AN EFFICIENT APPROACH FOR PREDICTION OF SUBJECTIVE WELL BEING FROM SOCIAL NETWORKS

Dissertation submitted in fulfilment of the requirements for the Degree of

MASTER OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

By

KANU PRIYA

Registration number

11502075

Supervisor

MR. DEEPAK KUMAR

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab (India)

May,2017

TOPIC APPROVAL PERFORMA

School of Computer Science and Engineering

Program : P172::M.Tech. (Computer Science and Engineering) [Full Time]

COURSE CODE : CSE546 **REGULAR/BACKLOG :** Regular **GROUP NUMBER :** CSERGD0291

Supervisor Name : Deepak Kumar **UID :** 17647 **Designation :** Assistant Professor

Qualification : _____ **Research Experience :** _____

SR.NO.	NAME OF STUDENT	REGISTRATION NO	BATCH	SECTION	CONTACT NUMBER
1	Kanu Priya	11502075	2015	K1519	09464618195

SPECIALIZATION AREA : System Architecture and Design **Supervisor Signature:** _____

PROPOSED TOPIC : An efficient approach for prediction of Subjective Well Being from social networks

Qualitative Assessment of Proposed Topic by PAC		
Sr.No.	**Parameter**	**Rating (out of 10)**
1	Project Novelty: Potential of the project to create new knowledge	6.60
2	Project Feasibility: Project can be timely carried out in-house with low-cost and available resources in the University by the students.	6.60
3	Project Academic Inputs: Project topic is relevant and makes extensive use of academic inputs in UG program and serves as a culminating effort for core study area of the degree program.	6.80
4	Project Supervision: Project supervisor's is technically competent to guide students, resolve any issues, and impart necessary skills.	6.60
5	Social Applicability: Project work intends to solve a practical problem.	6.80
6	Future Scope: Project has potential to become basis of future research work, publication or patent.	6.60

PAC Committee Members		
PAC Member 1 Name: Gaurav Pushkarna	UID: 11057	Recommended (Y/N): Yes
PAC Member 2 Name: Mandeep Singh	UID: 13742	Recommended (Y/N): Yes
PAC Member 3 Name: Er.Dalwinder Singh	UID: 11265	Recommended (Y/N): Yes
PAC Member 4 Name: Balraj Singh	UID: 13075	Recommended (Y/N): Yes
PAC Member 5 Name: Harwant Singh Arri	UID: 12975	Recommended (Y/N): Yes
PAC Member 6 Name: Tejinder Thind	UID: 15312	Recommended (Y/N): NA
DAA Nominee Name: Kanwar Preet Singh	UID: 15367	Recommended (Y/N): NA

Final Topic Approved by PAC: **An efficient approach for prediction of Subjective Well Being from social networks**

Overall Remarks: Approved

PAC CHAIRPERSON Name: 11011::Dr. Rajeev Sobti **Approval Date:** 26 Oct 2016

4/25/2017 4:31:46 PM

ABSTRACT

In today's world, each individuals are now addicted to the online social media. Social media sites have been used by millions of people, individuals globally. Every individual expresses his thoughts, daily life events and opinions on social media. In social media, individuals expressions are mostly in the form of non photographic images, text form, growing importance of linguistic text analysis etc. This paper provide an overview of the growing literature on subjective well being which is also known as Happiness. We use different types of approaches, one approach is that to measure wellbeing is to use of objective indicator which is also known as Gross Domestic Product(GDP).In this approach we classify the content of images and correlated with socio-economic indicator. The another approach is to using subjective measure by asking people to report on their happiness and life satisfaction. This paper represent the happiness indicator and create some limitation as measure of wellbeing like social media is not limited to text. In this research paper can be using the image classification with SVM and c-mean algorithm for image classification. Unlike text, the analysis of images does not require specific linguistic database.

ACKNOWLEDGEMENT

I would like to present my deep gratitude to concerned people who helped me out to learn this technology. I would like to thanks my mentor **Mr. Deepak Kumar,** during this project who helped me as instructor on each and every step from the day one during the period of Dissertation. He also helped me in searching and downloading the research papers of good journals related to my proposed plan so I can do my work qualitatively. His suggestions always help to do my work effectively.

I am thankful for his active support, valuable time and advice, whole-hearted guidance, sincere cooperation and pains-taking involvement during the study and in completing the Dissertation within the stipulated time.

I thank to all my friends, who are always ready to help me. All my friends encouraged me a lot.

I thank the marvelous and excellent work atmosphere at **Lovely Professional University, Phagwara, Punjab.**

Kanu Priya

11502075

DECLARATION STATEMENT

I hereby declare that the research work reported in the dissertation entitled "**AN EFFICIENT APPROACH FOR PREDICTION OF SUBJECTIVE WELL BEING FROM SOCIAL NETWORKS** " in partial fulfillment of the requirement for the award of Degree for Master of Technology in Computer Science and Engineering at Lovely Professional University, Phagwara, Punjab is an authentic work carried out under supervision of my research supervisor Mr. Deepak Kumar. I have not submitted this work elsewhere for any degree or diploma.

I understand that the work presented herewith is in direct compliance with Lovely Professional University's Policy on plagiarism, intellectual property rights, and highest standards of moral and ethical conduct. Therefore, to the best of my knowledge, the content of this dissertation represents authentic and honest research effort conducted, in its entirety, by me. I am fully responsible for the contents of my dissertation work.

Signature of Candidate

Name: Kanu Priya

Reg.No 11502075

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the M.Tech Dissertation entitled "**AN EFFICIENT APPROACH FOR PREDICTION OF SUBJECTIVE WELL BEING FROM SOCIAL NETWORKS"**, submitted by **Kanu Priya** at **Lovely Professional University, Phagwara, India** is a bonafide record of his / her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree.

Signature of Supervisor
Name: Deepak Kumar
Date:

Counter Signed by:

1) **Concerned HOD:**
   HoD's Signature: _____
   HoD Name: _____
   Date: _____

2) **Neutral Examiners:**
   External Examiner
   Signature: _____
   Name: _____
   Affiliation: _____
   Date: _____
   Internal Examiner
   Signature: _____
   Name: _____
   Date: _____

TABLE OF CONTENTS

CONTENTS **PAGENO.**

Inner First Page i

PAC Approval ii

Abstract iii

Acknowledgement iv

Declaration Statement v

Supervisor's Certificate vi

Table of Contents vii

List of Figures ix

CHAPTER 1:Introduction……1-18

1.1 Introduction to Data Mining 1

1.1.1 Applications of Data Mining 2

1.1.2 Image classification 2

1.1.3 Approaches of Image Classification 4

1.2Sentiment Analysis of twitter data 6

1.2.1 Characteristics of the Data 8

1.2.2 System Architecture 9

1.2.3 Social Media Analytics for Real time 12

1.3 Sentimental Analysis of Images 13

1.3.1 Framework 15

CHAPTER 2: Literature Review…….....20-27

CHAPTER 3: Present Work…….28-32

3.1 Scope of the Study 28

3.2 Objective 29

3.3 Research Methodology 30

3.4 Proposed Flowchart 32

CHAPTER 4: Results and Analysis…….33-44

4.1 Introduction to MATLAB 33

4.2 Results 34

CHAPTER 5: Summary and conclusions…….45-46

REFERENCES……..……47-49

APPENDIX……..……..50

LIST OF FIGURES

FIG NO.	FIGURE DESCRIPTION	PAGE NO.
Fig 1:	KDD process	3
Fig 2:	Well Being Classification	13
Fig 3:	Framework of Positive and Negative images being processed	16
Fig 4:	Proposed Flowchart	32
Fig 5:	Default Interface	34
Fig 6:	First Folder Images	35
Fig 7:	Second Folder Images	35
Fig 8:	Input First Image	36
Fig 9:	Input Second Image	36
Fig 10:	Clustering of Image	37
Fig 11:	Result Analysis	37
Fig 12:	Clustering of image	38
Fig 13:	Histogram of First Image	38
Fig 14:	First Image Classification	39
Fig 15:	Histogram of second image	39
Fig 16:	Second Image Classification	40
Fig 17:	Final Output Clustering	40
Fig 18:	PSNR Comparison	41
Fig 19:	MSE Comparison	41
Fig 20:	Accuracy Comparison	42
Fig 21:	Execution Time	42
Fig 22:	Pixel Range & Intensity value	43
Fig 23:	Base Paper Accuracy	43
Fig 24:	Final output image with accuracy	44

CHAPTER 1

INTRODUCTION

1.1 Introduction to Data Mining

Data mining may be defined as to find out the hidden patterns from the huge amount of data. Various companies and organization that have been focused and retrieved the information and knowledge from the data warehouse according to our requirement. The amount of data can be collected and refined by almost every company and business in today's world. Different types of databases can done when the high performance client and server computers are implemented. The growth of different economical and commercial databases at very high rates. Business data of each new step being evolved to built on the basis of the previous step which may be further has helped in converting the business data into business information.[22]

The mostly used business model for the purpose of navigate data warehouse is called the On-line Analytical Processing. OLAP is market oriented and is used for knowledge workers which includes manages, executives, analyzers etc. This system manage the large amount of historic data, and also provide the summarization at different level of granularity. The different techniques in data mining are:

1.Decision tree: A decision tree is a flowchart like tree structure or architecture where each internal node denote a teat on an attribute. Each branch represent an outcome of the test, the top most node is a tree called root node. A path is traced from the root to a leaf node, which hold the class prediction. Decision tree can handle the multidimensional data. The main advantage of decision tree is easy to understand by humans.

2.Mining frequent patterns: Frequent patterns are those patterns that appear frequently in a data set. For example, a set of items such as milk and bread that appear frequently together in a transaction data set is a frequent item set.

1.1.1 Applications

Data mining may be used and utilize into different types of company and business. It can manage the large amount of data also find out the historic data for the prediction of future trends. The various application applied in data mining can be briefly illustrate in the following:

- Data mining is used in the banking and finance. It may be used to identify the customer loyalty by analyses of data of customer purchase and buy activities such as frequency of purchase in a specific period of time, total monetary value. It can also used to retain the credit cards customers. By analyze the past data it can help to change their credit card applications. The hidden correlations can be identified using the data mining.
- Data mining also help to enables to characterize patients activities to see incoming visit and helps to identify the patterns of successful medical therapies for different illness.
- In business point of view, it provide the historical, current and predictive views of business operations for example, it include reporting, online analytical processing, business performance and many more.
- A web search engine is a specialized computer server that search information on the web. It can provide the web pages, images, and various different types of files. Web search engine are essentially very large data mining application. They have to handle the large amount of data and also deal with the online data.

1.1.2 Image Classification

There are rapid growth of object classification due to increase of high capacity PC, high quality and low priced camera. There are different types of steps which are involved in image classification such as sensors, image preprocessing, ETL(extraction transform loading), data integration and many more. we can classify the data based on the training set and the values in a classifying attributes and used to classify the new data. Classes are predefined. There are further two broad classes of image classification are:

- Unsupervised classification: The unsupervised classifications in which the outcomes greatly independent upon the software analysis of an image in which user given the sample class is known as unsupervised learning. The number of pixels which are related to each other are combined with the help of different techniques. It can be done by the observation and experience. In Unsupervised classification, when the pixels have the some common properties are combined, the user must be notified this values.
- Supervised Classification: This method provide to select the those samples image of the pixel which represent the certain classes. These images of sample can give the reference of image processing software to represent the training sites. On the basis of the user knowledge, the different training sets are selected, image can be classified is also given by the user only.

