages introduced by this recommender when generating potentially attractive suggestions.

A recommender system for researchers based on bibliometrics

Recommender systems (RSs) exploit past behaviors and user similarities to provide personalized recommendations. There are some precedents of usage in academic environments to assist users finding relevant information, based on assumptions about the characteristics of the items and users. Even if quality has already been taken into account as a property of items in previous works, it has never been given a key role in the re-ranking process for both items and users.

In this paper, we present REFORE, a quality-based fuzzy linguistic REcommender system FOr REsearchers. We propose the use of some bibliometric measures as the way to quantify the quality of both items and users without the interaction of experts as well as the use of 2-tuple linguistic

approach to describe the linguistic information. The system takes into account the measured quality as the main factor for the re-ranking of the top-N recommendations list in order to point out researchers to the latest and the best papers in their research fields. To prove the accuracy improvement, we conduct a study involving different recommendation approaches, aiming at measuring their performance gain. The results obtained proved to be satisfactory for the researchers from different departments who took part on the tests.

A peer-to-peer recommender system for self-emerging user communities based on gossip overlays

Gossip-based peer-to-peer protocols proved to be very efficient for supporting dynamic and complex information exchange among distributed peers. They are useful for building and maintaining the network topology itself as well as to support a pervasive diffusion of the information injected into the network. This is very useful in a world where there is a growing need to access and be aware of many types of distributed resources like Internet pages, shared files, online products, news and information. Finding flexible, scalable and efficient mechanisms addressing this topic is a key issue, even with relevant social and economic aspects. In this paper, we propose the general architecture of a system whose aim is to exploit the collaborative exchange of information between peers in order to build a system able to gather similar users and spread useful suggestions among them.

Social and Content Hybrid Image Recommender System for Mobile Social Networks

One of the advantages of social networks is the possibility to socialize and personalize the content created or shared by the users. In mobile social networks, where the devices have limited capabilities in terms of screen size and computing power, Multimedia Recommender Systems help to present the most relevant content to the users, depending on their tastes, relationships and profile. Previous recommender systems are not able to cope with the uncertainty of automated tagging and are knowledge domain dependant. In addition, the instantiation of a recommender in this domain should cope with problems arising from the collaborative filtering inherent nature (cold start, banana problem, large number of users to run, etc.). The solution presented in this paper addresses the abovementioned problems by proposing a hybrid image recommender system, which combines collaborative filtering (social techniques) with content-based techniques, leaving the user the liberty to give these processes a personal weight. It takes into account aesthetics and the formal characteristics of the images to overcome the problems of current techniques, improving the performance of existing systems to create a mobile social networks recommender with a high degree of adaptation to any kind of user.

A Personalized Recommender System Based on a Hybrid Model

Recommender systems are means for web personalization and tailoring the browsing experience to the users' specific needs. There are two categories of recommender systems; memory-based and model-based systems. In this

paper we propose a personalized recommender system for the next page prediction that is based on a hybrid model from both categories. The generalized patterns generated by a model based techniques are tailored to specific users by integrating user profiles generated from the traditional memory-based system's user-item matrix. The suggested system offered a significant improvement in prediction speed over traditional model-based usage mining systems, while also offering an average improvement in the system accuracy and system precision by 0.27% and 2.35%, respectively.

CHAPTER 3

3.1 PURPOSE

The purpose of this document is opinion mining for online customer reviews using machine learning algorithms. In detail, this document will provide a general description of our project, including user requirements, product perspective, and overview of requirements, general constraints. In addition, it will also provide the specific requirements and functionality needed for this project - such as interface, functional requirements and performance requirements.

3.2 SCOPE

The scope of this SRSdocument persists for the entire life cycle of the project. This document defines the final state of the software requirements agreed upon by the customers and designers. Finally at the end of the project execution all the functionalities may be traceable from the SRS to the product. The document describes the functionality, performance, constraints, interface and reliability for the entire life cycle of the project.

3.3 EXISTING SYSTEM

- In existing system the data include surveys, interviews, questionnaires, classroom activities about the student educational experiences and problems they are facing.
- But these traditional methods are time consuming and very limited in scale. The manual analysis does not make sense of analyzing student learning experiences which are huge in volume with different Internet slang and the timing of the student posting on the web.
- The sentiment analysis of the tweets does not cover much relevant experience because even for a human judge to determine what student problems a tweet indicates is a more complicated task than to determine just the sentiment of a tweet.
- These traditional methods are very time consuming and very limited in scale.

3.4 PROPOSED SYSTEM

The proposed system has to perform the qualitative analysis using classification algorithm instead of sentiment analysis. Sentiment analysis

considers the opinion of the user about a system or product and categorizes it to neutral, negative or positive mood. In the proposed system, there will be a workflow to mine the data which integrates both qualitative analysis and large scale machine learning technique. Based on the different prominent themes tweets will be categorized into different groups. Machine learning classifier will be implemented on mined data for qualitative analysis purpose to get the deeper understanding of the data. It uses multi label classification technique as each label falls into different categories and all the attributes are independent to each other. Label based measures will be taken to analyze the results and comparing them with the existing sentiment analysis technique.

CHAPTER 4

SYSTEM REQUIREMENTS

4.1 HARDWARE REQUIREMENTS

- System : i3 Processor
- Hard Disk : 500 GB.
- Monitor : 15"LED
- Input Devices : Keyboard, Mouse
- Ram : 4GB.

4.2 SOFTWARE REQUIREMENTS

- Operating system : Windows
- Coding Language : .Net
- IDE : Visual Studio

4.3 LANGUAGE SPECIFICATION

INTRODUCTION TO .NET

FEATURES OF .NET

Microsoft .NET is a set of Microsoft software technologies for rapidly building and integrating XML Web services, Microsoft Windows-based applications, and Web solutions. The .NET Framework is a language-neutral platform for writing programs that can easily and securely interoperate. There's no language barrier with .NET: there are numerous languages available to the developer including Managed C++, C#, Visual Basic and Java Script. The .NET framework provides the foundation for components to interact seamlessly, whether locally or remotely on different platforms. It standardizes common data types and communications protocols so that components created in different languages can easily interoperate.

