

Sentiment Analysis Using Contextual Approach For E-Commerce Reviews

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Engineering degree in Computer Science and Engineering

by

G Sri harsha (17SCS8042)



**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 | 12B Status by UGC | Approved by
AICTE**

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

March – 2021



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with “A” grade by NAAC | 12B Status by UGC | Approved by
AICTE

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this project report is the bonafide work of **G SRI HARSHA (17SCS8042)** who carried out the project entitled “**Sentiment Analysis Using Contextual Approach For E-Commerce Reviews**” under my supervision from **January 2021** to **March 2021**.

Internal Guide

Ms. Yogitha.R M.E.,

Head of the Department

Submitted for Viva voce Examination held on_____

Internal Examiner

External Examiner

DECLARATION

I **G SRI HARSHA** hereby declare that the Project Report entitled “**Sentiment Analysis Using Contextual Approach For E-Commerce Reviews**” is done by me under the guidance of **Ms. Yogitha.R M.E.**, Department of Computer Science and Engineering at Sathyabama Institute of Science and Technology is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my 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.

I convey my thanks to **Dr. T.Sasikala M.E., Ph.D .**, Dean, School of Computing and **Dr.L.Lakshmanan M.E., Ph.D.,and Dr.S.Vigneshwari M.E., Ph.D.**, Heads of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Ms. Yogitha.R M.E.**, for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my 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.

ABSTRACT

Sentiment analysis makes help people understand the opinion of other customers to other people. There are many methods for sentiment analysis but were not accurate because there are many input variables for training the model which takes more time. In this, work we will use feature selection technique which will reduce the computation cost. we will use clustering technique for grouping the opinions of the customers. the extracted opinions from the reviews are given as input for classification which can be used for prediction of remaining reviews. The reviews from internet like google, facebook are used as variant. The random forest and naive bayes are used while classifying

LIST OF ABBREVIATIONS

SVM	SUPPORT VECTOR MACHINE
NN	NEURAL NETWORKS
GUI	GRAPHICAL USER INTERFACE
SQL	STRUCTURED QUERY LANGUAGE

TABLE OF CONTENTS

CHAPTER No.	TITLE	PAGE No.
	ABSTRACT	v
	LIST OF ABBREVIATIONS	vi
1	INTRODUCTION	1
	1.1 INTRODUCTION	1
	1.2 OBJECTIVES OF .NET	2
2	AIM & SCOPE OF THE PROJECT	3
	2.1 AIM	3
	2.1.1 OBJECTIVE	3
	2.2 SCOPE	3
	2.3 PROPOSED SYSTEM	4
	2.4 LITERATURE SURVEY	4
3	SYSTEM DESIGN AND METHODOLOGY	7
	3.1 EXISTING SYSTEM	7
	3.2 MODULES	7
	3.2.1 MODULES DISCRIPTION	8
	3.3 REQUIREMENT SPECIFICATIONS	9
	3.4 SYSTEM ARCHITECTURE	10
	3.5 DATA FLOW DIAGRAMS	11
	3.6 ALGORITHMS AND TECHNIQUES	16
4	SOFTWARE DESCRIPTION AND DEVELOPMENT	22
	4.1 VISUAL STUDIO	22
	4.2 SQL SERVER	25

	4.2.1 SQL SERVER MANAGEMENT STUDIO	26
	4.2.2 TABLE	28
	4.2.3 FEASIBILITY STUDY	28
	4.4 TESTING	29
	4.4.1 TESTING PROCESS	29
	4.4.2 TESTING STRATEGY AND APPROCH	30
5	RESULTS AND DISCUSSIONS	33
6	6.1 CONCLUSION	40
	6.3 FUTURE WORK	40
	REFERENCES	41
	SOURCE CODE	42

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

.NET simplifies application development in highly distributed environment using internet. Machine learning has the ability to automatically learn and improves from the experience without explicitly programmed. Supervised machine learning helps to predict the incoming data by previously learned data that is labelled. Unsupervised machine learning algorithms will try to find the similarity in data .

Bayesian network algorithm which plays chief role in sentiment analysis is the main aspect of research in machine learning and python. Bayesian network is a model which tells set of variables and their dependencies. we will use decision tree algorithm also to create a training model.

First step is to extract the opinion from the reviews and creating a wordnet dictionary where positive and negative opinions are will be written out separately. the second step is to classifying and predicting the reviews.

This Process of sentiment analysis consist of steps like feature selection, word And polarity identification, classification and reviews analysis.

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

CHAPTER 2

AIM & SCOPE OF THE PROJECT

2.1 Aim of Project

The primary aim of the proposed system is to classify whether that event is given positive or negative feedback so that the other users can decide whether to go to that event or not. And to socialize the auction so that people from far & wide and even across the continent can participate in it.

2.1.1 Objective

Our objective is to propose an algorithm for sentiment analysis based on rules and a dimensional sentiment lexicon for reducing the computational cost.

2.2 Scope

The application is based on sentiment analysis in big data to become common computing power is one of the main factors that led to techniques for analysis new and divers digital data spreading emotion and response to ward local events analysis social media is to understand.

2.2.1 ADVANTAGES

- Collaborative Filtering
- Content Based Filtering
- Clustering
- Categorization

2.3 Proposed System

- Recommender System (RS): Special type of information filtering system that provides a prediction that assists the user in evaluating items from a large collection that the user is likely to find interesting or useful.
- Status update (micro post): Short message, shared in an online social platform, expressing an activity, state of mind or opinion.
- Folksonomy: Whole set of tags that constitutes an unstructured collaborative knowledge classification scheme in a social tagging system.
- This step involves identifying and extracting those specific product features and the opinions on them.

2.4 Literature survey

This paper based on the interaction between users through collaborative applications such as wikis, blogs. the concept wave a common space where the Resources and users can work together. if there are more number of users or resources then problem will arises. For this a fuzzy linguistic recommender system based on the google wave capabilities as a tool for communicating researchers. the system allows the creation of a common space by means a wave as a way for exchanging ideas between several researchers interested in common research line[1]

This paper is about recommender systems which are not accurate due to uncertainties exist with in both product and customer data it is very challenge. To achieve high recommendation accuracy. This develops a hybrid recommendation approach which combines user-Based and item based collaborative filtering techniques with fuzzy set technique.[2]

This paper based on many researchers who have ignored the social relationship among users. a social regularization is proposed that incorporates social network information to benefit recommender systems. user's friendships and rating records are employed to predict the missing values in user-item matrix. Bi-clustering algorithm is used for identifying the most dutiable group of friends for generating final recommendations. The correlation between user and items is taken in to consideration.[3]

This paper tells about how the life-long learning and development of information technologies has benefited various online community of practices This study proposes a hybrid, trust-based recommender system to mitigate the issues in online cops. recommender system can facilitates the formation of personalized learning communities.[4]

This paper states on a recommender system in the leisure domain for movie showtimes. the proposed system Recom-Metz which is a context-aware mobile recommender system based on sematic web technologies. Ontology serves semantic similarity metric adjusted to the concept of packages of single items was developed.[5]

This paper based on how the performances of recommender system are effected by issues like instance, attribute selection and scalability. To make this effects a hybrid recommender system relevance based recommender is proposed.it exploits individual measures of perceived relevance computed by each user for each instance of interest and to obtain best precision.[6]

this paper based on refore and quality based fuzzy linguistic recommender system for researchers. the use of some bibliometric measures to qualify the both items and users without interaction of experts by the use of 2-tuple linguistic approach to describe the linguistic information. the system takes measured quality as main factor for the re-ranking of the top-N.[7]

This paper is about previous recommender systems are not able to cope with the uncertainty of automated tagging and knowledge domain dependent. The proposed by a hybrid image recommender system, which combines collaborative filtering with content-based techniques.it improves the performance of existing system for creating a mobile social networks recommender .[8]

This paper has gossip-based peer-to-peer protocols which are very efficient for dynamic and complex information exchange among distributed peers. These peers are useful for building and maintaining the network topology and to support pervasive diffusion of information injection into network. we will proposed general architecture of a system whose aim is to exploit the collaborative exchange of information between peers in order to build a system.[9]

This paper is about proposing a personalized recommender system for next page prediction based on for hybrid model for both memory-based and model-based system. the proposed system improves the speed over traditional model-based mining and accuracy.[10]

CHAPTER 3

SYSTEM DESIGN AND METHODOLOGY

3.1 EXISTING SYSTEM

- E-commerce reviews reveal the customers' attitudes on the products, which are very helpful for customers to know other people's opinions on interested products. Meanwhile, producers are able to learn the public sentiment on their products being sold in E-commerce platforms.
- Generally, E-commerce reviews involve many aspects of products, e.g., appearance, quality, price, logistics, and so on.
- Therefore, sentiment analysis on E-commerce reviews has to cope with those different aspects.
- The problem with public auction is that the participation of the general public is very limited.