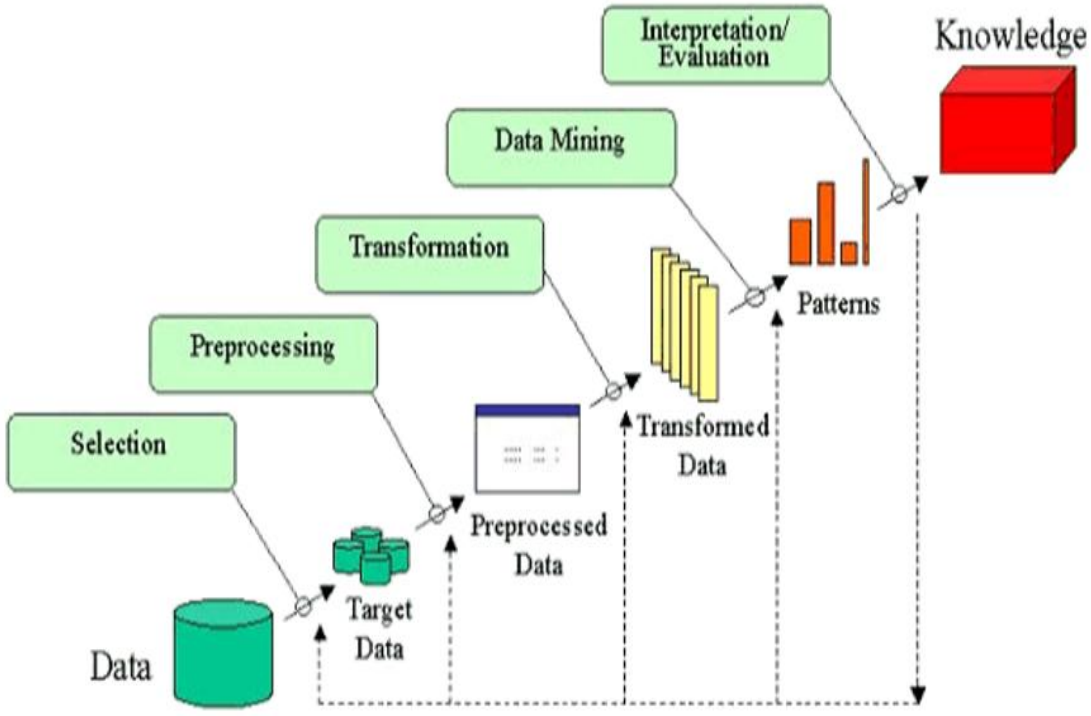

Fig 1: KDD Process[22]

1.1.3 Approaches of Image Classification

1) Characteristic Used on the basis of shape & motion:

a. On the basis Shape:- In the shape based we use 2D spatial information can be utilized from an object to conclude the characterstics. It include the centroid points of different types of primitive geometric shapes such as rectangle, the skeleton, circle, traingle etc. which can be utilized the features of this method.
b. On the basis of Motion:- In the motion based we can do the classification of these types of methods including the temporal tracked objects can be utilized from this features.

2) Used of Training Sample by supervised and unsupervised classifications:

a. Supervised Classification: In this technique we can sample the information classes which can be known and that can be utilized to classify the number of pixels that does not have any identity. Mostly it provide the different examples that are means algorithm to find the minimum distance parallelepiped algorithm & algorithm of maximum likelihood .
b. Unsupervised Classification: In this technique we can analyze the unknown pixels and their centered division the number of pixels of classes on the basis of image values is known as the unsupervised classification. The Example of unsupervised learning is the K-mean clustering algorithm.

3) Assumption on the basis of Parameter on Data with parametric and non parametric classifier:

a. Assumption of parametric classifier:- During assumption of parametric values which include mean vector & co- variance matrix that can be utilized with various parametric prediction of Gaussian distribution that can help the generation of parameter values of training samples such as for example linear discriminate analysis.

b. Assumptions of non parametric classifier:-It includes the related data with no prediction that can avoid the utilization of statistical parameters to calculating the class separation. examples of non parametric classifier is neural network, SVM, decision tree classifier.

4) Used of Pixel Information as per pixel and sub-pixel classifiers :

a. Pixel classifier:- In pixel classifier, we can generate the signature with the combination of spectra and with the help of conventional classifier. These type of combination that can provide the training set pixels. At the end the accuracy is not appropriate according to the requirement of user. For example SVM etc.

b. Sub-pixel classifiers:- It may include the linear or non-linear combination that can be predicted for the each pixel. It contain the mixed pixel issue of sub-pixel classifier with various spatial resolution images. The approach includes are fuzzy-set classifiers.

Introduction to Clustering:- clustering is on the basis of similarity and dissimilarity or on the basis of distance .In clustering, classes are not predefined. The process of grouping a set of physical or abstract objects into class of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another and are dissimilar to another objects. Clustering is an example of unsupervised learning that is in the form of learning by observation. Cluster analysis can widely used in many applications such as business intelligence, image pattern recognition, web search etc.

Measuring Well Being: We can measure the well being in two different approaches that is objective measure and subjective measure. In objective measure, people wellbeing is accessed by using some certain facts such as economic, social and environmental statistics. On the other hand subjective measure of well being by using the people feelings or life satisfaction. This approach which aims to capture the well being through the measurement of brain waves.

1.2 Sentiment Analysis of Twitter data

Sentimental Analysis likewise called opinion mining, is the field of study that analysis people, individuals appraisal, evaluations, assessments, evaluations, states of mind, emotions and feelings towards entities, for example, items, services, product administrations, event, people, issues, occasions, points, and their qualities. The basic task in estimation of sentimental analysis examination is classify the extremity of a given content at the report, document, sentence, or features/perspective level—whether the communicated conclusion in an archive, a sentence or a entity feature/aspects include is sure positive, negative, or neural. Progressed, "beyond extremity" assumption sentimental classifications looks, for example, at emotional states, for example, "sad", "happy", and "angry". Whenever you use the choice, we have to need to know another sentiments. In today reality, business, organizations and different associations dependably want to discover customer/general sentiments about their items, choices and privileges administrations.

Currently at this stage where number and different kinds of human needs and requirements, information can be found in services online. To Add the developing demand and choice, a significant part of the government, services, military trades, medical guide and private precisely information that should be accessible online. Because of that rapid growth of information and services on the internet, attacks have increased as of late. These types of groups are viewed as a danger, that can be committed to disable the information systems that should motivate mostly by politician or web social issues. The information of web services that can be added to changed the way online users communicate with another. A few web and internet services that permit individuals to interact or interfere to each other continuously. Most commonly type of communication is the purported social systems, that have based on a small-scale blog organize & permit clear content passages, opinions, loading pictures and files, and interactive user-to-user communication through different types of chat messages. A significant part of the information in social websites demonstrates the opinions and views of many users with respect to a different issues. Many users feelings, opinions or sentiments, in social websites is valuable for forecasting and monitor the different types of occasions, for

example, many market tendency which can be identified, different types of political views. The benefit of working with sentiment analysis in social networking is that allows for fast and efficient processing of information.

Not at all like traditional methods for forecasting macro-economic prediction of future or present stock investment is performed merging the result of sentimental order of tweets and stock investment or profit. The sentimental analysis may also called opinion mining which is extracting the different types of information from a textual input. With a specific end goal we merged the different types of strategies from natural language processing, opinion mining and textual analysis. Mining permit determining whether given textual input is objective & subjective. Subjective well being that gives the input is grouped together in either positive or negative analysis. With a specific end goal to perform analysis characterization a model must be describe and constructed by providing training and test different datasets .

The method that can perform datasets opinion and sentiments detection of messages automatically. The large amount of information to be characterized by the textual, the process Naive Bayes strategy may have to choose because it can give the efficient result with huge volumes of information and the way that is it is incremental or decrement. The large, complex and huge volumes of information is the choice to apply a Naive Bayes strategy calculation. Opinion sentiments analysis that covers of words a model and values must be constructed and test datasets for combining different groups of information. The datasets can have two unique types of methods. The first one is to apply an automatic detection of text and messages.

Some other different strategy was a giving naming of sentimental tweets. Every conversation is pointed to as positive(+ve), negative(-ve) or neutral types. There are mostly two different datasets. At the starting stage containing of 500 hundred tweets of text and messages. Secondly, dataset consist of 2.4 million chat messages. The 80% of dataset that can utilized which may have training data-set the other 20% was utilized for the testing. At the end 2 classifiers can utilize manually dataset. The different dataset can utilize the result for stock market prediction: 'supposition term' (which can be denoted as ε) developed - it is a logarithm at base 10, ratio of +ve to -ve tweets .

$$\varepsilon = \log_{10} \frac{\text{num} - \text{of} - \text{tweets(positive values)}}{\text{num} - \text{of} - \text{tweets(negative values)}} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots . \text{(i)}$$

If ε is positive number then it is normal that a stock price will be increased and if it is negative number ε it indicates price decreases. The mail goal of supposition term that can estimate price of stock, characterized results are merged with a linear regression of past prices which are imaginary values.

Challenges of Sentimental analysis: The most impact of notions that can be sentiment words, additionally called the sentiment words. The words that are regularly used to indicate the positive or negative types of assessments. For instance, great, brilliant, good, amazing and wonderful are certain slant words, and awful, bad, poor, and unpleasant are negative sentiment words.

Assessment opinion spamming or sentimental analysis has turned into a big issue. Aside from people who give fake suppositions in audits and gathering exchanges, there are additionally business organizations that are in the matter of composing fake surveys, fake review and false online journals for their customers. A few different instances of fake reviews have been accounted for in the news. It is important to recognize or detect such spamming exercises to guarantee that the feelings on the Web are a trusted source of valuable data.

1.2.1 Characteristics of the Data

The social networking and micro-blogging administration of twitter that provide the users to update the original time messages, which referred to as tweets. Different chats and tweets may contain huge unique characteristics value, that may come to face new difficulties and the method for resulting the investigation of various areas of various types. Taking after are some key attributes of tweets:

- Length of the message or chat : Twitter message is having maximum 130 characters that can be unique in relation to previous or present sentiment arrangement research which focused on classified the longer texts, for example, product reviews and reviews of movies.

- Techniques of writing the slang spellings: Many in-correct words & spelling and cyber different slang in tweets is all the more regularly of differentiate with different domains types of information. These type of messages may have give fast, efficient and small in size, individuals and users utilize acronyms words, mis-spelling, and utilize different emoticons messages and different character that consist of some characterize meanings.
- Measure of Availability: Measure of information which should be available in different manners. In public domain mostly the user tweet as compared to face book or twitter (that have huge amount of privacy details) making information with high availability. The consideration of tweets for training may given to twitter API facility.
- Subjects/topic: The users who having using Twitter, they can post messages of a different types of topics that may not designed for a specific type of topic. They give the difference of an different types of fraction of previous research, that specific domain can focus for example, reviews of the movies.
- Real time intervals: Many tweet and blogs that are update at greater intervals of time with blogs may consist of greater and write them investment profit and benefit. Tweets are limited capacity to 130 letters that have updated all the time. that can give the actual feelings of time and represents events reactions.

There are mainly different definitions about the twitter:

- The Emoticons: It can give the pictorial representations of facial expressions of different types of emotions that can be expressed by the individual and user moods.
- Target and Hash Tags: The users of twitter may use '@'which define to tag a different users which are present on your profile. Users also uses the '#' symbol that used to topic stamps.

1.2.2 System Description Architecture:

The users use Sentiment/opinions words to express their opinion (negative, neutral and positive). The semantic occurrence that we have to locate of the sentiment character in

tweets, the hybrid approach is user for both structure based & dictionary based techniques is considered. the different features like emoticons and hash tags are added to considered the part of cyber language.

In the system description architecture, opinion/sentimental direction in which to extraction of the opinion words of the different tweets of different things and after that discover their orientation method that have to choose either opinion word that can reflects the , negative(-ve) sentiment, positive(+ve) sentiment and neutral sentiment. The combination of the adjectives, verbs and adverbs are in the opinion/sentiment words. The structure based approach may have use to locate the adjectives and dictionary based approach that have employed to locate or organize. The general tweets opinion/sentiment is evaluated with a linear equation.

There are following different approaches of the proposed system :

a. Pre-processing of Tweets

The file that that can transaction may contains opinion indicators, specifically adjective, verb and adverb with many emotions. similarly some emotion are efficient, specific to some part, the tweet in caps, length of repeated sequences of exclamation marks that can be identified in the form of percentage. The method pre-processing of the data and tweets as follows:

A) Eliminate all URLs that is uniform resource locater (e.g. www.punjabi.com), hash-tags (e.g. #structure), targets (@name person), and special Twitter words ("e.g. EG").

B) Then we have to calculate the tweet with caps percentage.

C) Spelling Correction: The linear sequence of repeatedly characters is tag by a weight. The difference between the regular/continuous usage and automatic used by the word.

D) The emotions should be replaced with all the emotions with sentiment .

E) Eliminate the different punctuations which count number of the exclamation marks.

F) The NL Processor linguistic , to tag we have require adjectives, adverbs and verbs.

b. Module Scoring

The different steps is to locate and utilize the semantic module score of the opinion analysis bearers that are the verbs, adjectives and adverbs. We have utilize different approaches to locate the semantic orientation of adjectives and the dictionary based approaches to locate and utilize the semantic words orientation of verbs and adverbs .