“.NET” is also the collective name given to various software components built upon the .NET platform. These will be both products (Visual Studio.NET and Windows.NET Server, for instance) and services (like Passport, .NET My Services, and so on).

4.4 THE .NET FRAMEWORK

The .NET Framework has two main parts:

1. The Common Language Runtime (CLR).
2. A hierarchical set of class libraries.

The CLR is described as the “execution engine” of .NET. It provides the environment within which programs run. The most important features are

- ◆ Conversion from a low-level assembler-style language, called Intermediate Language (IL), into code native to the platform being executed on.
- ◆ Memory management, notably including garbage collection.
- ◆ Checking and enforcing security restrictions on the running code.
- ◆ Loading and executing programs, with version control and other such features.
- ◆ The following features of the .NET framework are also worth description:

Managed Code

The code that targets .NET, and which contains certain extra Information - “metadata” - to describe itself. Whilst both managed and unmanaged code can run in the runtime, only managed code contains the information that allows the CLR to guarantee, for instance, safe execution and interoperability.

Managed Data

With Managed Code comes Managed Data. CLR provides memory allocation and Deal location facilities, and garbage collection. Some .NET languages use Managed Data by default, such as C#, Visual Basic.NET and JScript.NET, whereas others, namely C++, do not. Targeting CLR can, depending on the language you’re using, impose certain constraints on the features available. As with managed and unmanaged code, one can have both managed and unmanaged data in .NET applications - data that doesn’t get garbage collected but instead is looked after by unmanaged code.

Common Type System

The CLR uses something called the Common Type System (CTS) to strictly enforce type-safety. This ensures that all classes are compatible with each other, by describing types in a common way. CTS define how types work within the runtime, which enables types in one language to interoperate with types in another language, including cross-language exception handling. As well as ensuring that types are only used in appropriate ways, the runtime also ensures that code doesn’t attempt to access memory that hasn’t been allocated to it.

Common Language Specification

The CLR provides built-in support for language interoperability. To ensure that you can develop managed code that can be fully used by developers using any programming language, a set of language features and rules for using them called the Common Language Specification (CLS) has been defined. Components that follow these rules and expose only CLS features are considered CLS-compliant.

THE CLASS LIBRARY

.NET provides a single-rooted hierarchy of classes, containing over 7000 types. The root of the namespace is called System; this contains basic types like Byte, Double, Boolean, and String, as well as Object. All objects derive from System. Object. As well as objects, there are value types. Value types can be allocated on the stack, which can provide useful flexibility. There are also efficient means of converting value types to object types if and when necessary.

The set of classes is pretty comprehensive, providing collections, file, screen, and network I/O, threading, and so on, as well as XML and database connectivity.

The class library is subdivided into a number of sets (or namespaces), each providing distinct areas of functionality, with dependencies between the namespaces kept to a minimum.

4.5 LANGUAGES SUPPORTED BY .NET

The multi-language capability of the .NET Framework and Visual Studio .NET enables developers to use their existing programming skills to build all types of applications and XML Web services. The .NET framework supports new versions of Microsoft's old favorites Visual Basic and C++ (as VB.NET and Managed C++), but there are also a number of new additions to the family.

Visual Basic .NET has been updated to include many new and improved language features that make it a powerful object-oriented programming language. These features include inheritance, interfaces, and overloading, among others. Visual Basic also now supports structured exception handling, custom attributes and also supports multi-threading.

Visual Basic .NET is also CLS compliant, which means that any CLS-compliant language can use the classes, objects, and components you create in Visual Basic .NET.

Managed Extensions for C++ and attributed programming are just some of the enhancements made to the C++ language. Managed Extensions simplify the task of migrating existing C++ applications to the new .NET Framework. C# is Microsoft's new language. It's a C-style language that is essentially "C++ for Rapid Application Development". Unlike other languages, its specification is just the grammar of the language. It has no standard library of its own, and instead has been designed with the intention of using the .NET libraries as its own.

Microsoft Visual J# .NET provides the easiest transition for Java-language developers into the world of XML Web Services and dramatically improves the interoperability of Java-language programs with existing software written in a variety of other programming languages.

Active State has created Visual Perl and Visual Python, which enable .NET-aware applications to be built in either Perl or Python. Both products can be integrated into the Visual Studio .NET environment. Visual Perl includes support for Active State's Perl Dev Kit.

Other languages for which .NET compilers are available include

- FORTRAN
- COBOL
- Eiffel

ASP.NET		Windows
XML SERVICES	WEB	Forms
Base Class Libraries		
Common Language Runtime		
Operating System		

Table 4.5.1. NET Framework

C#.NET is also compliant with CLS (Common Language Specification) and supports structured exception handling. CLS is set of rules and constructs that are supported by the CLR (Common Language Runtime). CLR is the runtime environment provided by the .NET Framework; it manages the execution of the code and also makes the development process easier by providing services C#.NET is a CLS-compliant language. Any objects, classes, or components that created in C#.NET can be used in any other CLS-compliant language. In addition, we can use objects, classes, and components created in other CLS-compliant languages in C#.NET .The use of CLS ensures complete interoperability among applications, regardless of the languages used to create the application.

Constructors and Destructors

Constructors are used to initialize objects, whereas destructors are used to destroy them. In other words, destructors are used to release the resources allocated to the object. In C#.NET the sub finalize procedure is available. The sub finalize procedure is used to complete the tasks that must be performed when an object is destroyed. The sub finalize procedure is called automatically when an object is destroyed. In addition, the sub finalize procedure can be called only from the class it belongs to or from derived classes.

Garbage Collection

Garbage Collection is another new feature in C#.NET. The .NET Framework monitors allocated resources, such as objects and variables. In addition, the .NET Framework automatically releases memory for reuse by destroying objects that are no longer in use. In C#.NET, the garbage collector checks for the objects that are not currently in use by applications. When the garbage collector comes across an object that is marked for garbage collection, it releases the memory occupied by the object.

Overloading

Overloading is another feature in C#. Overloading enables us to define multiple procedures with the same name, where each procedure has a different set of arguments. Besides using overloading for procedures, we can use it for constructors and properties in a class.