3.2 MODULES

- Collection of User's Reviews
- Pre-Processing

- Feature Selection
- Sentiment Word Identification
- Sentiment Polarity Identification
- Sentiment Classification
- Analysis of Reviews

3.2.1 Modules Description

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.

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

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.

Sentiment Word Identification

Sentiment word identification is a fundamental work in numerous applications of sentiment analysis and opinion mining, such as review mining, opinion holder finding, and review classification. Sentiment words can be classified into positive, negative and neutral words.

Sentiment Polarity Identification

The basic task in SA is classifying the polarity of a given text at the document, sentence, or feature. The polarity is in three category i.e. Positive, Negative and Neutral. Polarity identification is done by using different lexicons which help to calculate sentiment score, sentiment strength etc.

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.

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.

3.3 REQUIREMET SPECIFICATION

This proposed software runs effectively on a computing system that has the minimum requirements. Undertaking all the equipment necessities are not satisfied but rather exist in their systems administration between the customer's machines already. So the main need is to introduce appropriate equipment for the product. The requirements are split into two categories, namely:

SOFTWARE REQUIREMENTS

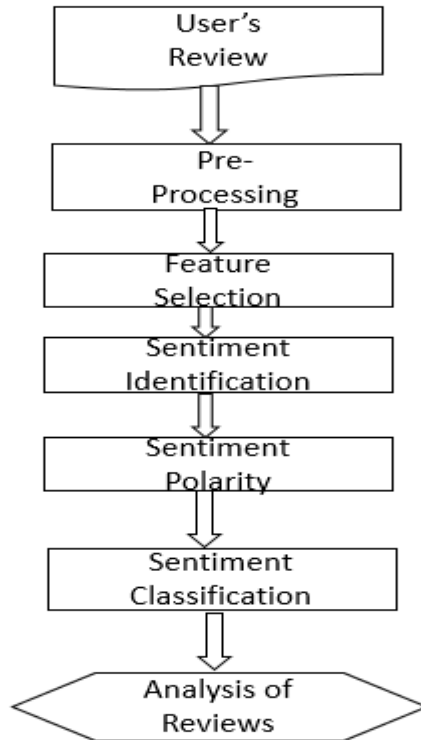
- Operating system : Windows
- Coding Language : .Net
- IDE : Visual studio2013
- Database : SQL Server

HARDWARE REQUIREMENTS

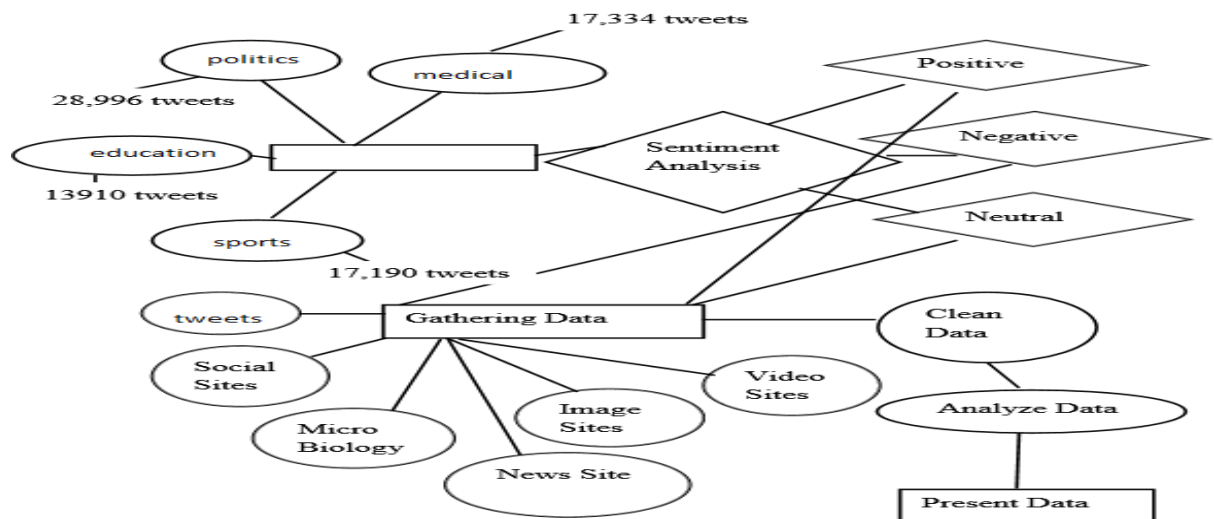
- System : i3 Processor
- Hard Disk : 500 GB.
- Monitor : 15"LED
- Input Devices : Keyboard, Mouse
- Ram : 4GB.

3.4SYSTEMARCHITECTURE

□ It permits analyst to isolate areas of interest in the organization and study them by examining the data that enter the process and viewing how they are altered when they leave.



ER DIAGRAMS



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.

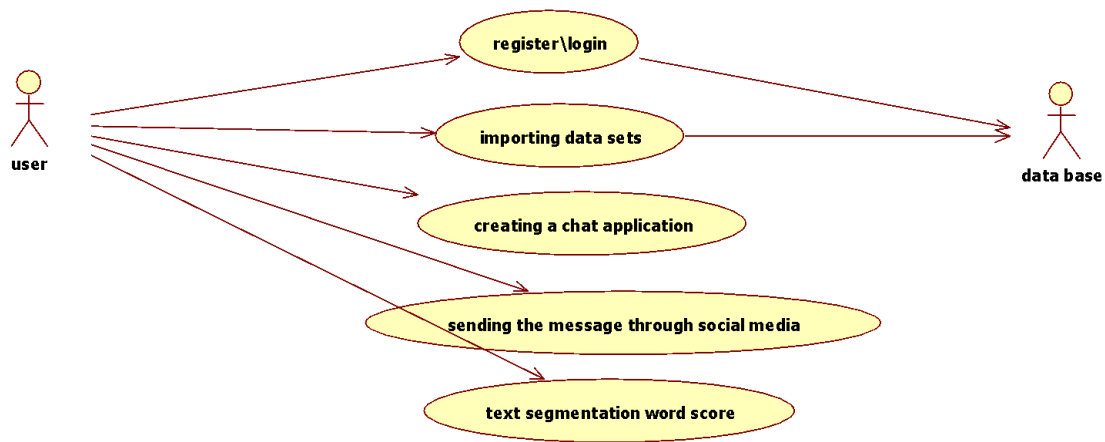
GOALS:

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.