- **Adjectives with Semantic score:** The adjective is to describe the word which may utilize of an object. Adjectives with semantic orientation that have a tendency to be in a domain specific features which we to utilize the semantic module score of the opinion analysis bearers that are the verbs, adjectives and adverbs specific domain.
- **Adverbs & Verbs with semantic score:** Further, we may compute the sentiment of specific texts based on the semantic orientation of the adjectives which includes adverbs. The primary goal of semantic score of verbs and adverbs that there are many adverbs in linguistics, (for e.g, 'no') that consist of the adjective which have to utilize a negative & positive orientation.

c. Scoring of tweet sentiments:

The verbs, adverbs and adjectives, we group the alternatives adverb & adjective together and we called as a adjective group. Similarly this method groups the corresponding verb and adverb merged and call it to a verb group. This adjective group strength is then calculate by the product/structure of adjective score and adverb score and the verb group strength of the product verb score and adverb score may utilize. In many conditions, no adverb may found in the opinion words, S (adv) is set as a default value 0.5.

There are various following steps which includes :

- The basic method is to allow unsupervised learning sentiment analysis for the social sites pictures and texts.
- The framework which may be created by unsupervised sentiment analysis for social media text and pictures that can captures visual and textual

information. This is the appropriate method of unsupervised sentiment analysis framework which contains social sites images.

- Different studies and survey using these datasets from real-world social media image-sharing sites may have to reporting performance, monitoring and leading existing methods which can serve as a benchmark for further exploration and evaluations.

1.2.3 Real Time Social Media Analytics

Real-Time analytics denotes the capacity to use all available data and resources when needed. The analysis of data is carried out dynamically and reports are generated with no delay. Mostly real-time analytics is used for geographic location and tracking purposes. Nowadays, people instantly share on social media about situations like natural disasters, hence the real-time analytics of social media may provide life-saving information. Well-being is the 'Good Life' that everybody wants to pursue and is influenced by the basic daily life objective factors like income, job, health.

These factors can be measured directly with the quantitative factors like GDP(Gross Development Product), HDI (Human Development Index) and GPI(Genuine Progress Indicator). However, these factors cannot measure the Quality of Life (QoL) of individuals. The indicator of measuring QoL is the Qualitative Well-Being or Subjective Well-Being. (SWB). SWB can be expressed as how the people experience their QoL, which includes judging life positively and feeling good. The main objective of this technique is to find the hidden patterns of huge amount of data by the various present data mining techniques such as classification, clustering, regression etc. The direct and indirect measurement techniques of well-being are given in the Figure 1.

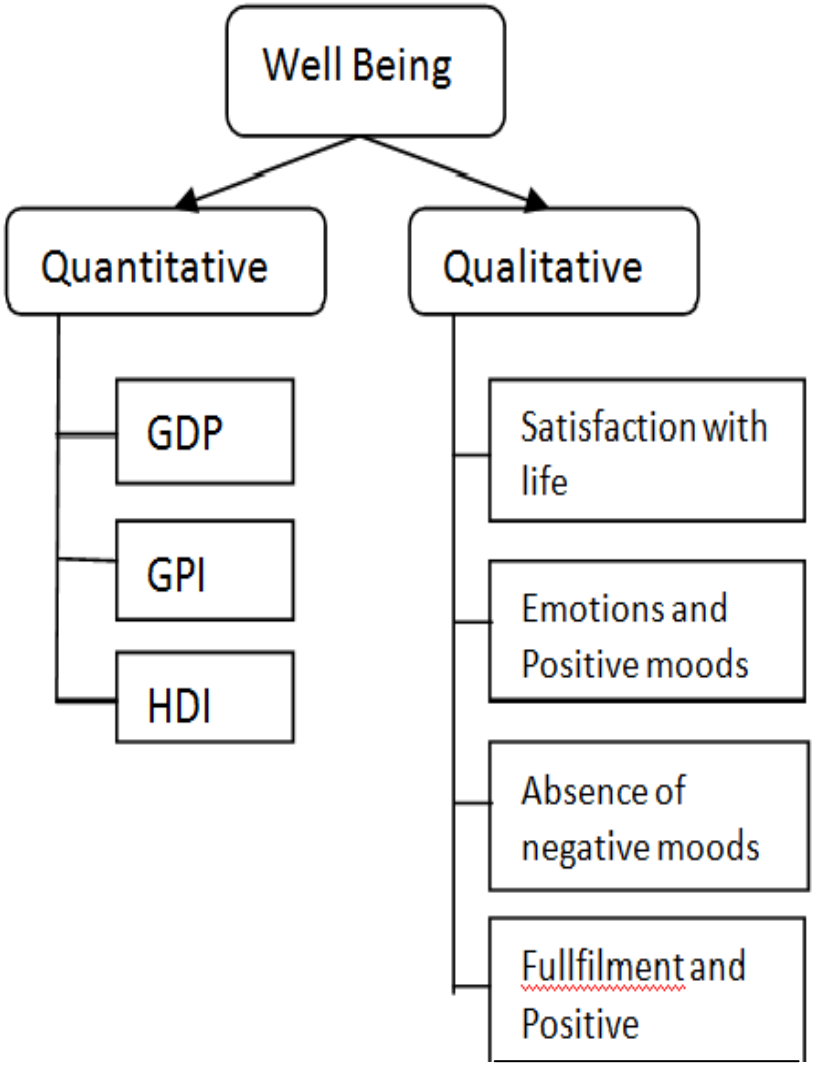

Fig 2: Well Being Classification

1.3 Sentimental analysis of Images

In the opinion mining, analysis of textual information areas which may developed, user decision making process, monitor the performance of brand lifestyles, prediction of stock markets, voting forecasts of political, and intelligence gathering. Similarly, the visual

information that can be analyzed in different areas for example, reviewing of the feeling, image information retrieves and the progress is moderately behind. In the research work, the demonstration of the images lies around 35 % that can be updated and share links on twitter which makes to a interesting and active area to investigate features that can create visual data mining process. A picture is justified which consist of many hundreds and thousand words. Most likely the text mining approach various research has done with respect to feel and emotions in pictures and images, has greatly depend upon been critical yet challenging. The awareness of social networks sites, images turn into a convenient way for information diffusion among different users online. To Plan a lead visual content-based sentiment analysis, we have expressed current approaches which incorporate employing low-level features which means of facial expression detection user intent resources and understanding images utilizing attribute learning types. The various Sentiment and opinion analysis method which contain low-level characteristics have the limitation that provide its undesirable for abnormal state utilize.

The various resources of data about data also known as meta data which contain the information of images in different form of learning techniques. All the data but not images which contain of sort data and researchers are attempting techniques and features. It becomes very difficult to predict the human behavior as there are many factors like genetic, socio-economic, etc. Since humans are very comfortable with each other and try to live a social life. Being social is the common characteristics and it can be useful to predict the human behavior with respect to certain conditions. The impact of the internet on the social life of humans can be seen well. Social media forcing the people to use it and has become part of everyday life.

To study and investigate the evidence of the most important factors, trust, and risk which destroy the user emotions behavior corresponds to social media platforms. The finding suggested that both risk and trust had significant measure, but trust had a stronger effect. The various effects of risk and trust have been clearly visible on the social media. Trust is closely related to the happiness of the human behavior and mostly happier persons are more trustworthy. Human living style, languages and behavior changes after a few hundred kilometers. It becomes very difficult to predict the human behavior as there are

many factors like genetic, socio-economic, etc. Since humans are very comfortable with each other and try to live a social life. Human living style, languages and behavior changes after a few hundred kilometers. Being social is the common characteristics and it can be useful to predict the human behavior with respect to certain conditions. The impact of the internet on the social life of humans can be seen well. Social media forcing the people to use it and has become part of everyday life.

1.3.1 Framework

a. Dataset: The method principally which contains 3 different types, the first type that can developed medium level attributes values, the classifier which is trained utilizing SUN Database and the second main attribute of database is to designed for abnormal state scene understanding and fine-grained scene recognition. The database includes more than 850 category & the 14,341 images which can labeled regarding the user studies of different attributes, the various Attributes labels are presented as zero to three votes which contain zero vote denoted as this picture is the min correlated attribute & three votes means it is the most correlated, because of this voting mechanism, the set of images that can be labeled as positive:-that pictures which have more than one vote, also known as as soft decision (SD) may classified ,the images which contain more than two votes, that can introduced have hard decision (HD), Soft decisions are less inclined to be over fitting over the mid level attributes and less accurate than the classifiers learned based on hard decisions and benefits.

The next step of algorithm is to predict the sentiment classifiers with images crawled from Twitter together that have 800 images with their textual data and information. The most well known micro blog platforms in the twitter currently which consist of sentiment truth is held from visual sentiment analysis with authorization of creators datasets . Different dataset may contain 1,341 positive, 222 negative and 542 neutral picture tweets. When we have to test the task, choose randomly select 811 images which containing positive (660 picture tweets) and negative (150 picture tweets) . Mostly the common and important final step is to detection of human facial emotion mechanism,

this method chose to utilize the direct face emotions dataset for the most part on the grounds that faces are all around aligned together and have reliable lighting, that makes its makes generating great Eigen confront much less demanding. At present study, the dataset which contains 70 girls & boys more than 2 days expressing 7 emotions (frightened, anger, sad, happy, negative, neutral, disgust, and surprised types) in 5 distinct modes like (front, left parole, right parole, left angle, right angle).

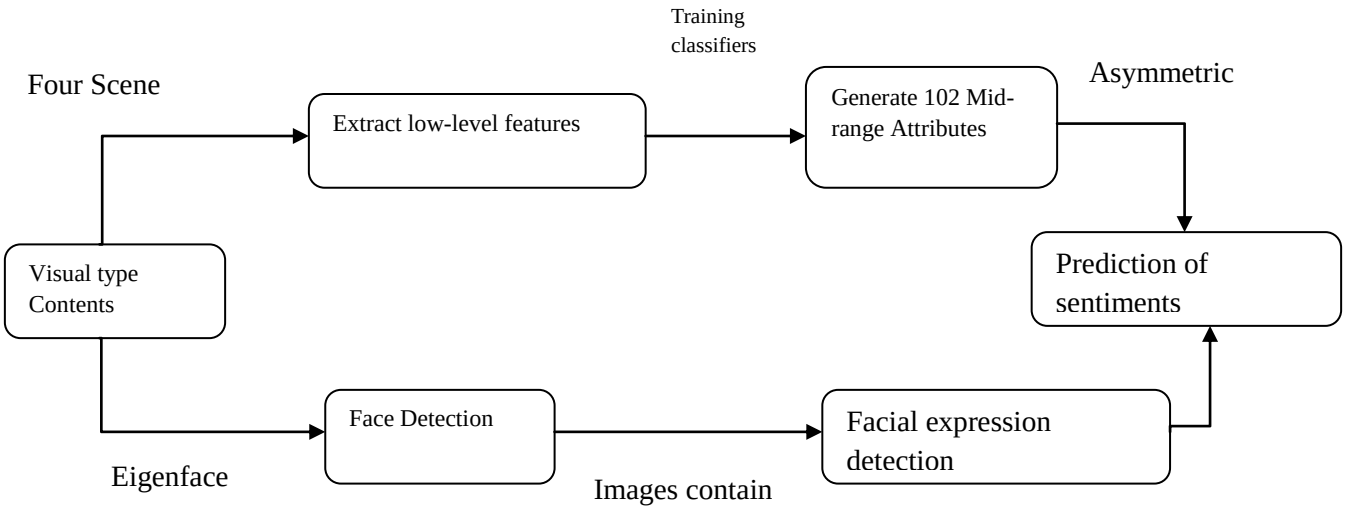

Fig 3: Framework of positive and negative images being processed

b. Selection features:- These part have to plans to select mid-level attribute of low level features & general scene descriptors which are GIST descriptor. The HOG 2x2 descriptor, self-similarity descriptor and a geometric context shading histogram features. There are many features that were chosen where each individually powerful and in light of the fact that they can depict distinct visual phenomena in a scene perspective other than utilizing specific object classifier. Many different types of features of scene descriptive that suffer neither from the inconsistent performance compared to commonly utilized object detectors for abnormal state semantics analysis of a picture that can be based to generate low level features.

c. Mid level Attribute generating:- When to choose a low level features they are training to our mid level attribute classifiers based on sun database contents that have to

15,341 pictures are training data sets & low level characteristics of every picture mean more than 171,000 dimensions and the visualization reflects that the intrinsic complexity and obscure characteristics of web data, sometimes, make difficult to establish relationships among certain attributes.

d. Classification of image sentiments:- There are many factors which influence the human behavior such as genetic, socio-economic, physical environment and psychological factors. The socio-economic and psychological factors play a vital role in social behavior. The factors like education, family, culture, self-concept, fear, anxiety, etc. directly associated with the social life. There are certain theories related to human behavior named as Bandura. Our dataset that have contain 650 positive samples and the different 150 negative samples that is to probably going to acquire classifier based on these samples. We can realize that various asymmetric structured data that can manage one sided dataset resources. The common facility and idea of asymmetric packing s that to contain building of one and several other classifiers are fabricated and trained with similar types of negative(-ve) samples merged together with various samples of positive(+ve) samples of a similar amount of resources of data.

e. Recognition of facial sentiments: There are many steps which consist of last decision mechanism by adjusting the Eigen based detection method of emotions approach. The various types of pictures which contain different faces facility with the awesome partition of different entirely images that is 372 pictures that contain dataset with different faces. The face emotion technique elimination and detection is not important that can give average overall performance of our algorithm method.

f. Visual Sentimental Analysis for Images: This analysis type of architecture which may can prediction of various sentiment that can be breaking down image data of different information, At these days individuals are uploading the different types of images in social networks in millions amount. The example include that Twitter, Facebook, you tube, and instagram. These images have some data in show the emotions of individuals in online social networks. Many sentiment analysis which contain the images having consideration in the field of online multimedia big data research. Many

researchers may focus on examining and investigating the sentiment of the textual contents [2].