Multithreading

C#.NET also supports multithreading. An application that supports multithreading can handle multiple tasks simultaneously, we can use multithreading to decrease the time taken by an application to respond to user interaction.

Structured Exception Handling

C#.NET supports structured handling, which enables us to detect and remove errors at runtime. In C#.NET, we need to use Try...Catch...Finally statements to create exception handlers. Using Try...Catch...Finally statements, we can create robust and effective exception handlers to improve the performance of our application.

The .Net Framework

The .NET Framework is a new computing platform that simplifies application development in the highly distributed environment of the Internet.

OBJECTIVES OF .NET FRAMEWORK

1. To provide a consistent object-oriented programming environment whether object codes is stored and executed locally on Internet-distributed, or executed remotely.
2. To provide a code-execution environment to minimizes software deployment and guarantees safe execution of code.
3. Eliminates the performance problems.

There are different types of application, such as Windows-based applications and Web-based applications.

4.6 SQL SERVER

Microsoft is the vendor of SQL Server. We have different editions of SQL Server, where SQL Server Express is free to download and use. SQL Server uses T-SQL (Transact-SQL). T-SQL is Microsoft's proprietary extension to SQL. TSQL is very similar to standard SQL, but in addition it supports some extra functionality, built in functions, etc.

T-SQL expands on the SQL standard to include procedural programming, local variables, and various support functions for string processing, date

processing, mathematics, etc. SQL Server consists of a Database Engine and a Management Studio (and lots of other stuff which we will not mention here). The Database engine has no graphical interface - it is just a service running in the background of your computer (preferable on the server). The Management Studio is graphical tool for configuring and viewing the information in the database. It can be installed on the server or on the client (or both).

SQL Server Management Studio can also be used to create a new database, alter any existing database schema by adding or modifying tables and indexes, or analyze performance. It includes the query windows which provide a GUI based interface to write and execute queries.

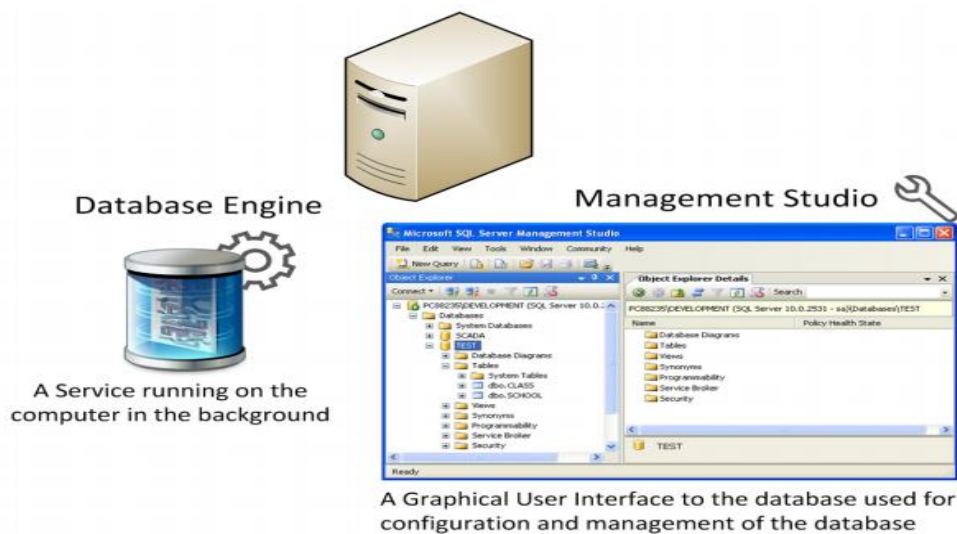


Fig 4.6.1

SQL SERVER MANAGEMENT STUDIO

SQL Server Management Studio is a GUI tool included with SQL Server for configuring, managing, and administering all components within Microsoft SQL Server. The tool includes both script editors and graphical tools that work with objects and features of the server. As mentioned earlier, version of SQL Server Management Studio is also available for SQL Server Express Edition, for which it is known as SQL Server Management Studio Express.

A central feature of SQL Server Management Studio is the Object Explorer, which allows the user to browse, select, and act upon any of the objects within the server. It can be used to visually observe and analyze query plans and optimize the database performance, among others. SQL Server Management Studio can also be used to create a new database, alter any existing database schema by adding or modifying tables and indexes, or analyze performance. It includes the query windows which provide a GUI based interface to write and execute queries.

When creating SQL commands and queries, the “Query Editor” (select “New Query” from the Toolbar) is used (shown in the figure above). With SQL and the “Query Editor” we can do almost everything with code, but sometimes it is also a good idea to use the different Designer tools in SQL to help us do the work without coding (so much).

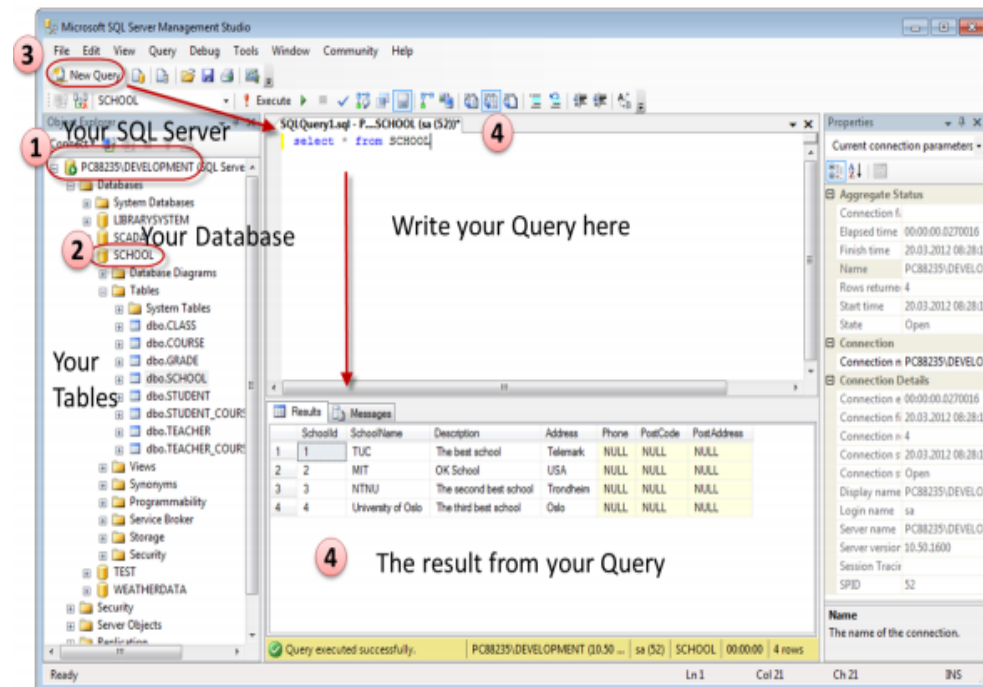


Fig 4.6.2

Create a new Database

It is quite simple to create a new database in Microsoft SQL Server. Just right-click on the “Databases” node and select “New Database...”