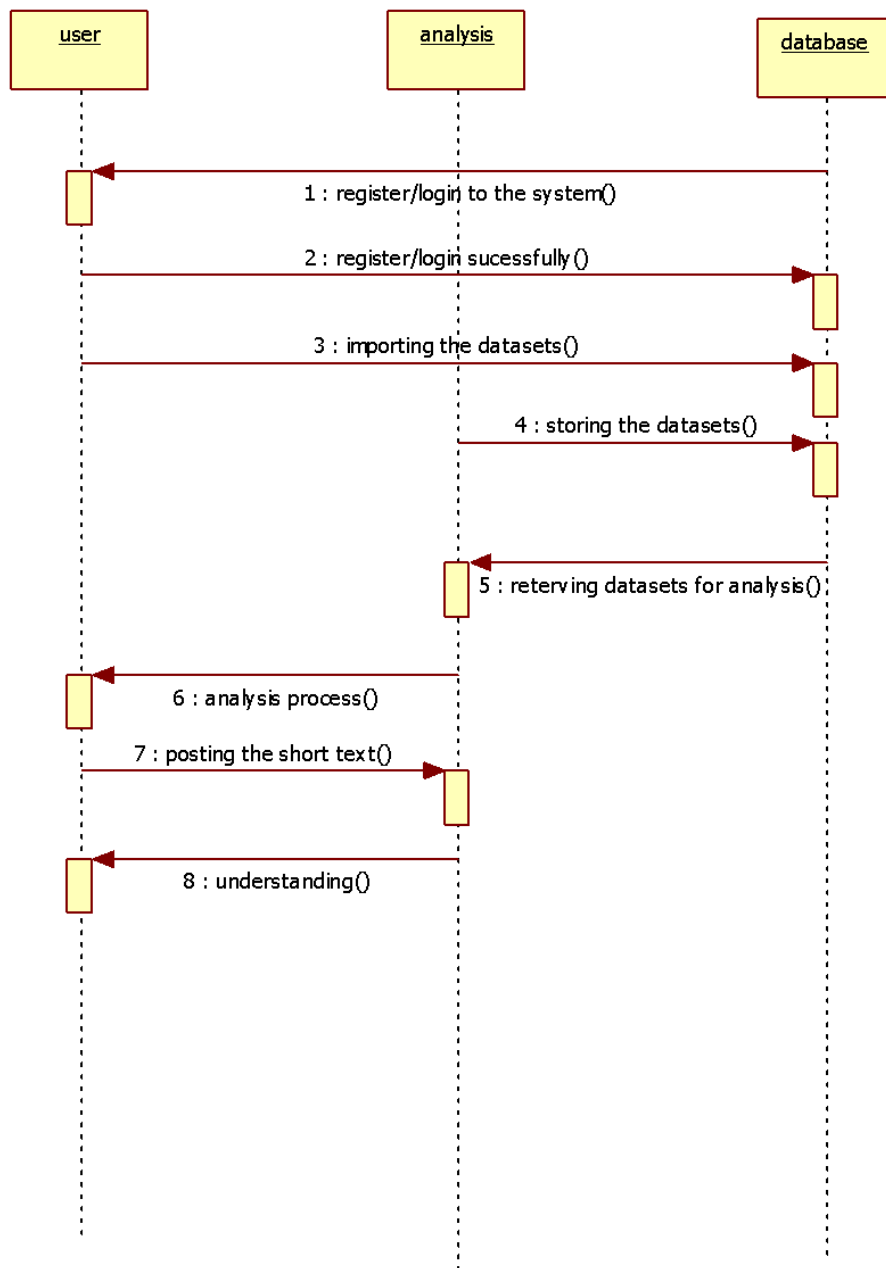
USE CASE DIAGRAM

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.

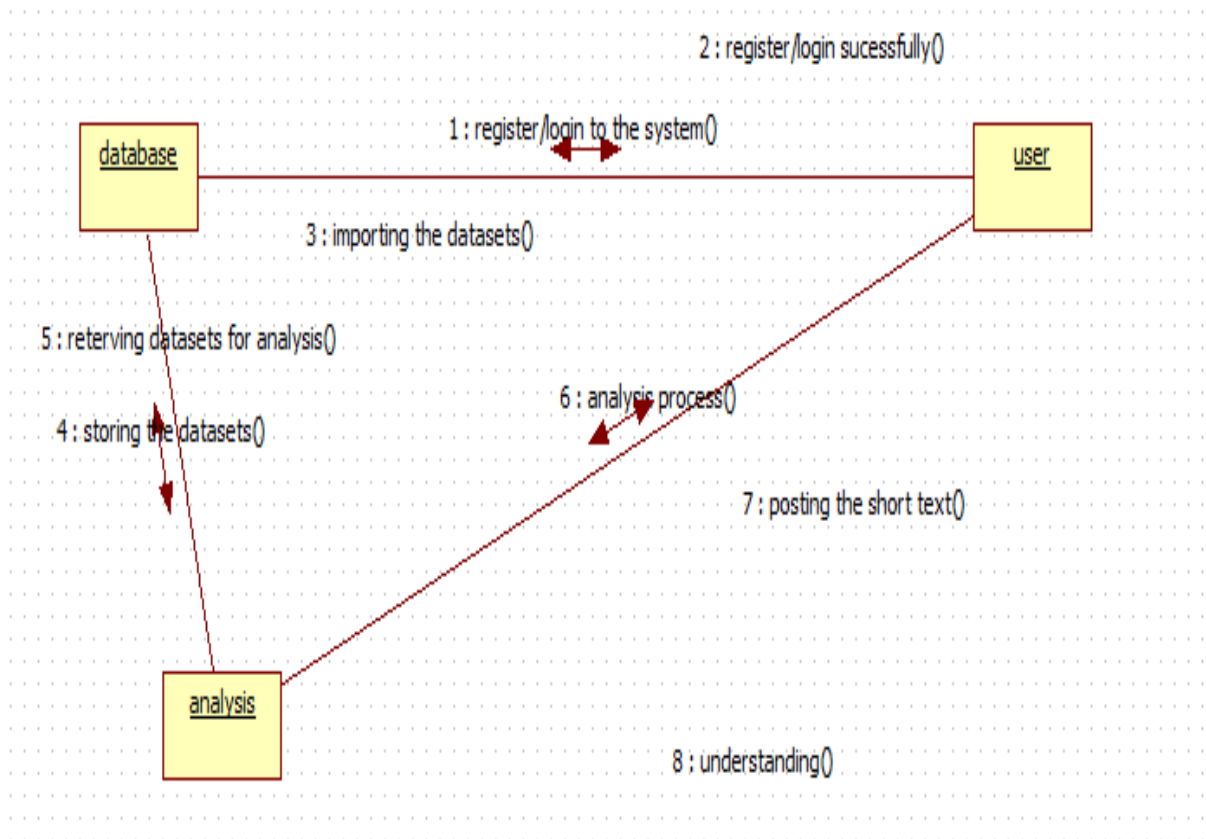


SEQUENCE DIAGRAM

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.



COLLABRATION DIAGRAM



3.6 Algorithms and techniques

Collaborative filtering

collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating). The underlying assumption of the collaborative filtering approach is that if a person *A* has the same opinion as a person *B* on an issue, *A* is more likely to have *B*'s opinion on a different issue than that of a randomly chosen person. For example, a collaborative filtering recommendation system for television tastes could make predictions about which television show a user should like given a partial list of that user's tastes (likes or dislikes). Note that these predictions are specific to the user, but use information gleaned from many users. This differs from the simpler approach of giving an average (non-specific) score for each item of interest, for example based on its number of votes.

In the more general sense, collaborative filtering is the process of filtering for information or patterns using techniques involving collaboration among multiple agents, viewpoints, data sources, etc. Applications of collaborative filtering typically involve very large data sets. Collaborative filtering methods have been applied to many different kinds of data including: sensing and monitoring data, such as in mineral exploration, environmental sensing over large areas or multiple sensors; financial data, such as financial service institutions that integrate many financial sources; or in electronic commerce and web applications where the focus is on user data, etc. The remainder of this discussion focuses on collaborative filtering for user data, although some of the methods and approaches may apply to the other major applications as well.

Clustering technique

clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including pattern,recognition, imageanalysis, retrieval, bioinformatics, data compression, computer graphics and machine learning.

Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their understanding of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances between cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings (including parameters such as the distance function to use, a density threshold or the number of expected clusters) depend on the individual data set and intended use of the results. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It is often

necessary to modify data preprocessing and model parameters until the result achieves the desired properties.