In this research paper we create model to prediction of sentiment of visual information. Visual Sentimental Analysis architecture may have to utilize to predict the visual sentiments by using the transfer learning approach. Various parameters which have learned that can considerable various convolution neural networks to prevent cure of over lifting of data from various data models. The Twitter image dataset and study various types of model to guarantee that the performance which should done prefer over the present state of the art. We can perform for monitor the emotional states of persons which are suffering from mental cause disorder. Various researchers might want to apply our model for dissecting sentiment from visual information. At end result, researchers need to scattered supervised and unsupervised approaches for the visual contents in different types of framework learning.

g. The Realistic Analysis of Data Warehousing and Data Mining Application in Education Domain: To construct data warehouse and to extract the important contents and mining open source tools. This analysis than can be used business intelligence for an educational campus, realistic data may be used in this type of analysis for the use of experiments. In this research paper we have to find the hidden patterns from the huge amount of data for different educational campus which can be made the decision making.

The different types of approaches may be used in data mining and data warehouse which may include the use of clustering and classification having large number of information. To improve decision making process which can be implemented in different form ton the basis of classification & clustering data. The individual may difficult to access the information from the different location which can be removed.

The technique of Data mining on the basis of education campuses may involve various categories such as Manager, workers, Students and Infrastructure. The worker database which consist of information of individual users information on the basis of career growth. To provide the data of different student's information regarding his/her academic and non academic involvement. These features are designed based on

psychological study and experiments. The greater range of SWB which may suggest that the many people to share the content on social sites like text messages, images and different types of emotions over the insta-gram and twitter with greater percentage. We have to study regarding all the approaches which contain the decision can be made with different types of information for the final result.

CHAPTER 2

LITERATURE REVIEW

In literature review, the research papers that I have studied are in the following that have the foundation to this study:

Predict Social Trends on Twitter from Non-photographic Images[23]

Humanists may have to utilize the history types of pictures as which contain the information related to the behavior, opinion, fashion, lifestyles and another various different details of particular tradition, periods and places such as such visuals demonstration the objects of interests of data, various examples includes social scenes of data, various city places and people lifestyles of various tradition. Many years may have to gain that researchers shown the various characteristics of the large scales of social media are correlated with number of different socio economic indicators characteristics which may be prediction of the huge volume of different social patterns and media. Various cases incorporate influenza patterns, achievement of movies, and measures of social prosperity of populaces. Interestingly, this strategy concentrates on images [23]. We have to research the character extracted the information from Tweeted non photographic kind of images can predict a number of socio economic indicator characteristics. Various content of these images is classified utilizing the cutting edge Neural Network convolution benefit of goggle net classifier and further the incremental category is chosen that is called "images of writings reviews" that are typically non-photographic images which contain the screen shots of websites and the various message data conversations. There are mainly two features of image recognition of non photographic images which are: to sharing rate every year per city or the sharing rate every hour over a 24-hour time commonly shared more than 1 year in each and every commonly city. These features we have to find that they are related with social self reported which give idea from Gallup type studies, about the median housing prices, salary, and education field of ranges and to final result that gives commonly types of

social media non photographic pictures which may be utilized to predict social characteristics of image detection.

Visual Sentiment Analysis for Social Images Using Transfer Learning Approach[12]

This analysis type of architecture which may can prediction of various sentiment that can be breaking down image data of different information, At these days individuals are uploading the different types of images in social networks in millions amount. The example include that Twitter, Facebook, you tube, and instagram. These images have some data in show the emotions of individuals in online social networks. Many sentiment analysis which contain the images having consideration in the field of online multimedia big data research. Many researchers may focus on examining and investigating the sentiment of the textual contents [12]. In this research paper we create model to prediction of sentiment of visual information. Visual Sentimental Analysis architecture may have to utilize to predict the visual sentiments by using the transfer learning approach. Various parameters which have learned that can considerable various convolution neural networks to prevent cure of over lifting of data from various data models. The Twitter image dataset and study various types of model to guarantee that the performance which should done prefer over the present state of the art. We can perform for monitor the emotional states of persons which are suffering from mental cause disorder. Various researchers might want to apply our model for dissecting sentiment from visual information. At end result, researchers need to scattered supervised and unsupervised approaches for the visual contents in different types of framework learning.

Invention of Visual Features for Recognizing User's Sentiments in Social media Images[13]

Recently, with the increasing of users and activities in social network service, an image sentiment analysis has been an important keyword for psychological study and commercial marketing. To perceive accurately user's sentiments of the image, it is essential to identify discriminative visual features and after that conduct analysis based on observed features. In this paper, authors propose two hand-designed features: color composition and SIFT-based shape descriptor [13]. These features are designed based on

psychological study and experiments. Initial, two visual dictionaries are worked by Kobayashi's color image scale and Hierarchical clustering. Next, color compositions and SIFT-based descriptors are extracted from image. At that point, the set of extracted features are independently transformed into a histogram representation by calculating the occurrences of the respective feature appointed to each visual word in the dictionary. To confirm the effectiveness of the proposed features, we apply them to image sentiment analysis for predicting user's polarity and influences.

Realistic analysis of data warehousing and data mining application in education domain[24]

This paper or research work may define the initialization information regarding data mining and data warehousing that has to be illustrate. To construct data warehouse and to extract the important contents and mining open source tools. This analysis than can be explained the need of data warehouse & business intelligence for an educational campus, realistic data may be used in this type of analysis for the use of experiments. In this research paper we have to find the hidden patterns from the huge amount of data for different educational campus which can be made the decision making. [24]

The different types of approaches may be used in data mining and data warehouse which may include the use of clustering and classification having large number of information. To improve decision making process which can be implemented in different form ton the basis of classification & clustering data. The individual may difficult to access the information from the different location which can be removed.

The technique of Data mining on the basis of education campuses may involve various categories such as Manager, workers, Students and Infrastructure. The worker database which consist of information of individual users information on the basis of career growth. To provide the data of different student’s information regarding his/her academic and non academic involvement. These features are designed based on psychological study and experiments. We have to study regarding all the approaches which contain the decision can be made with different types of information for the final result.

Towards a Selfie Social Network with Automatically Generated Sentiment-Bearing Hashtags[14]

The primarily category which have focus on content which can be generated by individuals and utilizing the data which can be analyzed in some meaningful ways and methods. Text conversation, screen shot of images, Selfies, and another type of content which have to turned into an most common feature of the visual graphic communication in social networks. The different images that can be post on many social media sites are are selfies, text conversation, screenshot of images etc., which contain the sentiment type of selfies of the person in the image that can post in different social network sites such as twitter, facebook, instagram. The researcher find that to add quantitative sentiment indicators to these selfies can provides the important information on social network individuals over the different sites of networks[14]. The researcher may represent social networks which automatically captures the content of the sentiments which are in form of selfies that have posted over the social sites. These sentiments are based on the selfies of different types of users along with the timeline of the sites. The system or computer which may automatically generates sentiment hash-tags. Various selfie that can automatically generated by hash tags which can be post by individual. Many users may have enjoy to take selfies and to upload them on the social sites.

Computational Challenges for Sentiment Analysis in Life Sciences[15].

The different types of users may have the base of social networks with data/information and connectivity. Different types of status, images, friendship, mimic of tweets, hash tags and tag which can generates the large amount of structured & unstructured data. The meaningful information of the data may extracting the behavior and emotions of each individual users and user groups, which provide the goal of sentiment analysis efficient representation of the data.there may be infinite types of applications of Social networks including the both research and commercial purposes which contain the wide range of areas, including life science [15]. By and by, collecting, sharing, putting away and breaking down data of social sites may represent the different difficulties of computer researchers. The best example is that management of different unstructured data, big data, data mining and the basic requirement for real time data evaluation. Many

application of these methods which provide social sites of data with different types of new knowledge increased in different fields of area.

Sentiment analysis using latent correlations among textual, visual and sentiment analysis views[16]

The user of the internet may upload the images increases in rate may have to post pictures every day to express the sentiments which can be developed by various applications and the method of image sentimental analysis may focus on the visual features which consist of information related to images frequently explored. In this research paper we have to propose special method that can correlate among the multiple views of content, visual and textual views constructed about the sentiments which utilizing Sentimental wordnet. [16].In this research the researcher tells that we locate a embedding space in which correlations among the three views are maximized in datasets. Mostly the features may utilize that to train a sentiment classifier analysis, the reciprocal information from different views may be analyzed. In this approach, results of practical which may organized on Flickr and Instagram images which tells that this approach goal to accept sentiment classification which considered the accuracy over methods that utilization the state of the art method of various modules.

Applying Sentiment and Emotion Analysis on Brand Tweets for Digital Marketing[19]

As digital marketing is turning out to be more popular, the number of customer views on brands is expanding rapidly. This makes it harder for companies to survey their brand picture or digitally showcase their products on the web [19]. A vocabulary based approach is shown to extracting sentiment and emotion from tweets for digital marketing purposes. Ten thousand tweets related to ten technology brands: Apple, Google, Microsoft, Samsung, GE, IBM, Intel, Facebook, Oracle and HP are collected. A sentiment analysis is performed utilizing Senti-Word-Net while emotions are detected utilizing the NRC Hash-tag Emotion Lexicon. The scores obtained from the two lexicons are compared and combined into one result for every tweet. The execution process of this experiment is depicted and it is demonstrated that the accuracy of the combined approach of sentiment and emotion analysis is upgraded over the independent approaches of

sentiment analysis or emotion analysis. This paper proposes a dictionary based approach to combine sentiment and emotion analysis of tweets related to technology brands. The output of this approach gives a better understanding about the polarity and emotional states of customers. This, thus, helps companies get a more detailed perspective of their brand picture on Twitter. The executed experiment demonstrates an improved accuracy of the combined approach (52.6%) over the independent sentiment analysis.

Paragraph2 Vec-Based Sentiment Analysis on Social Media for Business in Thailand[20]

This paper proposes the sentiment analysis system in Thai language. It aims to use for the three business types (Retail, Banking and Telecommunication) to monitor their brand picture by means of social media. Pantip.com is the most popular online community in Thailand, which many customers posted the comments about their business [20]. Typically, three sentiments must be identified (positive, negative and neutral), yet four sentiments (positive, negative, neutral and need) are presented in our proposed system in light of the fact that the need sentiment can be utilized for creating new business openings. The unsupervised deep learning feature extraction for text, called Paragraph2Vec, paragraph vector or Doc2Vec, was connected in this paper, compared to the established TF-IDF. Various results which illustrate that our present work performs better method. As indicated by the experimental results, the Paragraph2Vec is extremely useful in text processing. It can generate the far lower dimensional feature space with embedding the request and semantics of words that makes the generated feature vector is more meaningful than TF and TFIDF in the standard framework of bag-of-words. This sentiment analysis utilizes Logistic regression as the classifier for the paragraph vectors. From the experimental results demonstrate that joining Paragraph2Vec with Logistic Regression obtains the best performances in accuracy with acceptable speed.