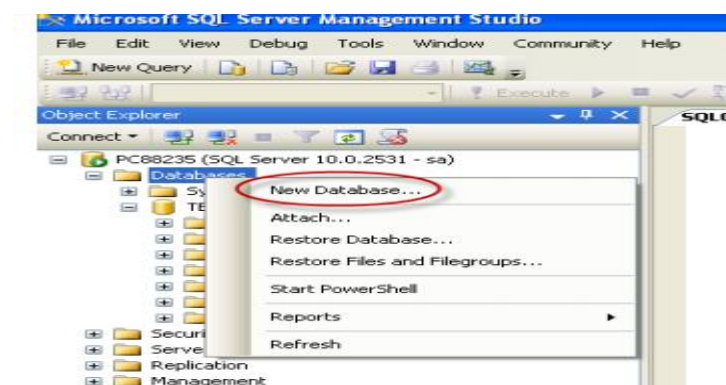


Fig 4.6.3

The OLAP Services feature available in SQL Server version 7.0 is now called SQL Server 2000 Analysis Services. The term OLAP Services has been replaced with the term Analysis Services. Analysis Services also includes a new data mining component. The Repository component available in SQL Server version 7.0 is now called Microsoft SQL Server 2000 Meta

Data Services. References to the component now use the term Meta Data Services. The term repository is used only in reference to the repository engine within Meta Data Services

SQL-SERVER database consist of six type of objects,

They are,

1. TABLE
2. QUERY
3. FORM
4. REPORT
5. MACRO

Table

A database is a collection of data about a specific topic.

Views of Table

We can work with a table in two types,

1. Design View
2. Datasheet View

Design View

To build or modify the structure of a table we work in the table design view. We can specify what kind of data will be hold.

Datasheet View

To add, edit or analyses the data itself we work in tables datasheet view mode.

Query

A query is a question that has to be asked the data. Access gathers data that answers the question from one or more table. The data that make up the answer is either dynaset (if you edit it) or a snapshot (it cannot be edited).Each time we run query, we get latest information in the dynaset. Access either displays the dynaset or snapshot for us to view or perform an action on it, such as deleting or updating.

4.7 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

The feasibility study investigates the problem and the information needs of the stakeholders. It seeks to determine the resources required to provide an information systems solution, the cost and benefits of such a solution, and the feasibility of such a solution.

The goal of the feasibility study is to consider alternative information systems solutions, evaluate their feasibility, and propose the alternative most suitable to the organization. The feasibility of a proposed solution is evaluated in terms of its components.

4.7.1 ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

4.7.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

4.7.3 SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity.

CHAPTER 5 SYSTEM DESIGN

5.1 INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.

- Methods for preparing input validations and steps to follow when error occur.

5.2 OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action

5.3 DATA FLOW DIAGRAMS

1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
2. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
4. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

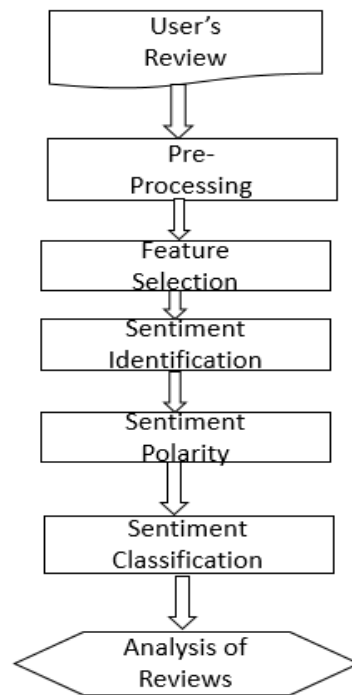


Fig 5.3.1

ER DIAGRAMS

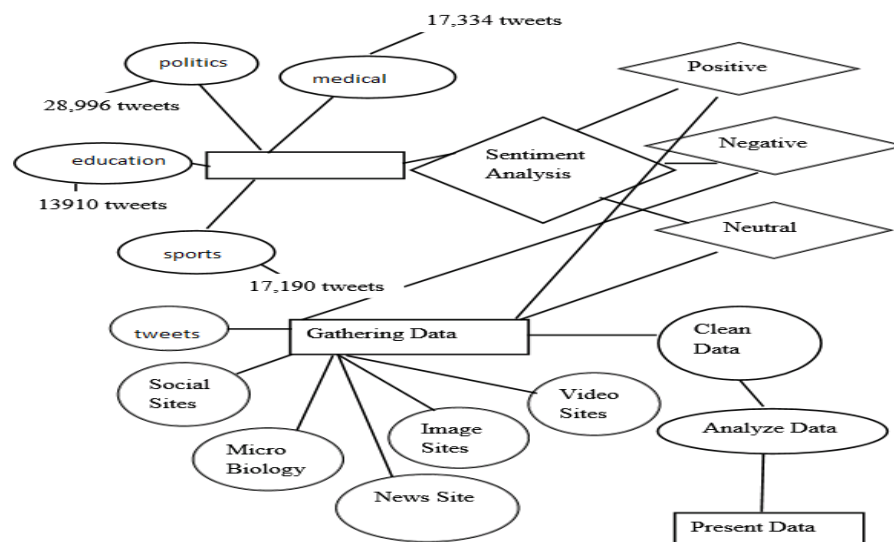


Fig 5.3.2

UML DIAGRAM

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software

engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

GOALS:

The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components.
7. Integrate best practices.