Besides the term *clustering*, there are a number of terms with similar meanings, including *automatic classification*, *numerical taxonomy*, *botryology* (from Greek βότρυς "grape"), *typological analysis*, and *community detection*. The subtle differences are often in the use of the results: while in data mining, the resulting groups are the matter of interest, in automatic classification the resulting discriminative power is of interest.

Bayesian network

Bayesian network (also known as a Bayes network, belief network, or decision network) is a probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG). Bayesian networks are ideal for taking an event that occurred and predicting the likelihood that any one of several possible known causes was the contributing factor. For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases.

Efficient algorithms can perform inference and learning in Bayesian networks. Bayesian networks that model sequences of variables (e.g. speech signals or protein sequences) are called dynamic Bayesian networks. Generalizations of Bayesian networks that can represent and solve decision problems under uncertainty are called influence diagrams.

Content based filtering

Another common approach when designing recommender systems is content-based filtering. Content-based filtering methods are based on a description of the item and a profile of the user's preferences. These methods are best suited to situations where there is known data on an item (name, location, description, etc.), but not on the user. Content-based recommenders treat recommendation as a user-specific classification

problem and learn a classifier for the user's likes and dislikes based on an item's features.

In this system, keywords are used to describe the items and a user profile is built to indicate the type of item this user likes. In other words, these algorithms try to recommend items that are similar to those that a user liked in the past, or is examining in the present. It does not rely on a user sign-in mechanism to generate this often temporary profile. In particular, various candidate items are compared with items previously rated by the user and the best-matching items are recommended. This approach has its roots in information retrieval and information filtering research.

To create a user profile, the system mostly focuses on two types of information:

- A model of the user's preference.
- A history of the user's interaction with the recommender system.

Naive bayes

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. There is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness, and diameter features.

For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian

probability or using any Bayesian methods.

- Calculate the prior probability for given class labels.
- Find Likelihood probability with each attribute for each class.
- Put these value in Bayes Formula and calculate posterior probability

Random forest

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set. Random forest is a type of supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of learning where you join different types of algorithms or same algorithm multiple times to form a more powerful prediction model. The random forest algorithm combines multiple algorithm of the same type i.e. multiple decision *trees*, resulting in a *forest of trees*, hence the name "Random Forest". The random forest algorithm can be used for both regression and classification tasks.

The following are the basic steps involved in performing the random forest algorithm:

- Pick N random records from the dataset.
- Build a decision tree based on these N records.
- Choose the number of trees you want in your algorithm and repeat steps 1 and 2.
- In case of a regression problem, for a new record, each tree in the forest predicts a value for Y (output). The final value can be calculated by taking the average of all the values predicted by all the trees in forest. Or, in case of a classification problem, each tree in the forest predicts the category to which

the new record belongs. Finally, the new record is assigned to the category that wins the majority vote.

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.

CHAPTER 4

SOFTWARE DESCRIPTION AND DEVELOPMENT

4.1 visual studio

Microsoft Visual Studio is an integrated development environment (IDE) from Microsoft. It is used to develop computer programs. It supports programming languages like visual basic .NET, C# and 34 more different languages.

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 Forms
XML WEB SERVICES	
Base Class Libraries	
Common Language Runtime	
Operating System	

Figure. 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.

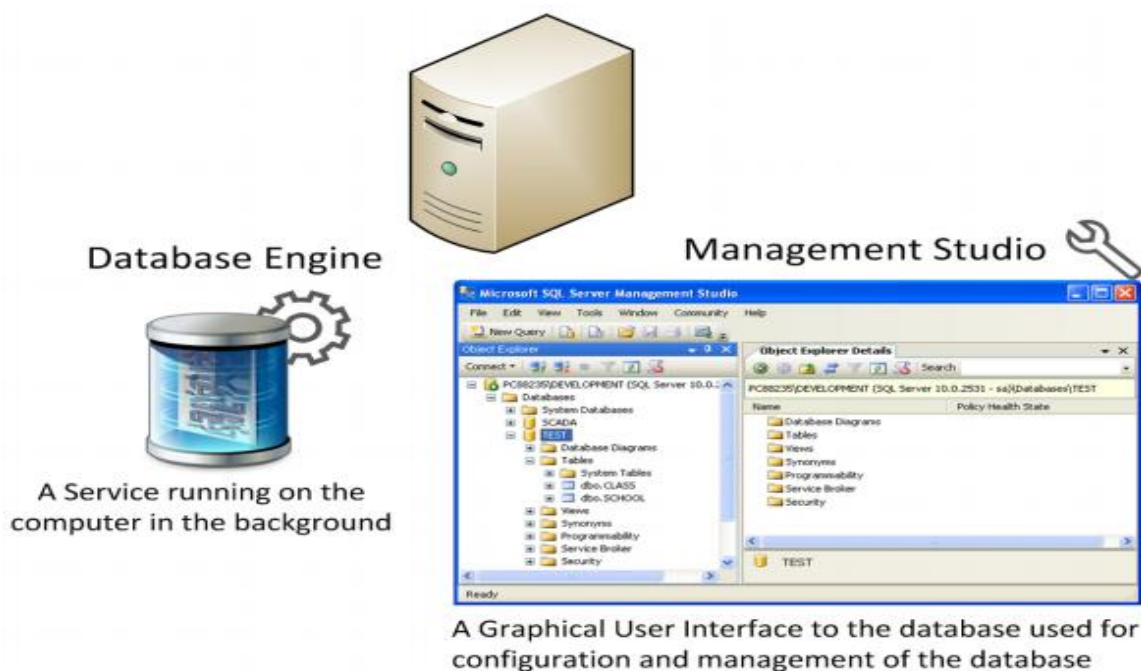
4.2 SQL SERVER

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.



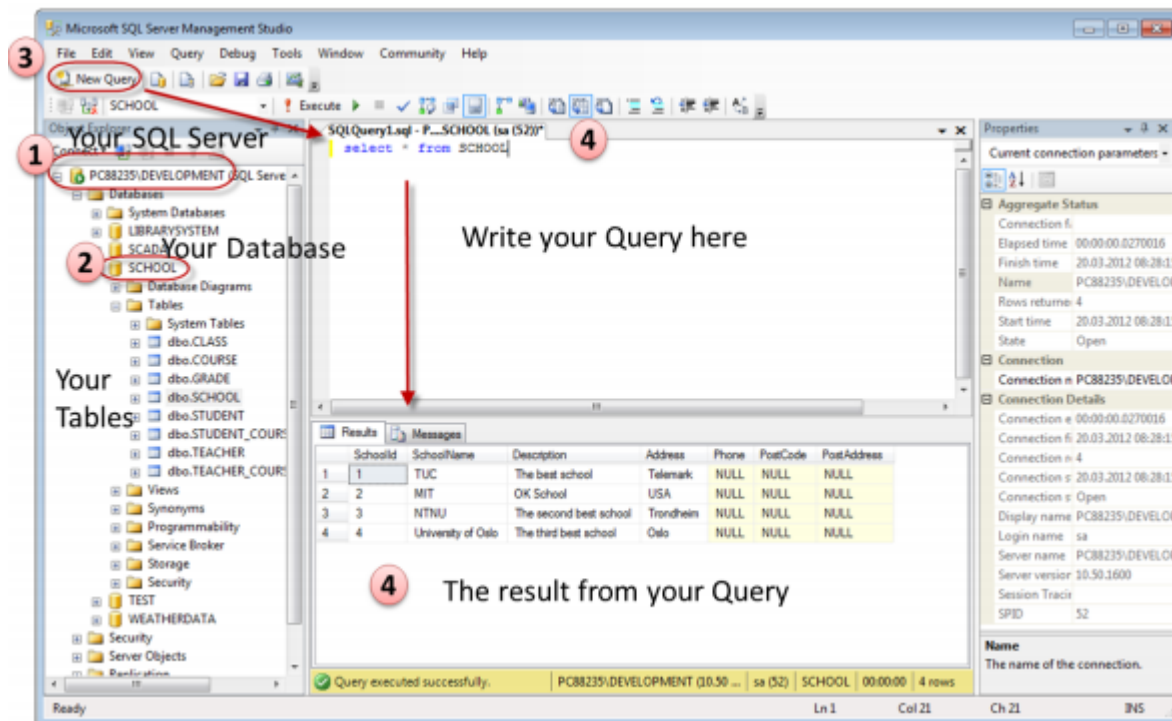
4.2.1SQL SERVER MANAGEMENT STUDIO

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.