Sentiment Analysis for Social Media Images[21]

In this proposition, the difficulties about the type of understanding different human sentiments/opinions is examined from the huge amount of scale collection of Internet pictures based on both picture features and contextual social network information, (for example, friend comments and user description). In spite of the colossal walks in

dissecting user sentiment based on text information, the analysis of sentiment behind the picture content has largely been ignored [21]. In this manner, the significant advances in text-based sentiment prediction tasks are advanced to the higher-level challenge of predicting the underlying sentiments behind the images. It is demonstrated that neither visual features nor the textual features are independent from anyone else adequate for accurate sentiment labeling. In this way, a method for utilizing them two is given, and sentiment prediction problem is figured in two situations: supervised and unsupervised. An optimization calculation is built up for finding a local-optima solution under the proposed framework. When we have to compare the large scale datasets values that can consider the demonstrated that the proposed method improves significantly over existing state-of-the-craftsmanship methods. Later on, the method is going to fuse more information on the social network and explore sentiment on signed social network. The proposed framework often obtains better performance than baseline methods. There are two noteworthy reasons. To start with textual information gives semantic meanings and sentiment signals for images. Second the visual and textual information for sentiment analysis is combined.

Data Mining which can be used in different branches[22]

[22] The author Smita and Priti Sharma studied the data mining which can be used in different branches. They provided a deep knowledge of data mining concepts including the various applications and techniques. This study does a review of data mining in various sectors. The sectors include educational area, marketing field, fraud detection and telecommunication sector. They described data mining as predictive and descriptive in This model describes the various categorization into which data mining works. The main objective of this technique is to find the hidden patterns of huge amount of data by the various present data mining techniques such as classification, clustering, regression etc.

Predicting the future with social media[25]

In today's world social media becomes more popular and mostly people can used the social media sites including facebook, instagram, twitter and many more. Social media is the fast growing day by day therefore the need of data increasing widely. It can also be

used to collect the different types of information online with huge amount of data. The content that can be generated is increasing again and again. In this paper, it illustrate that how social media content can be used to predict real world outcomes. We can built a tweets that are create by a particular topic and then demonstrate extraction of twitter sentiments. In this paper it demonstrate about the movies for 2 different reasons. The first reason is that the social media use the user community that can be characterized by large number of users that can discuss about movies and ask for the opinions. The second reason is that outcomes of the real world that can be easily observed from box office revenue for movies.

CHAPTER 3

PRESENT WORK

3.1. Scope of the Study

This work is based to predict the user interest for uploading the images on the twitter. To predict the user interest, features the images which are uploading on the twitter. The features of the input images are broadly classified into date of uploading, image type and image features like color and texture. The technique of neural networks is been applied which learn from the previous experience and drive new values. The unsupervised learning technique is been applied which will learn the user behaviors and driver the patterns of the image.

In this work, googlet classifier is been applied which will classify the image and predict the behavior of the users. The googlet is the online classifier which predicts the images on the basis of text, aggregated shared rate per year.

3.2. Objectives of study

1. The first objectives is to analyzing and Study the various image prediction techniques and approaches for the twitter data.

2. The second objectives to propose improvement in existing GoogleNet classifier for image classification .

3. The third objective is to proposed improvement is based on wavelet transformation and SVM classifier .

4. The fourth objectives is to implement proposed and existing techniques and compare results in terms of accuracy & time .

3.3. Research Methodology

The user behavior should be analyses from the images which are uploaded on the twitter. It can be analyses or classified into positive or negative based on the user behavior. This type of technique can be used to increase the performance of existing algorithm of the image data classification. In proposed research methodology of this work, technique will be proposed c-mean clustering is used with the SVM classification in which we can analyze the behavior of the user. In this type of technique, c-mean clustering the similar and dissimilar type of data can be clustered into two clusters on which SVM technique can be applied for classifications. Each data point is assigning to membership that should be related to each cluster center with relevant to the distance amongst the cluster center and data point is known as c-mean clustering. The cluster center that directly depend upon on the distance of the data towards its cluster center. In the fuzzy c-mean clustering algorithm that has been utilized in both industry and academic applications that can gain the huge amount of popularity.

The c-mean clustering algorithm can provide the best results for overlapped data sets. The c-mean clustering algorithm is also better than the k-mean clustering algorithm in terms of performance. here, the data points belongs to more than one cluster center this is because the dataset assigned to each cluster center. In this work, there are some steps that can analyse the user behavior in the following:

1. In the first step of proposed technique the images are taken as input on which clustering need to be performed

2. In the second the c-mean clustering is applied which will cluster the similar and dissimilar type of data of the images

3. The technique of histogram is been applied which will analyze the features of the images

4. In the last step the technique of SVM classification is applied which will classify the data according to their similarity .

Pseudo Code of proposed algorithm,

Input : Input images for the classification

Output : Classified images

1. Initialize the centriods points from U1 to Un randomly

2. Repeat for all points in the dataset {

For every point I set

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2,$$

For every point J set

$$D_{m=} \sqrt{\left(X - X(i)\right) + (Y - Y(i)_{)}}$$

}

3. Apply technique of SVM classification to classify similar and dissimilar type of data

4. Return classified data

3.4 Proposed Flowchart

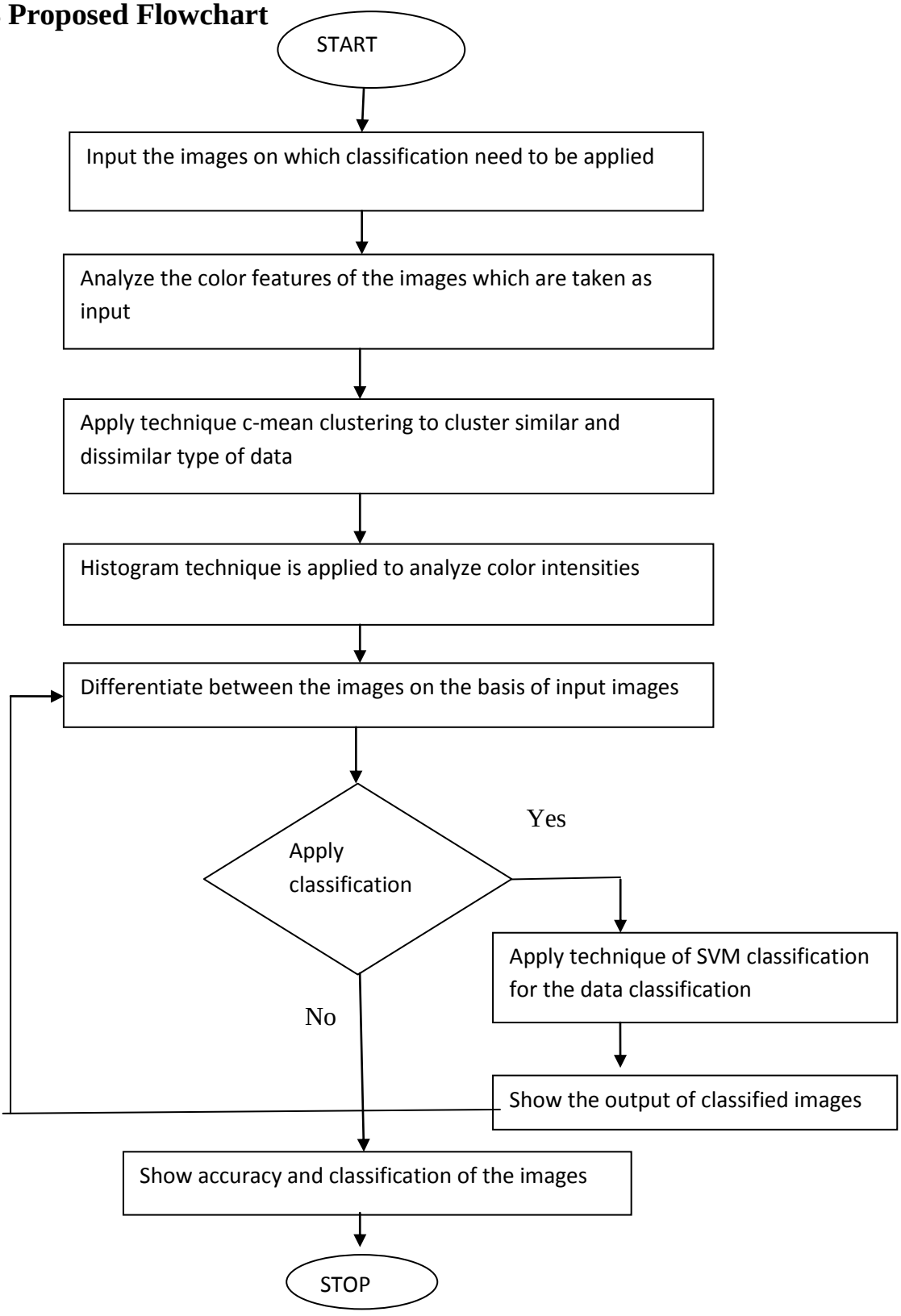

Fig 4: Flowchart

CHAPTER 4

RESULTS AND ANALYSIS

4.1 Introduction of MATLAB

The MATLAB tool can be used to perform the complex mathematical computations. In the MATLAB has many in-built functions, guide, toolboxes, GUI drag and drop based options, image processing and many more. It can be mostly used to implement the different types of algorithm, plotting graph and different design interfaces according to the requirement of the user. Matlab tool can process the elements in the form of matrix and different types of language such as JAVA, PYTHON etc. The default interface of MATLAB have the following components:

1. **Command Window:-** This is important part of the MATLAB tool, in which the user can display the output of saved code data and to execute the codes.
2. **Workspace** :- It can be used to display the allocation and de-allocation of different types of variables, type of variable and value of variable.
3. **Command History** :-In this part we can write the different types of commands that user want to execute also it display the commands.
4. **Current Folder Path** :- It can show the path of the data whre it can be stored in your device or computers. It can capture the data while giving the path of the folder which should be saved.
5. **Current Folder Data**: - This part provide the current folder path to the user which we want to execute the code.

MATLAB it can various types of structures with different data types and functions including in it. Ot can be support the classes, packages, inheritance and many more features. It can be support the developing application that can be designed by Graphical user interface(GUI).

4.2 Results

The results and discussion are in the following:

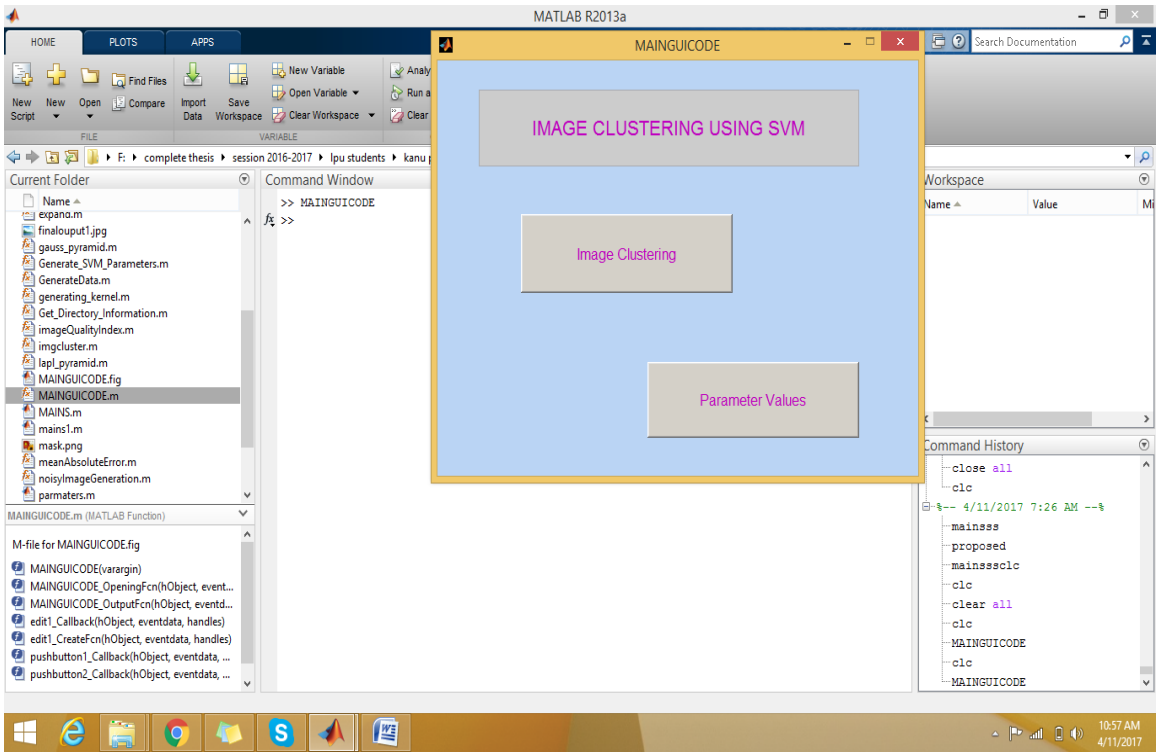

Fig 5: Default Interface

As shown in figure 5, the default interface is designed in which two buttons are shown in which image execution algorithm and parameter calculation button is shown.