USE CASE DIAGRAM

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

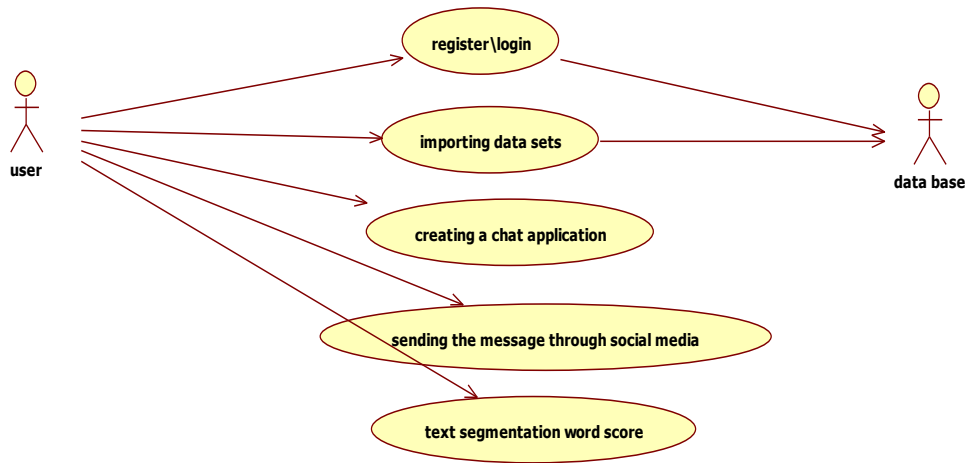


Fig 5.3.3

SEQUENCE DIAGRAM

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

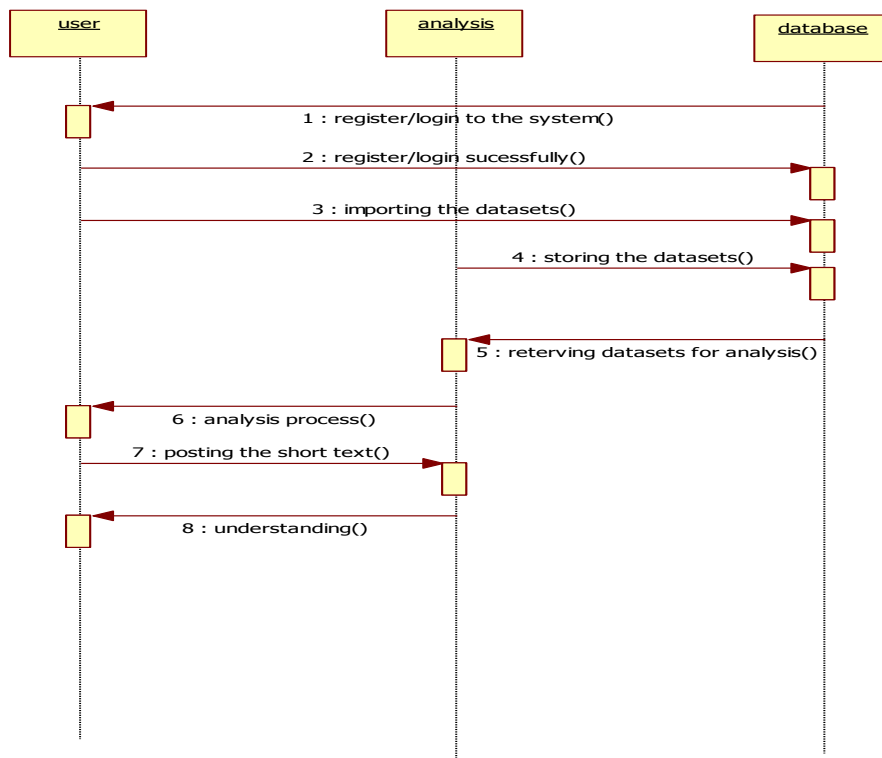


Fig 5.3.4

COLLABRATION DIAGRAM

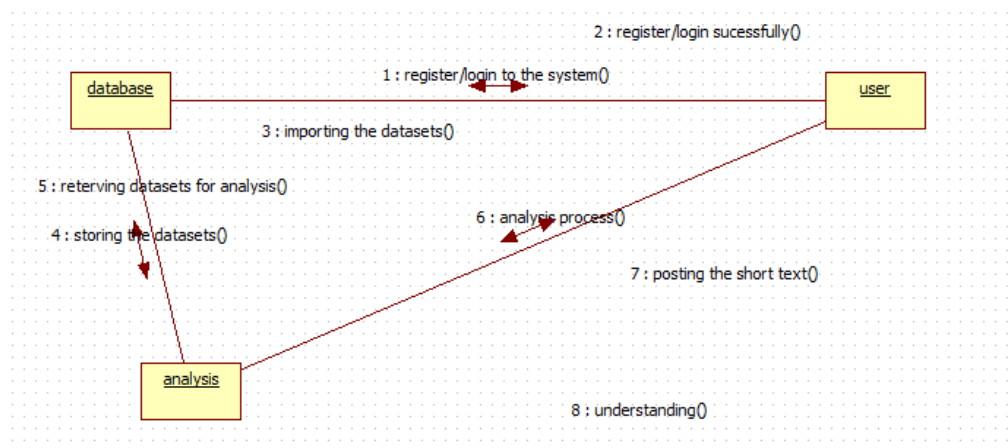


Fig 5.3.

CHAPTER 6

MODULES

- Collection of User's Reviews
- Pre-Processing
- Feature Selection
- Sentiment Classification
- Analysis of Reviews

Modules Description

6.1 Collection of User's Reviews

Reviews are necessary for doing the Sentiment Analysis Task. For the Collection of reviews there are different techniques which are used in this survey. The reviews can be a structured, semi-structured and unstructured type. Sentiment Analysis research, there are open source framework where researcher can get their data for the research purpose. R is one of the programming language and software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. By installing required packages and authentication process of social website, to crawl the reviews from that site is easy task. Once we have our text data with us then we can use that data for Pre-processing purpose.