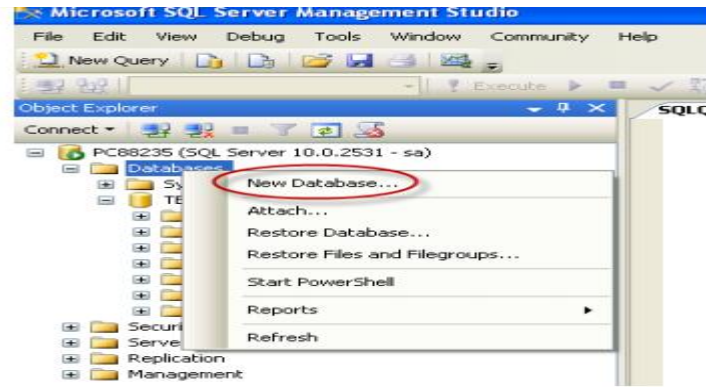
A central feature of SQL Server Management Studio is the Object Explorer, which allows the user to browse, select, and act upon any of the objects within the server. It can be used to visually observe and analyze query plans and optimize the database performance, among others. SQL Server Management Studio 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).



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...”



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

4.2.2Table

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.

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.

4.2.3Testing

TESTING PROCESS

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product it is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TEST STRATEGY AND APPROACH

Field testing will be performed manually and functional tests will be written in detail.

Test Objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.
- Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct

Integration Testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional Testing

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation and user manuals.

Functional testing is centered on the following items:

Valid Input is used to identified classes of valid input must be accepted.

Invalid Input is used to identified classes of invalid input must be rejected.

Functions is used to identified functions must be exercised.

Output is used to identify classes of application outputs.

Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications.

Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

ALPHA TESTING

In software development, alpha test will be a test among the teams to confirm that your product works. Originally, the term alpha test meant the first phase of testing in a software development process. The first phase includes unit testing, component testing, and system testing. It also enables us to test the product on the lowest common denominator machines to make sure download times are acceptable and preloads work.

CHAPTER 5

RESULTS AND DISCUSSION

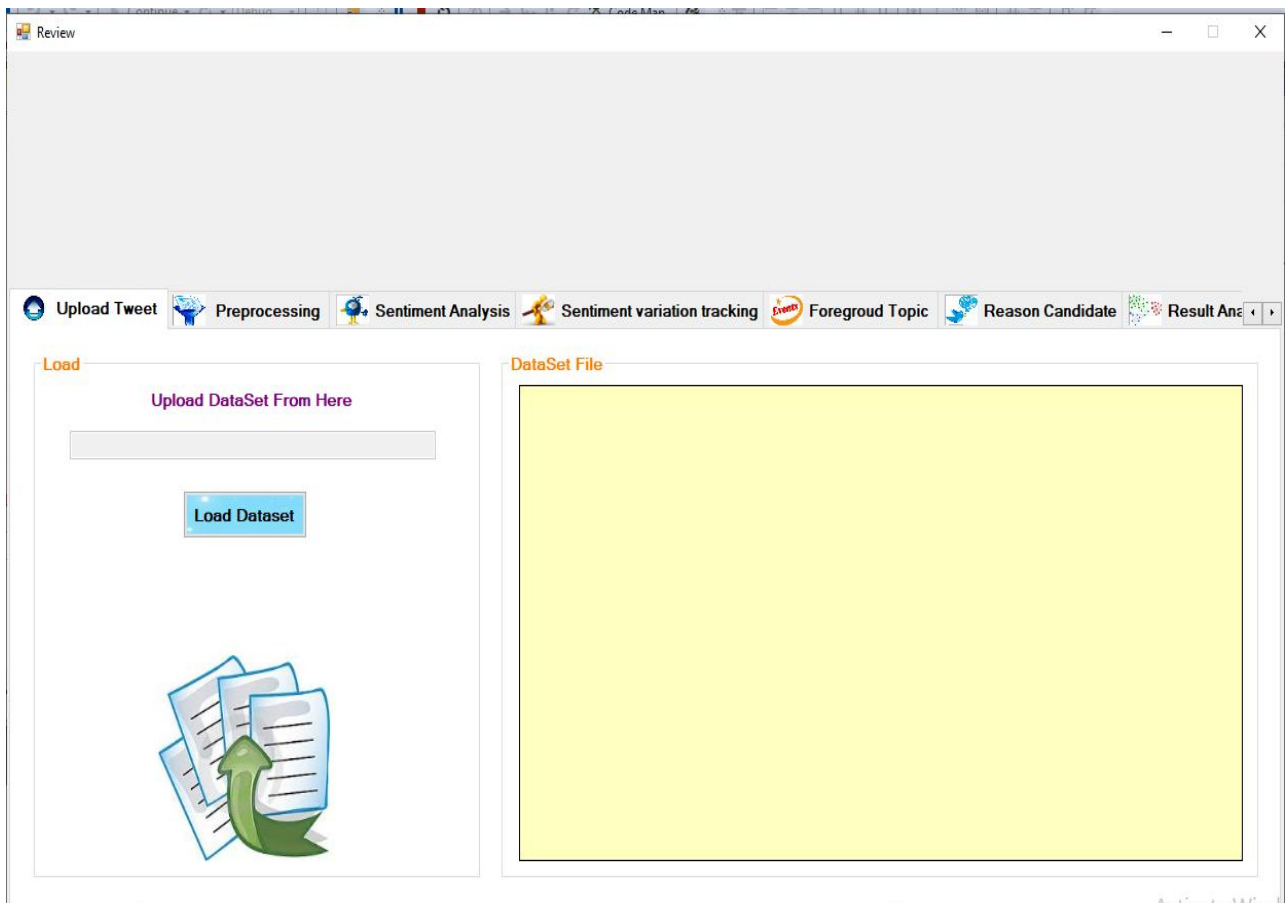


Fig 5.1 Home Page

The Fig 4.1 the home page of the project. This page will appear once the user run the program in Microsoft visual studio. This has a load dataset button where the reviews collected from the twitter will have to be selected from the pc where it's stored. After clicking on the dataset wait until the white progress bar will be filled with the green shade, which indicated that the dataset has been uploaded. The format of the dataset has to be excel sheet.