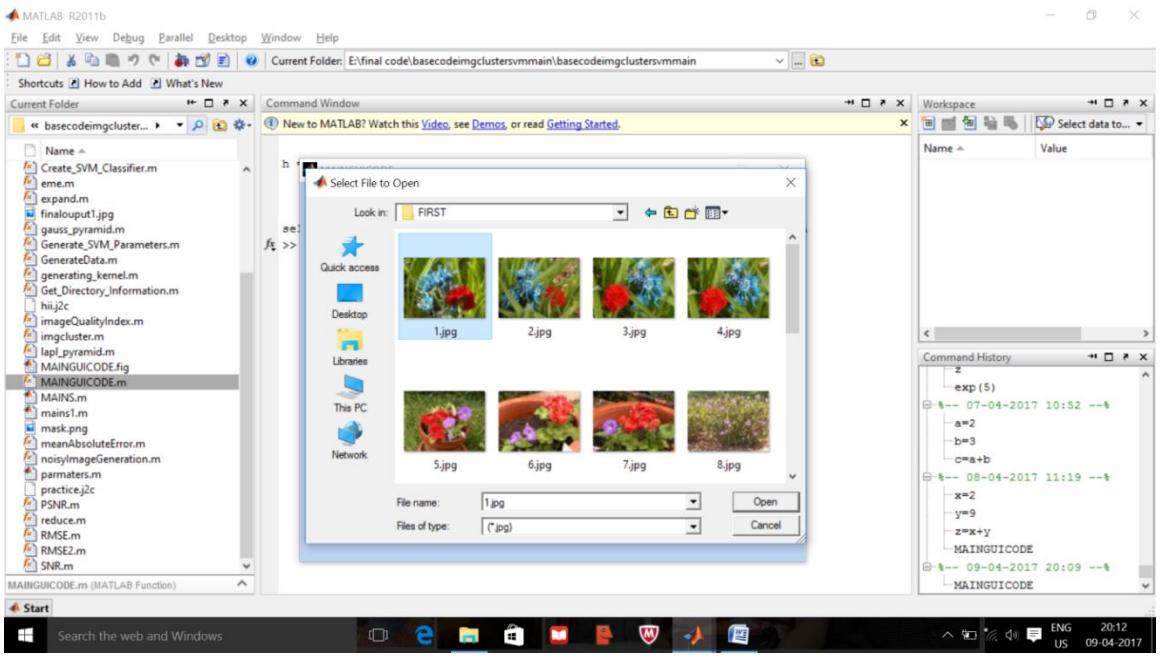

Fig 6: First Folder Images

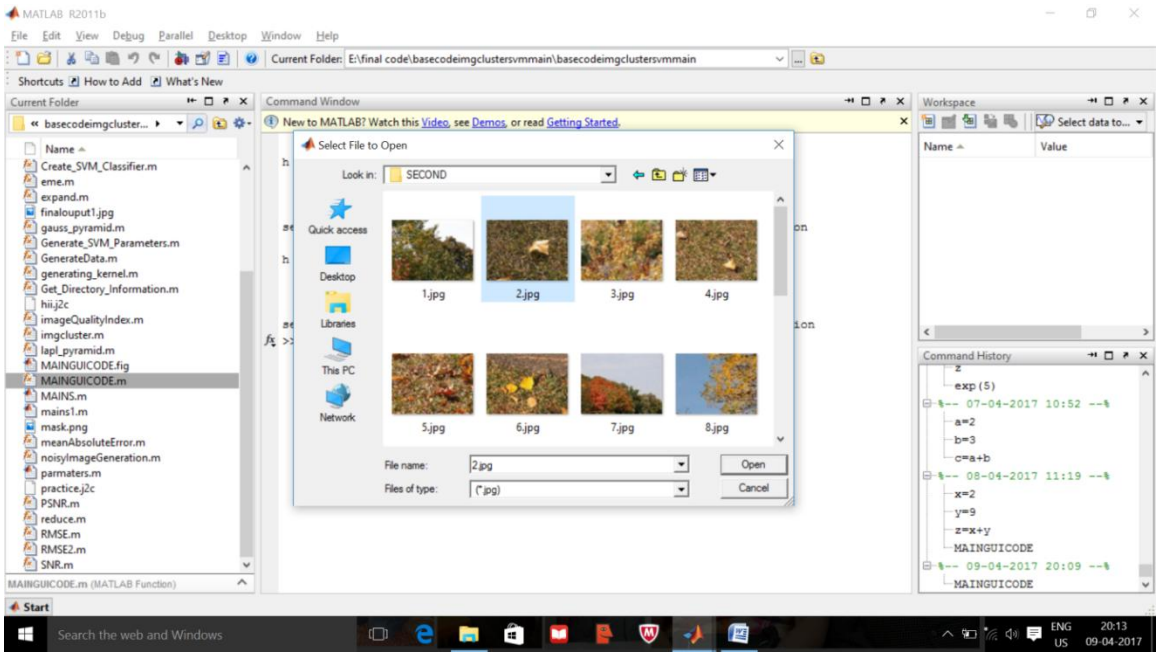

Fig 7: Second Folder Images

As shown in Fig6 and fig 7, the two images it can show in your matlab window by showing name first image and second image.

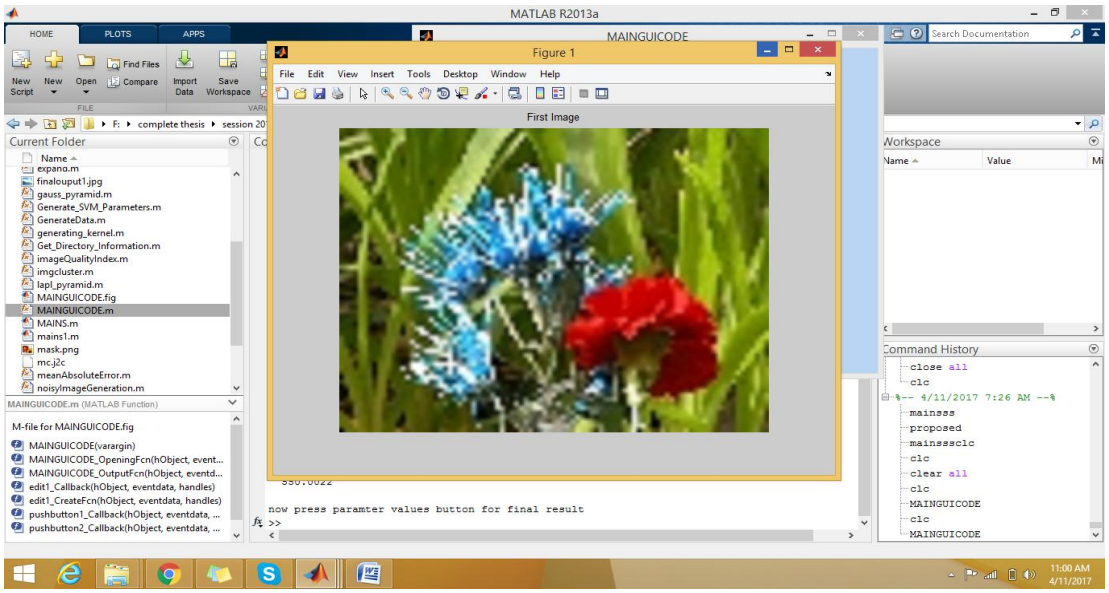

Fig 8: Input first image

As shown in the figure 8, the first image is taken as input on which sentiment analysis need to be performed.

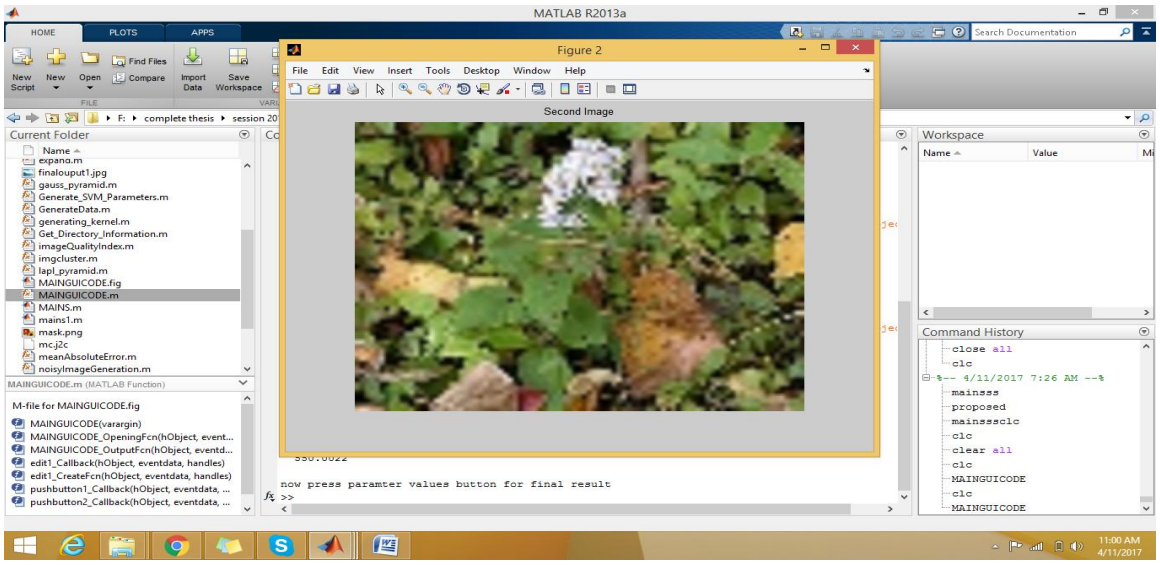

Fig 9: Input second image

As shown in the figure 9, the second image is taken as input on which sentiment analysis need to be performed.

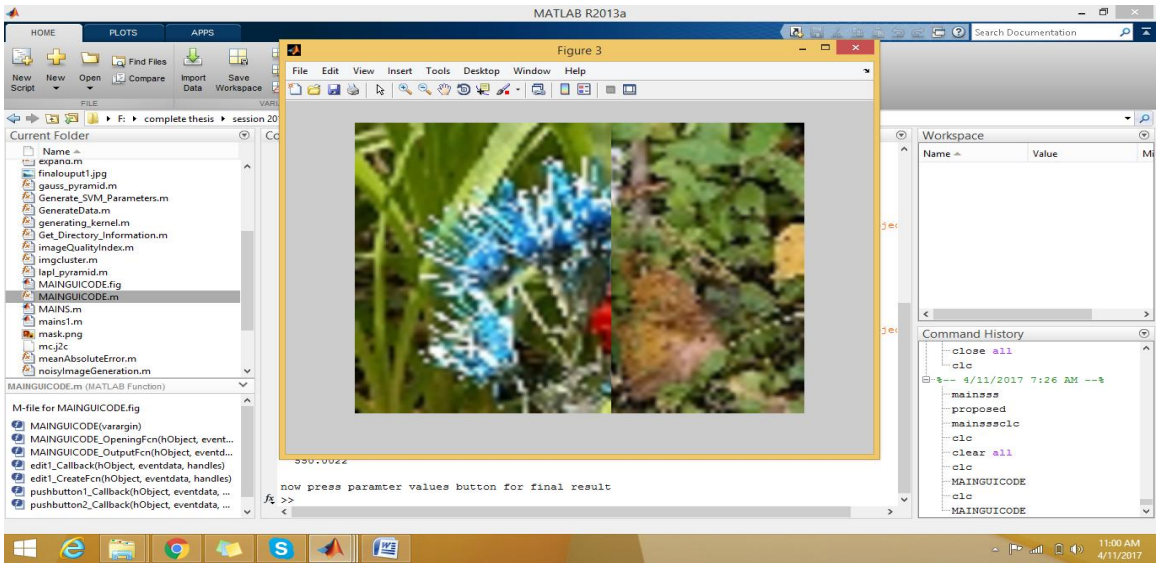

Fig 10: Clustering of image

As shown in figure 10, the two images are taken as input and these images are merged together to form one image which need to be analyzed .