6.2 Pre-Processing

In pre-processing following are some tasks:

- Removing URLs, Special characters, Numbers, Punctuations etc.
- Removing Stop words
- Removal of Retweets (in case of twitter dataset)

- Stemming
- Tokenization

6.3 Feature Selection

Feature selection from pre-processed text is the difficult task in sentiment analysis. The main goal of the feature selection is to decrease the dimensionality of the feature space and thus computational cost. Feature selection will reduce the over fitting of the learning scheme to the training data. In different machine learning algorithms were analysed on a news review dataset with different feature selection techniques features are usually unigrams, bigrams and grams. POS tagging is used in feature selection techniques.

6.4 Sentiment Classification

Sentiment classification of news review dataset and product review dataset is done using supervised machine learning approaches like naïve Bayes, SVM, Maximum Entropy etc. Accuracy is depends on which dataset is used for which classification methods. In the case of Supervised machine learning approaches Training dataset is used to train the classification model which then help to classify the test data.

6.5 Analysis of Reviews

Finally Analysis of result is important to make decision to individual and industry. In case of news reviews if more result is positive then user can decide to go that news event. Analysis is used in business intelligence.

CHAPTER 7

ARCHITECTURE DIAGRAM

7.1 SYSTEM ARCHITECTURE

Describing the overall features of the software is concerned with defining the requirements and establishing the high level of the system. During architectural design, the various web pages and their interconnections are identified and designed. The major software components are identified and decomposed into processing modules and conceptual data structures and the interconnections among the modules are identified. The following modules are identified in the proposed system.

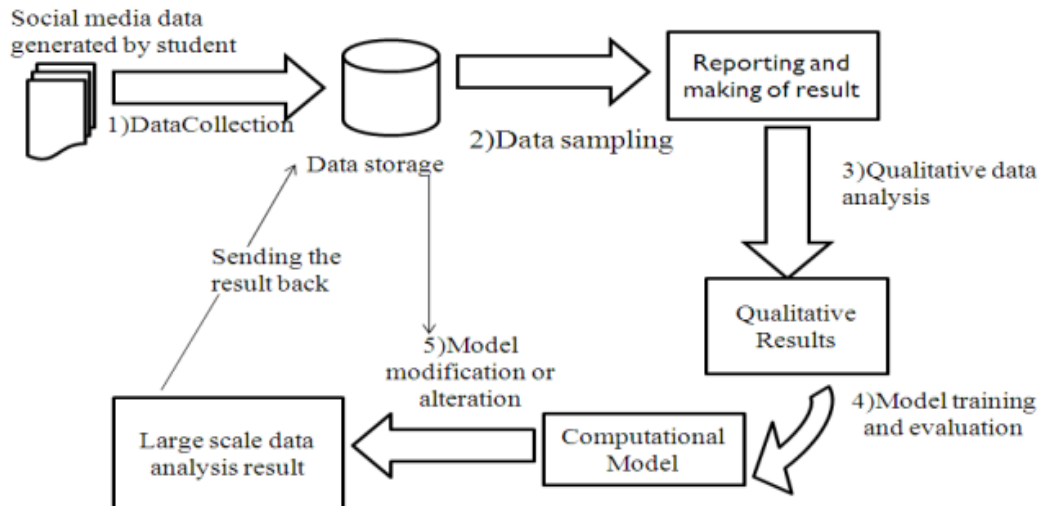


Fig 7.1

CHAPTER 8

SYSTEM TESTING

8.1 Test plan

Software testing is the process of evaluation a software item to detect differences between given input and expected output. Also to assess the feature of a software item. Testing assesses the quality of the product. Software testing is a process that should be done during the development process. In other words software testing is a verification and validation process.

8.2 Verification

Verification is the process to make sure the product satisfies the conditions imposed at the start of the development phase. In other words, to make sure the product behaves the way we want it to.

8.3 Validation

Validation is the process to make sure the product satisfies the specified requirements at the end of the development phase. In other words, to make sure the product is built as per customer requirements.

8.4 Basics of software testing

There are two basics of software testing: black box testing and white box testing.

8.5 Black box Testing

Black box testing is a testing technique that ignores the internal mechanism of the system and focuses on the output generated against any input and execution of the system. It is also called functional testing.

8.6 White box Testing

White box testing is a testing technique that takes into account the internal mechanism of a system. It is also called structural testing and glass box testing. Black box testing is often used for validation and white box testing is often used for verification.

8.7 Types of testing

There are many types of testing like

- Unit Testing
- Integration Testing
- Functional Testing
- System Testing
- Stress Testing
- Performance Testing
- Usability Testing
- Acceptance Testing
- Regression Testing
- Beta Testing

8.7.1 Unit Testing

Unit testing is the testing of an individual unit or group of related units. It falls under the class of white box testing. It is often done by the programmer to test that the unit he/she has implemented is producing expected output against given input.

8.7.2 Integration Testing

Integration testing is testing in which a group of components are combined to produce output. Also, the interaction between software and hardware is tested in integration testing if software and hardware components have any relation. It may fall under both white box testing and black box testing.

8.7.3 Functional Testing

Functional testing is the testing to ensure that the specified functionality required in the system requirements works. It falls under the class of black box testing.

8.7.4 System Testing

System testing is the testing to ensure that by putting the software in different environments (e.g., Operating Systems) it still works. System testing is done with full system implementation and environment. It falls under the class of black box testing.

8.7.5 Stress Testing

Stress testing is the testing to evaluate how system behaves under unfavorable conditions. Testing is conducted at beyond limits of the specifications. It falls under the class of black box testing.

8.7.6 Performance Testing

Performance testing is the testing to assess the speed and effectiveness of the system and to make sure it is generating results within a specified time as in performance requirements. It falls under the class of black box testing.

8.7.7 Usability Testing

Usability testing is performed to the perspective of the client, to evaluate how the GUI is user-friendly? How easily can the client learn? After learning how to use, how proficiently can the client perform? How pleasing is it to use its design? This falls under the class of black box testing.

8.7.8 Acceptance Testing

Acceptance testing is often done by the customer to ensure that the delivered product meets the requirements and works as the customer expected. It falls under the class of black box testing.