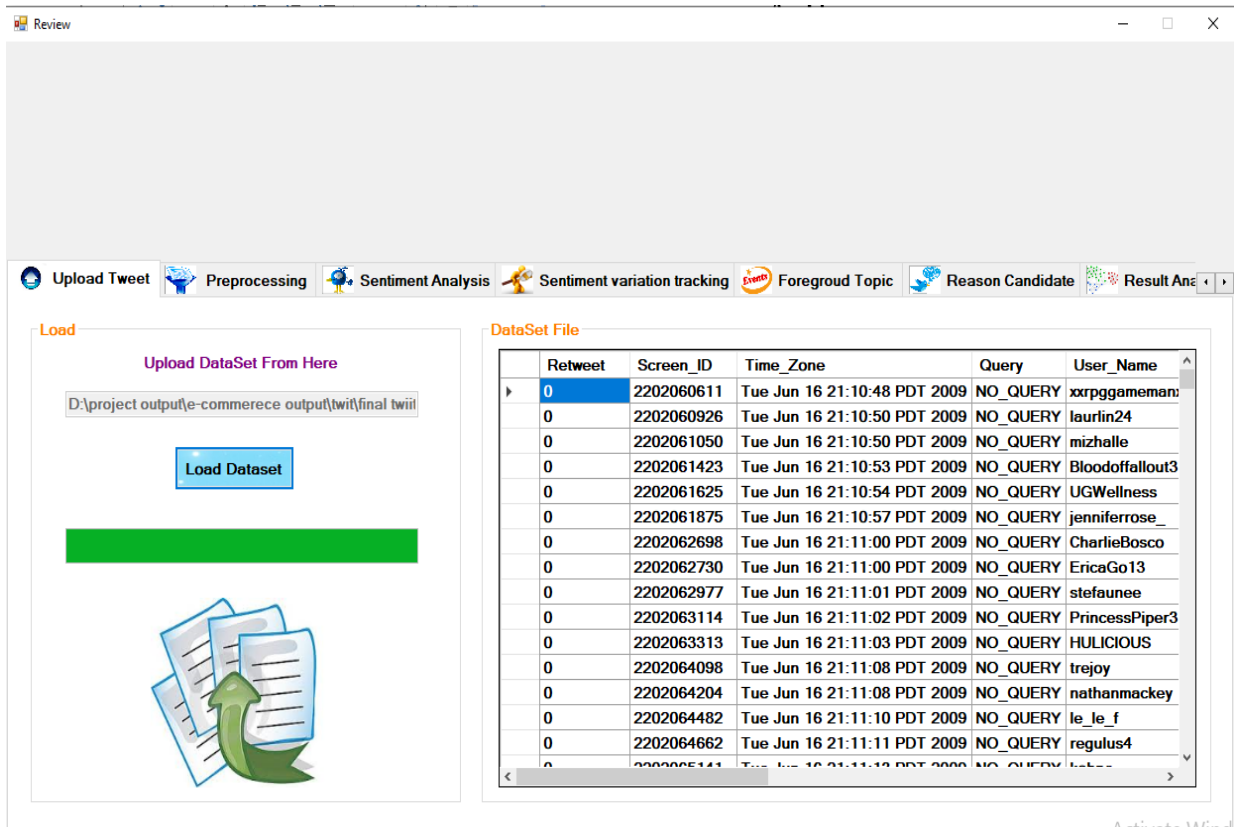


Fig 5.2 dataset uploaded

The Fig 4.2 shows the dataset and all the parameters which have been uploaded from the system. And the dataset location where it has been stored is also shown above the load dataset button. Retweet, Screen_ID ,Time_Zone ,Query ,User_Name ,Tweets are all the parameters which are used. These tweets are collected using R tool which Is a open source.

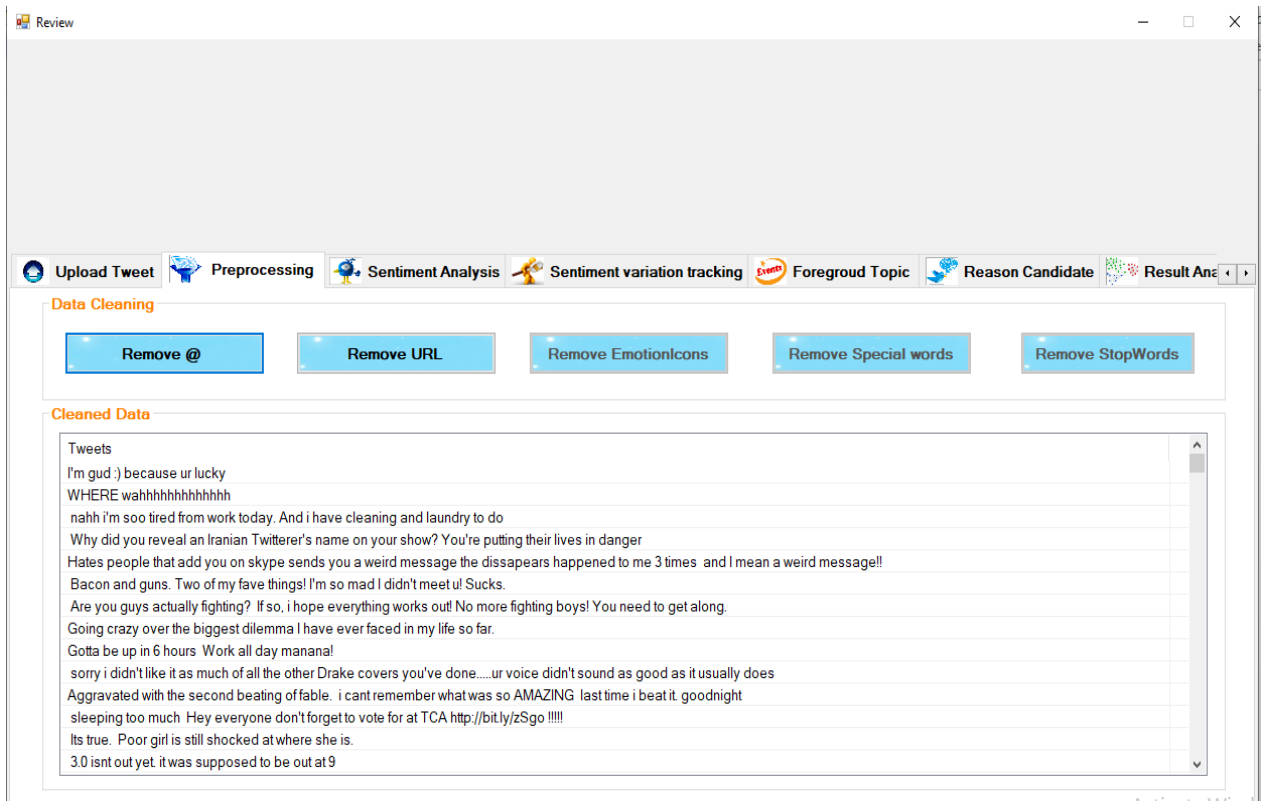


Fig 5.3 Pre-processing

This Fig 4.3 is the pre-processing place where the data line inconsistent and noisy will be removed. First @ symbol from all the tweets will be removed then followed by url or weblinks, and emotion icons like smiles symbols will be removed, and words like special words ,stop words are those words like what, do where it is already defined by used in the dictionary format.