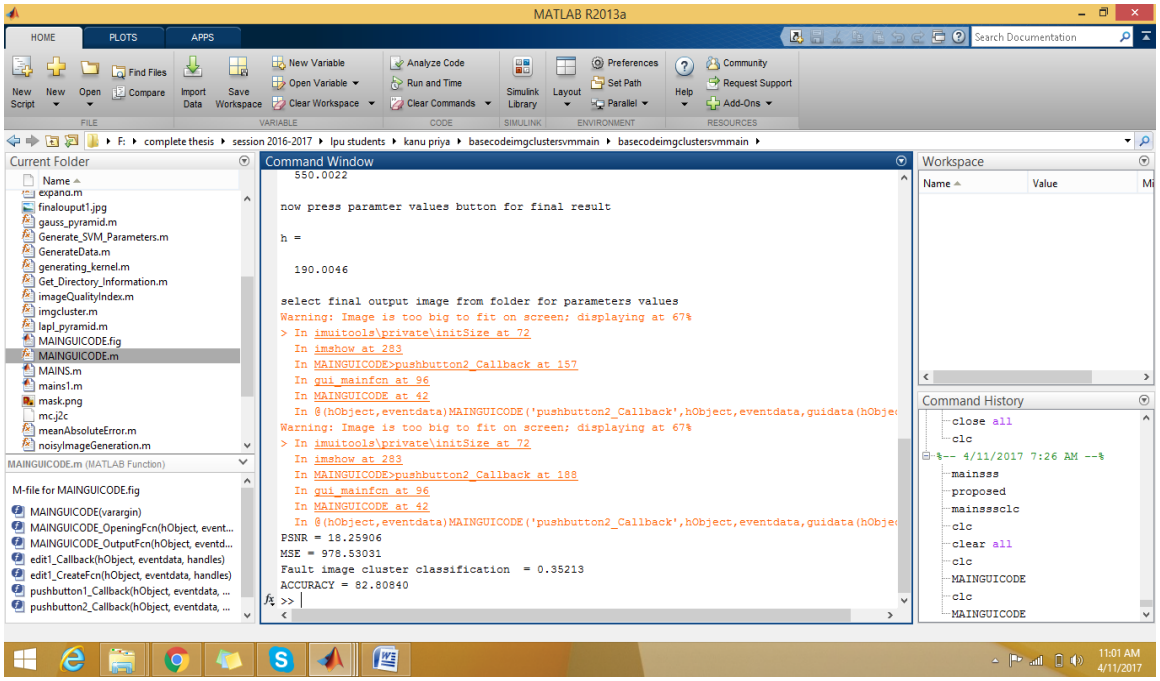

Fig 11: Result analysis

As shown in figure 11, the result of the final clustered image is shown in the form of PSNR, MSE, fault rate and accuracy

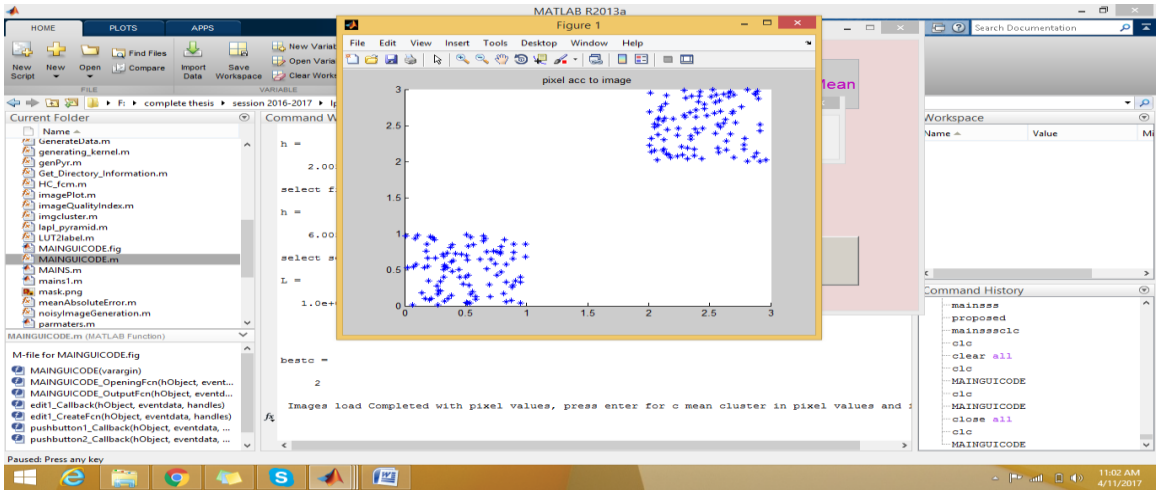

Fig 12: Clustering of image

As shown in figure 12, the images which are taken as input and two images are merged together to form one image. The technique of c-mean clustering is applied on which final result of image clustering is shown

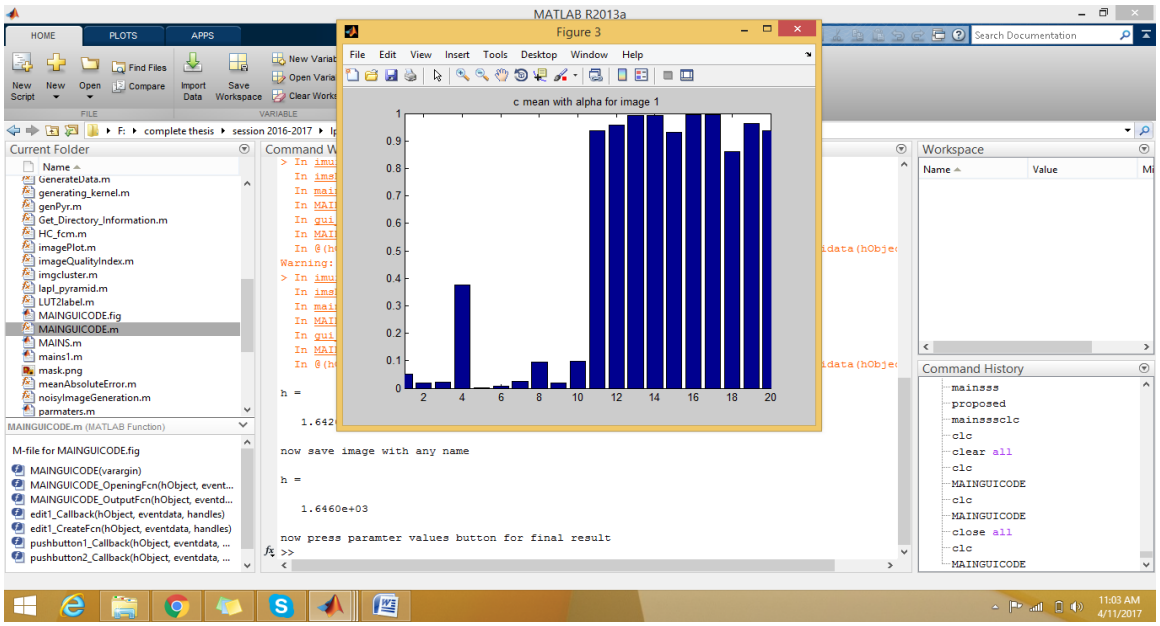

Fig 13: Histogram of first image

As shown in figure 13, the first image which is taken as input on that image technique of histogram is applied which show the color intensity of the image

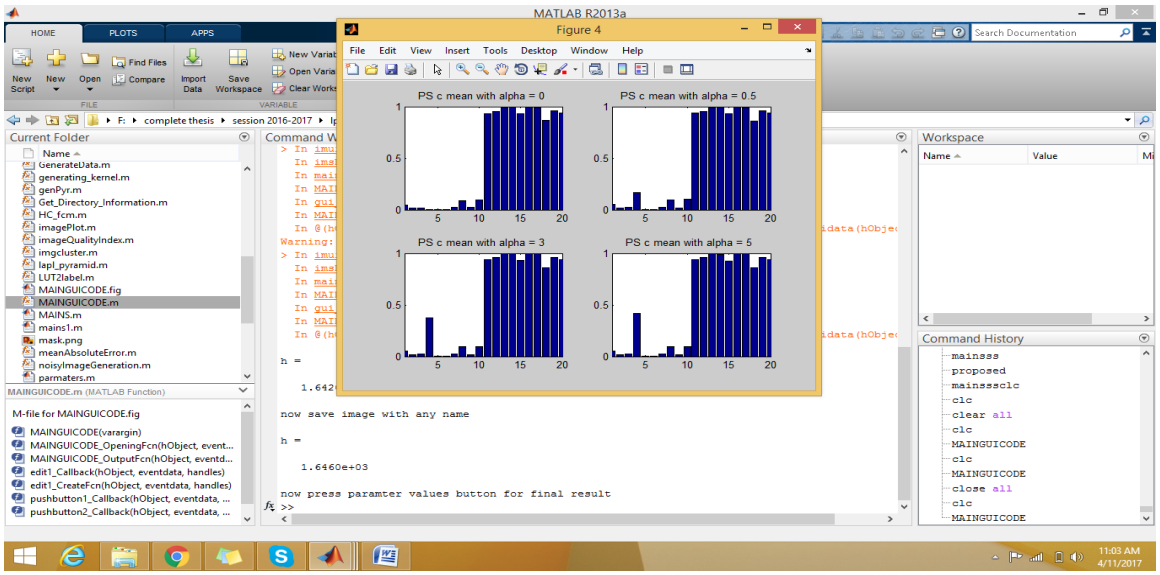

Fig 14: First Image classification

As shown in figure 14, the technique of histogram is applied which will analyze color features of the image and technique of SVM is applied which will classify the similar pixels of the images

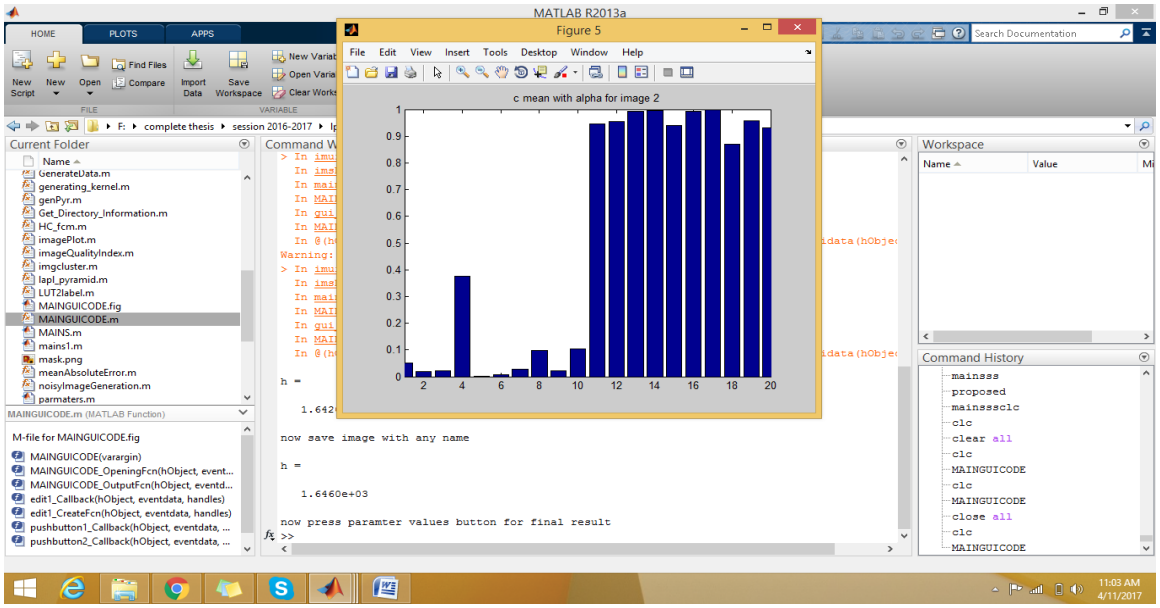

Fig 15: Histogram of second image

As shown in fiigue 15, the second image which is taken as input on that image technique of histogram is applied which show the color intensity of the image