8.7.9 Regression Testing

Regression testing is the testing after modification of a system, component, or a group of related units to ensure that the modification is working correctly and is not damaging or imposing other modules to produce unexpected results. It falls under the class of black box testing

REQUIREMENT ANALYSIS

Requirement analysis, also called requirement engineering, is the process of determining user expectations for a new modified product. It encompasses the tasks that determine the need for analyzing, documenting, validating and managing software or system requirements. The requirements should be documentable, actionable, measurable, testable and traceable related to identified business needs or opportunities and define to a level of detail, sufficient for system design.

FUNCTIONAL REQUIREMENTS

It is a technical specification requirement for the software products. It is the first step in the requirement analysis process which lists the requirements of particular software systems including functional, performance and security requirements. The function of the system depends mainly on the quality hardware used to run the software with given functionality.

Usability

It specifies how easy the system must be use. It is easy to ask queries in any format which is short or long, porter stemming algorithm stimulates the desired response for user.

Robustness

It refers to a program that performs well not only under ordinary conditions but also under unusual conditions. It is the ability of the user to cope with errors for irrelevant queries during execution.

Security

The state of providing protected access to resource is security. The system provides good security and unauthorized users cannot access the system there by providing high security.

Reliability

It is the probability of how often the software fails. The measurement is often expressed in MTBF (Mean Time Between Failures). The requirement is needed in order to ensure that the processes work correctly and completely without being aborted. It can handle any load and survive and survive and even capable of working around any failure.

Compatibility

It is supported by version above all web browsers. Using any web servers like localhost makes the system real-time experience.

Flexibility

The flexibility of the project is provided in such a way that is has the ability to run on different environments being executed by different users.

Safety

Safety is a measure taken to prevent trouble. Every query is processed in a secured manner without letting others to know one's personal information.

NON- FUNCTIONAL REQUIREMENTS

Portability

It is the usability of the same software in different environments. The project can be run in any operating system.

Performance

These requirements determine the resources required, time interval, throughput and everything that deals with the performance of the system.

Accuracy

The result of the requesting query is very accurate and high speed of retrieving information. The degree of security provided by the system is high and effective.

Maintainability

Project is simple as further updates can be easily done without affecting its stability. Maintainability basically defines that how easy it is to maintain the system. It means that how easy it is to maintain the system, analyse, change and test the application. Maintainability of this project is simple as further updates can be easily done without affecting its stability.

FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

The feasibility study investigates the problem and the information needs of the stakeholders. It seeks to determine the resources required to provide an information systems solution, the cost and benefits of such a solution, and the feasibility of such a solution.

The goal of the feasibility study is to consider alternative information systems solutions, evaluate their feasibility, and propose the alternative most suitable to the organization. The feasibility of a proposed solution is evaluated in terms of its components.

ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity.

CHAPTER 9

CONCLUSION

In this proposed work, we used Twitter API using R tool which is open source. Tweets from twitter has been collected and gives to pre-processing task in that tool. R open source tool is used in text mining and also to crawl streaming data from social media like twitter and Facebook etc. Movie reviews data also pre-processed in R tool for sentiment analysis and opinion mining. There are different supervised and unsupervised approaches and different lexicons, dictionaries and corpus based methods which are very helpful in Sentiment Analysis. Different dataset are available for movie review, product review, Epinions dataset etc. In this method sentiment score has been calculated and counted number of positive, negative and neutral tweets for given Hashtag and can predict the public opinion of particular event. As per above analysis of different #Hashtags tweets for sentiment analysis, individual and industry can find the public opinion behind that event.

CHAPTER 10

REFERENCES

- [1] J. Serrano-Guerrero et al., "A google wave-based fuzzy recommender system to disseminate information in University Digital Libraries 2.0," Information Sciences, Vol. 181, no.9, pp. 1503-1516, May. 2011.
- [2] Z. Zhang et al., "A hybrid fuzzy-based personalized recommender system for telecom products/services," Information Sciences, Vol. 235, pp. 117-129, Jun. 2013
- [3] Z.B. Sun et al., "Recommender systems based on social networks," Journal of Systems and Software, Vol. 99, pp. 109-119, Jan. 2015
- [4] X.L. Zheng et al., "A Hybrid Trust-Based Recommender System for Online Communities of Practice," IEEE Transactions on Learning Technologies, Vol. 8 no. 4, pp. 345-356. Apr. 2015.
- [5] L.O. Colombo-Mendoza et al., "RecomMetz: A context-aware knowledge-based mobile recommender system for movie showtimes," Expert Systems with Applications, Vol. 42, no. 3, pp. 1202-1222, Feb. 2015.
- [6] G.M.L. Sarne, "A novel hybrid approach improving effectiveness of recommender systems," Journal of Intelligent Information Systems, Vol. 44, no. 3, pp. 397-414, Jun. 2015.

- [7] A. Tejeda-Lorente et al., "REFORE: A recommender system for researchers based on bibliometrics," *Applied Soft Computing*, Vol. 30, pp. 778-791, May 2015.
- [8] R. Baraglia et al., "A peer-to-peer recommender system for self-emerging user communities based on gossip overlays," *Journal of Computer and System Sciences*, Vol. 79, no. 2, pp. 291-308, Mar 2013.
- [9] F. Sanchez et al., "Social and Content Hybrid Image Recommender System for Mobile Social Networks," *Mobile Networks & Applications*, Vol. 17, no. 6, pp. 782-795, Dec. 2012.
- [10] W. Hussein et al., "A Personalized Recommender System Based on a Hybrid Model," *Journal of Universal Computer Science*, Vol. 19, no.15, pp. 2224-2240, Sep. 2013.
- [11] M. Tavakolifard and K.C. Almeroth, "Social Computing: An Intersection of Recommender Systems, Trust/Reputation Systems, and Social Networks," *IEEE Network*, Vol. 26, no. 4, pp. 53-58, Jul-Aug 2012.
- [12] B. Amini et al., "A reference ontology for profiling scholar's background knowledge in recommender systems," *Expert Systems with Applications*, Vol. 42, no. 2, pp. 913-928, Feb. 2015.