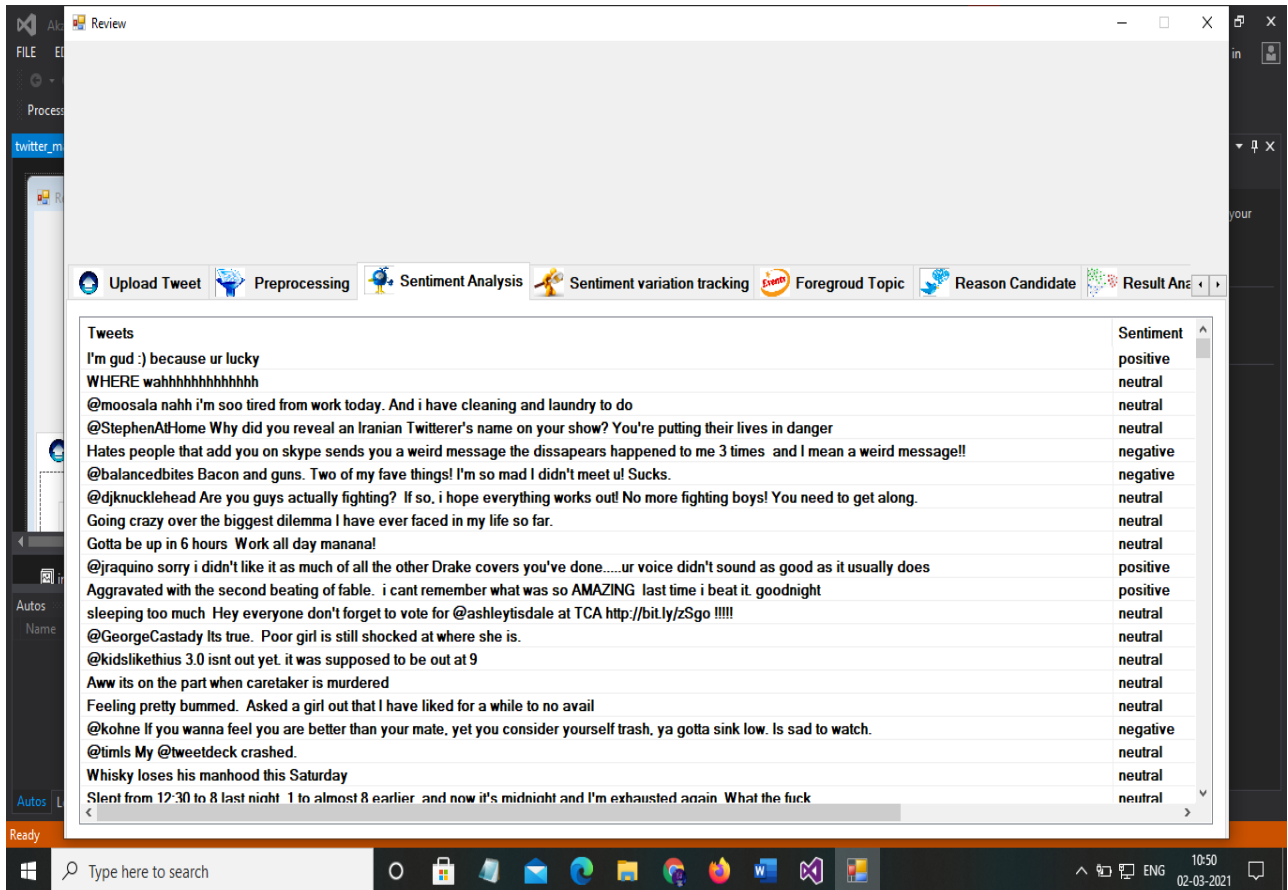


Fig 5.4 Sentiment Analysis

The Fig 4.4 The sentiment of each tweet is shown in the above figure with tweets and sentiment. This is done using the sentiment lexicons which are already give by the user. In polarity identification lexicons are used to calculate the sentiment score and strength which is helpful for finding the sentiment. The sentiment is classified in to three positive which means the event is good, negative which means the event is not good, neutral which is average nor good not bad.

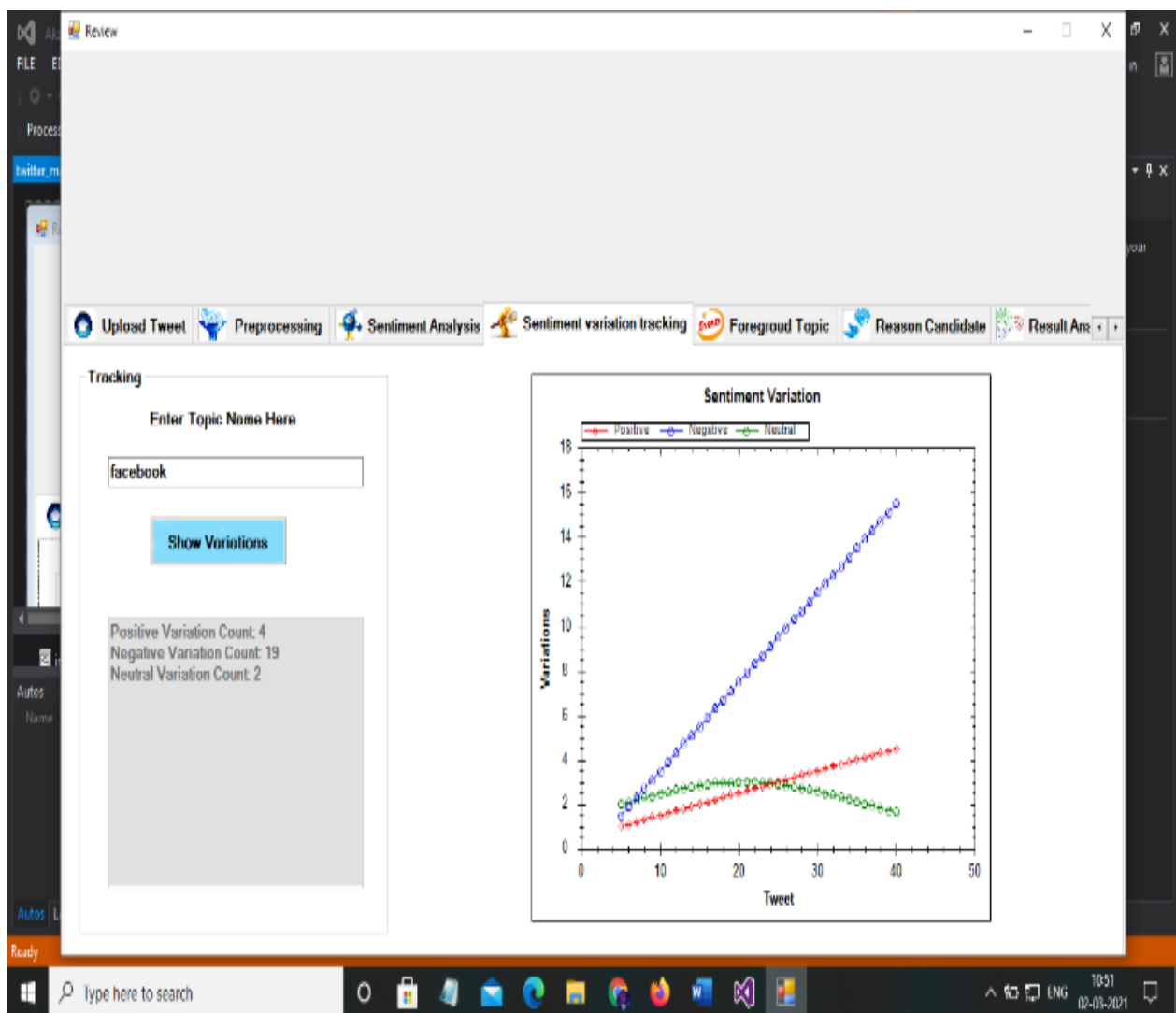


Fig 5.5 sentiment Variation

This Fig 4.5 has two sections one is tracking where the topic related reviews

Will be shown for example you need about facebook event. The type facebook and click on show variation button where the below the button it will display count of all positive ,negative ,netural reviews related to the topic which have been enterd and graph is also shown between the tweets. Positive are drawn with red color, negative is drawn with blue and neutral are drawn with green.

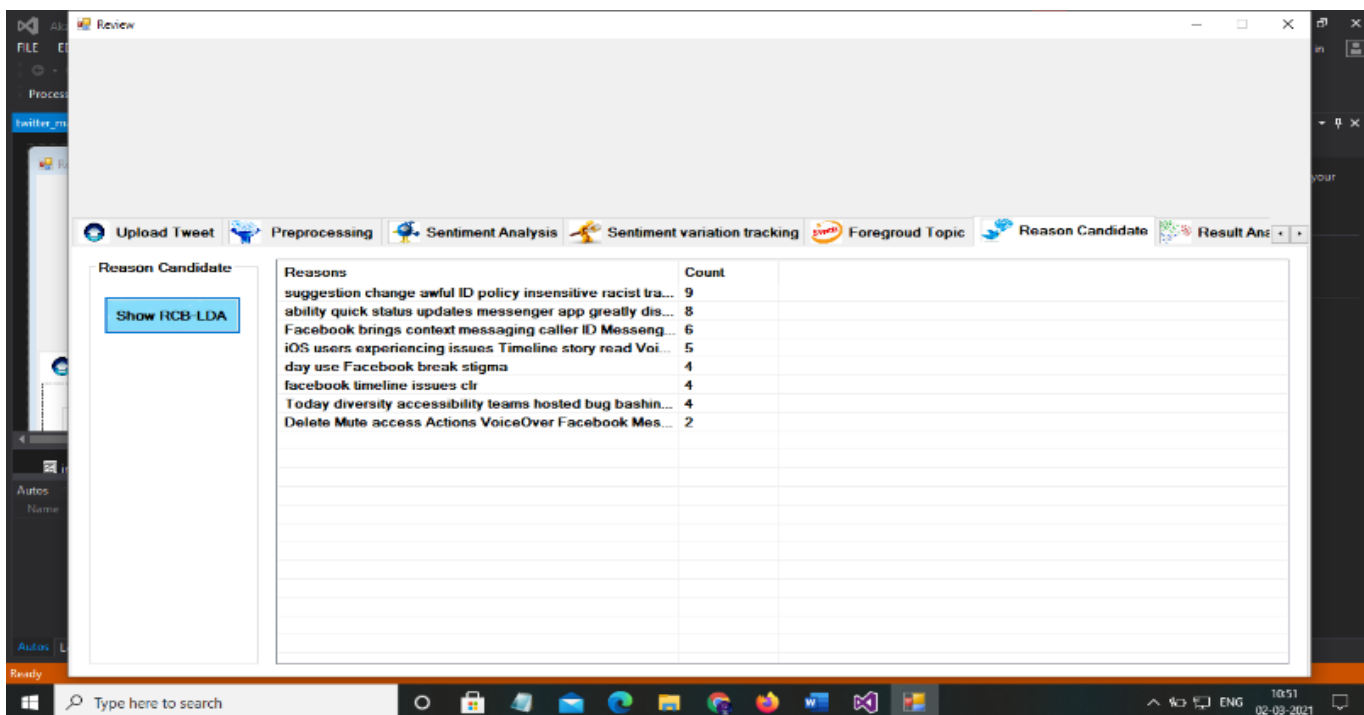


Fig 5.6 Reason candidate

This Fig 4.6 shows how many similar reviews are there in the whole dataset and reasons with similar meaning count is also displayed to the topic which ha been entered in sentiment variation tracking.

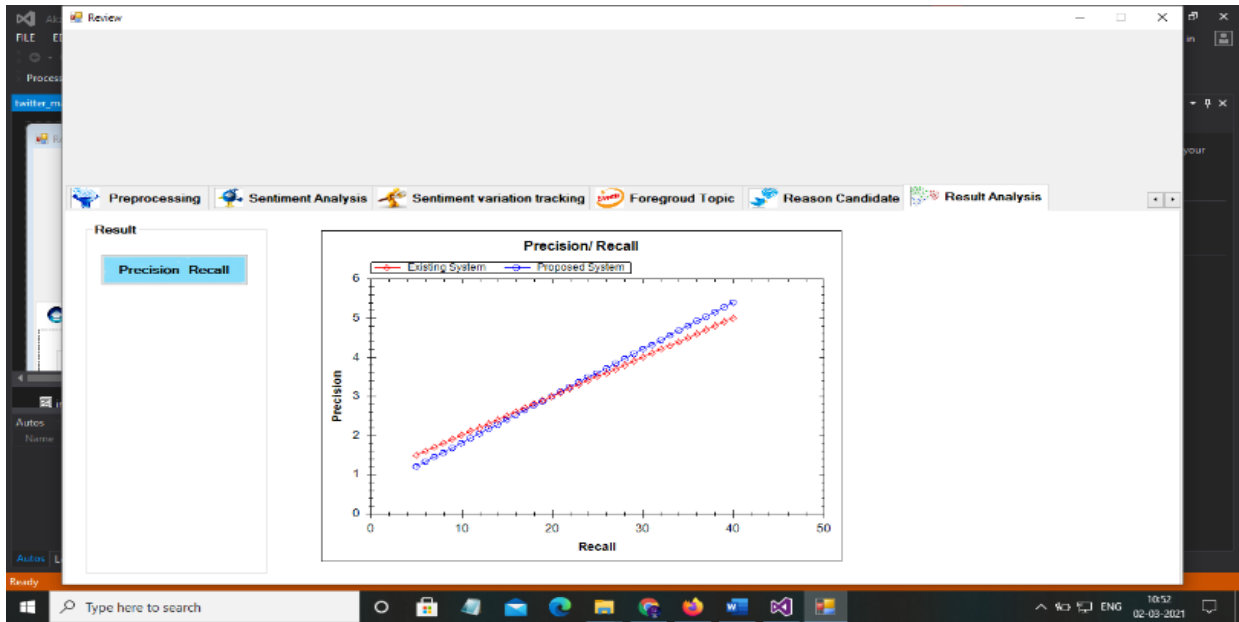


Fig 5.7 result analysis

The Fig 5.7 has the comparison between random forest and naive bayes. In this, the precision on the y-axis represents relevant among retrieved instances, and recall on the x-axis represents fraction of relevant instances retrieved.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 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, opinions 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.

6.2 FUTURE WORK

Future work about product review sentiment analysis is find out aspects and their polarity of the product which helps for consumer to take decision to buy online products on e-commerce site. Aspect level sentiment analysis gives detail information about that product and about movie review, director of that movie should know what exactly user know from particular movie which is possible in aspect based sentiment analysis. In Hotel reviews same like movie, hotel owner should know what items people like from their hotel and what other items need for customers.

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.

SOURCE CODE:-

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;

namespace Twitter_Event_Detection
{
    class NB
    {
        public enum FeatureType
        { Regular,
          RegularRemovingSmallWordsAndNumbers,
          AugmentDictionaryWithWordPairs10,
          AugmentDictionaryWithWordPairs30,
          AugmentDictionaryWithWordTriads5,

        }

        private FeatureType featType;
        public void NaiveBayes()
        {
            this.k = 1;
            this.featType = FeatureType.RegularRemovingSmallWordsAndNumbers;
        }
    }
}
```



```

TotalSamples++;
sample = sample.ToLower();
Classification = Classification.ToLower();

if (curCateg < 0)
{
    AllCategories.Add(Classification);

    _Dictionaries.Add(new SortedDictionary<string, int>());

    CatOccurrences.Add(0);
    TotalWords.Add(0);
    curCateg = AllCategories.Count - 1;
}

CatOccurrences[curCateg]++;

//Current category name and dictionary
string strCateg = AllCategories[curCateg];

SortedDictionary<string, int> curDict = _Dictionaries[curCateg];
//List<string> curDic = Dictionaries[curCateg];
//List<int> curWordOccur = WordOccurrences[curCateg];

//Splits sample words
string[] words = sample.Split(cc, StringSplitOptions.RemoveEmptyEntries);
words = GetFeatures(words);

TotalWords[curCateg] += words.Length;

//adds words to the dictionary
foreach (string s in words)
{
    if (!curDict.ContainsKey(s)) curDict.Add(s, 1);
    else curDict[s] = curDict[s] + 1;

    if (!GlobalDictionary.ContainsKey(s)) GlobalDictionary.Add(s, 1);
    else GlobalDictionary[s]++;
}
}
#endregion

#region Joint probabilities and classification

/// <summary>Gets Probab(Classification)</summary>

```

```

/// <param name="ind">Index of category</param>
public float GetCategoryProbability(int ind)
{
    int catOccurrences = 0;
    if (ind >= 0) catOccurrences = CatOccurrences[ind];

    return ((float)catOccurrences + k) / ((float)TotalSamples + k * (float)AllCategories.Count);
}

/// <summary>Computes  $P(\text{word} \mid \text{Category})$ </summary>
/// <param name="word">Word</param>
/// <param name="ind">Index of category</param>
public float GetWordProbability(string word, int ind)
{
    if (ind < 0) return 0;

    int wordOccur = 0;
    if (_Dictionaries[ind].ContainsKey(word)) wordOccur = _Dictionaries[ind][word];

    return ((float)wordOccur + k) / ((float)TotalWords[ind] + k * (float)GlobalDictionary.Count);
}

/// <summary>Gets category index number</summary>
/// <param name="cat">Classification to look for</param>
public int GetCategoryIndex(string cat)
{
    return AllCategories.IndexOf(cat);
}

```