CHAPTER 11

OUTPUTS

HOMEPAGE

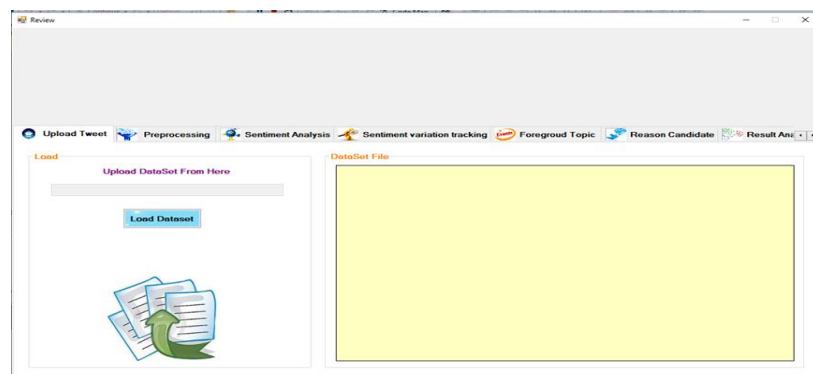


Fig 11.1

UPLOAD TWEET DATASET

Load

Upload DataSet From Here

D:\project output\commerce output\twit\final twit

Load Dataset

DataSet File

Retweet	Screen_ID	Time_Zone	Query	User_Name
0	2202060611	Tue Jun 16 21:10:48 PDT 2009	NO_QUERY	xrpggamemans
0	2202060926	Tue Jun 16 21:10:50 PDT 2009	NO_QUERY	laurin24
0	2202061050	Tue Jun 16 21:10:50 PDT 2009	NO_QUERY	mizhalle
0	2202061423	Tue Jun 16 21:10:53 PDT 2009	NO_QUERY	Bloodoffallout3
0	2202061625	Tue Jun 16 21:10:54 PDT 2009	NO_QUERY	UGWellness
0	2202061875	Tue Jun 16 21:10:57 PDT 2009	NO_QUERY	jenniferrose_
0	2202062698	Tue Jun 16 21:11:00 PDT 2009	NO_QUERY	CharlieBosco
0	2202062730	Tue Jun 16 21:11:00 PDT 2009	NO_QUERY	EricaGo13
0	2202062977	Tue Jun 16 21:11:01 PDT 2009	NO_QUERY	stefaunee
0	2202063114	Tue Jun 16 21:11:02 PDT 2009	NO_QUERY	PrincessPiper3
0	2202063313	Tue Jun 16 21:11:03 PDT 2009	NO_QUERY	HULICIOUS
0	2202064098	Tue Jun 16 21:11:08 PDT 2009	NO_QUERY	trejoy
0	2202064204	Tue Jun 16 21:11:08 PDT 2009	NO_QUERY	nathanmackey
0	2202064482	Tue Jun 16 21:11:10 PDT 2009	NO_QUERY	le_le_f
0	2202064662	Tue Jun 16 21:11:11 PDT 2009	NO_QUERY	regulus4

Fig 11.2

PRE-PROCESSING:

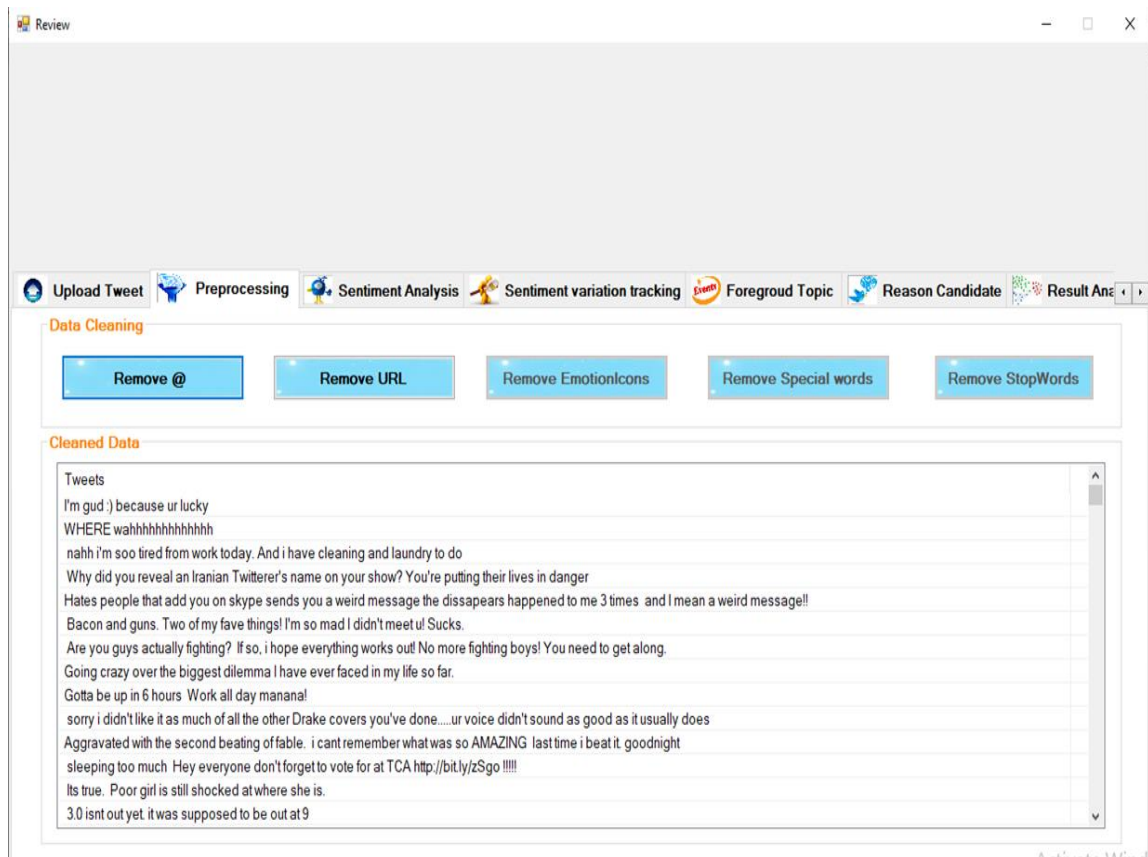


Fig 11.3