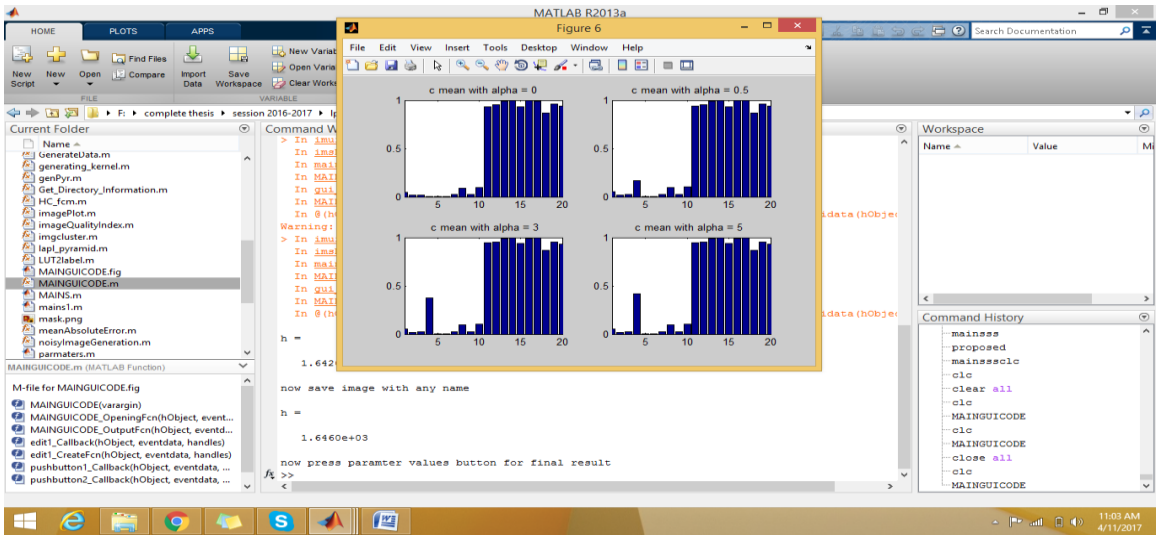

Fig 16:Second Image classification

As shown in figure 16, the technique of histogram is applied which will analyze color features of the image and technique of SVM is applied which will classify the similar pixels of the images

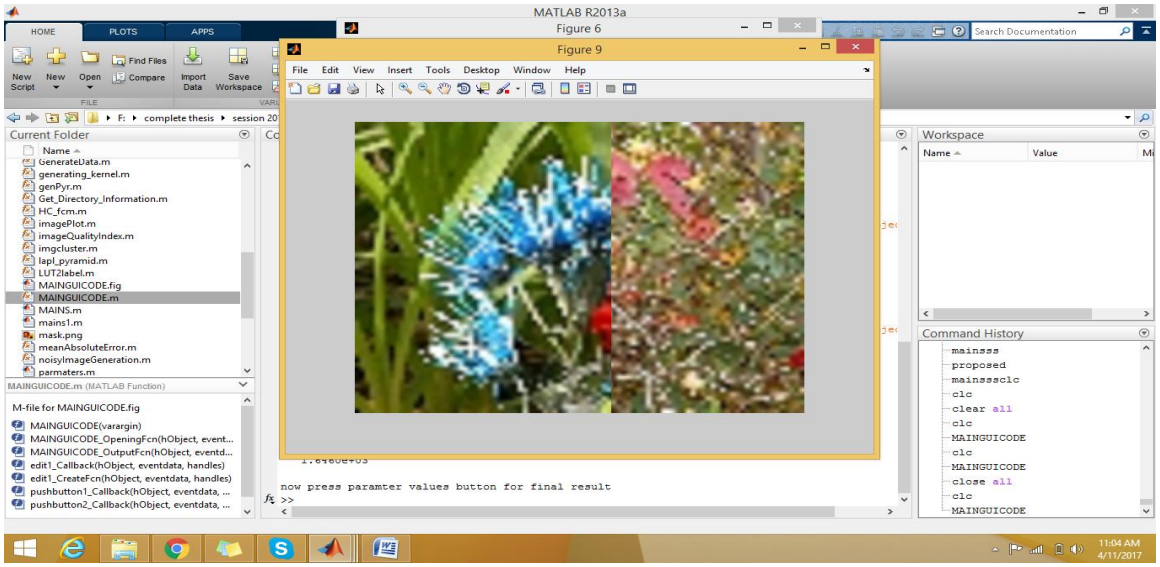

Fig 17: Final output of clustering

As shown in figure 17, the two images are joined together according to their color intensities and formed the final image as shown in the figure

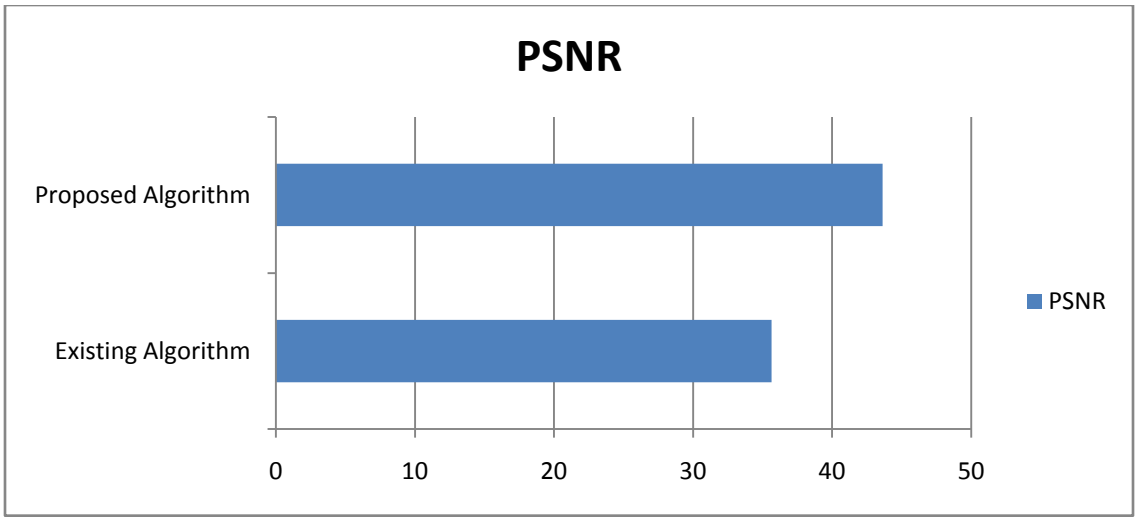

Fig 18: PSNR comparison

As shown in figure 18, the PSNR value of the proposed and existing algorithm is compared and it is been analyzed that PSNR of proposed algorithm is increased as compared to existing one

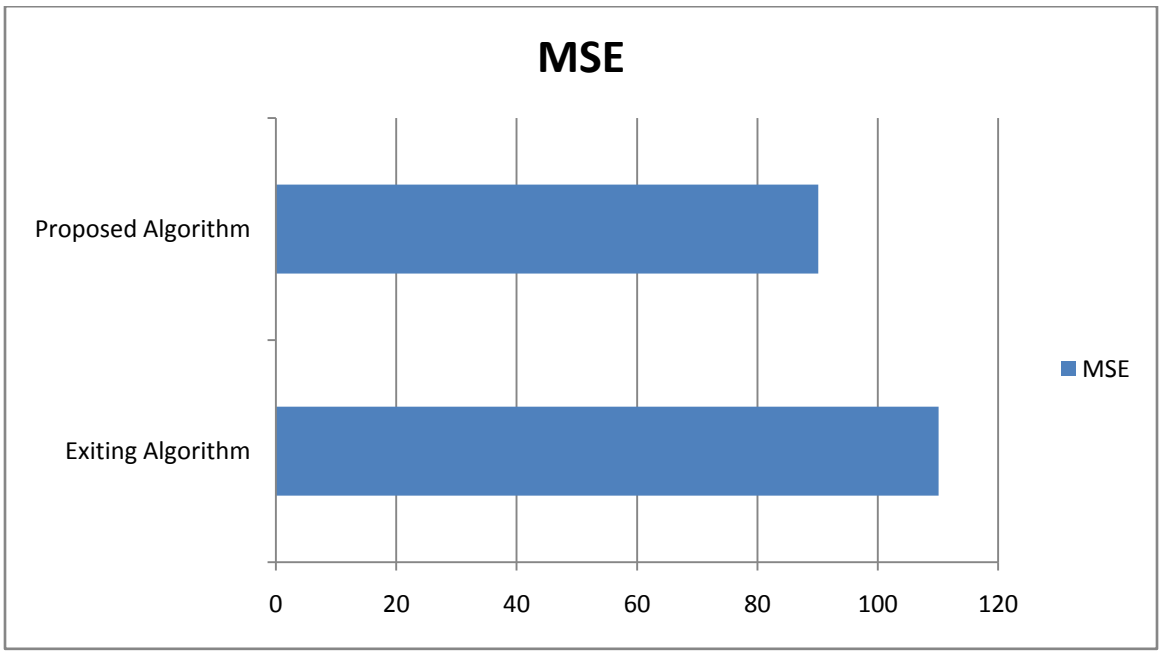

Fig 19: MSE comparison

As shown in figure 19, It is been analyzed that MSE value of the image which is generated by the proposed algorithm is less as compared to existing algorithm

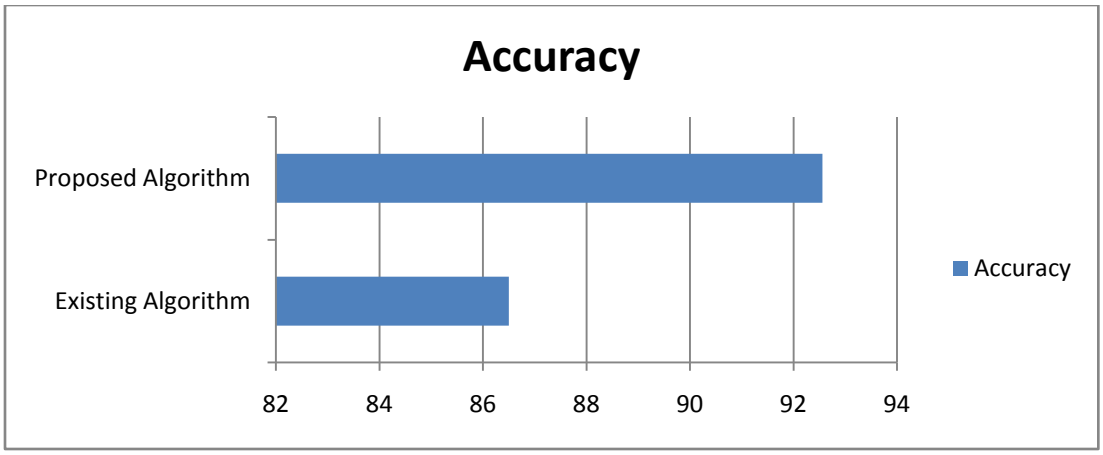

Fig 20: Accuracy Comparison

As shown in the figure 20, the accuracy of proposed algorithm is high as compared to existing algorithm

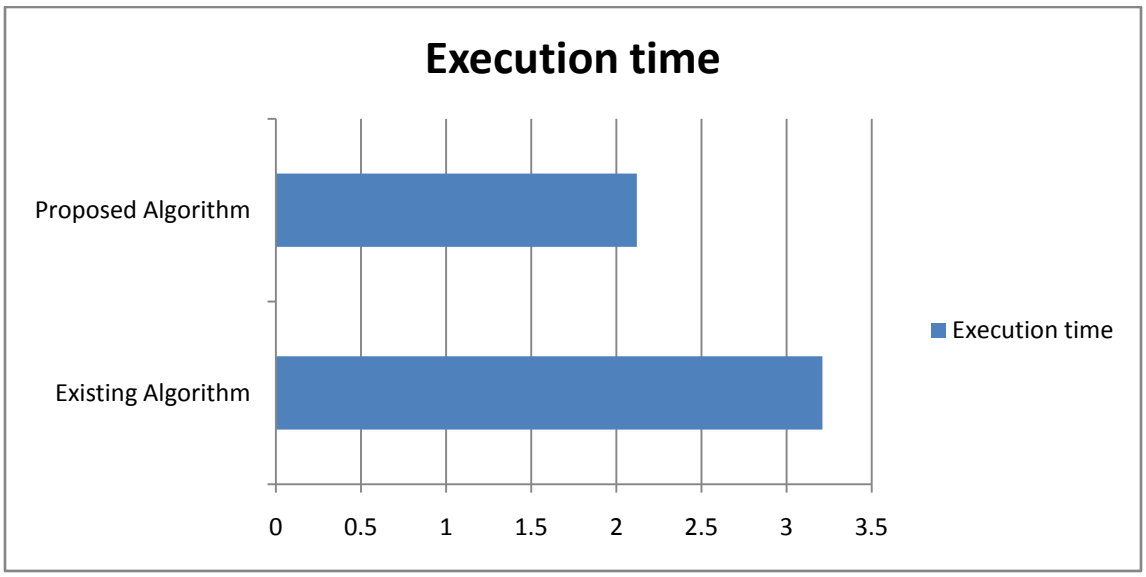

Fig 21: Execution time

As shown in figure 21, the execution time of proposed technique is less as compared to existing algorithm due to c-mean clustering used for classification

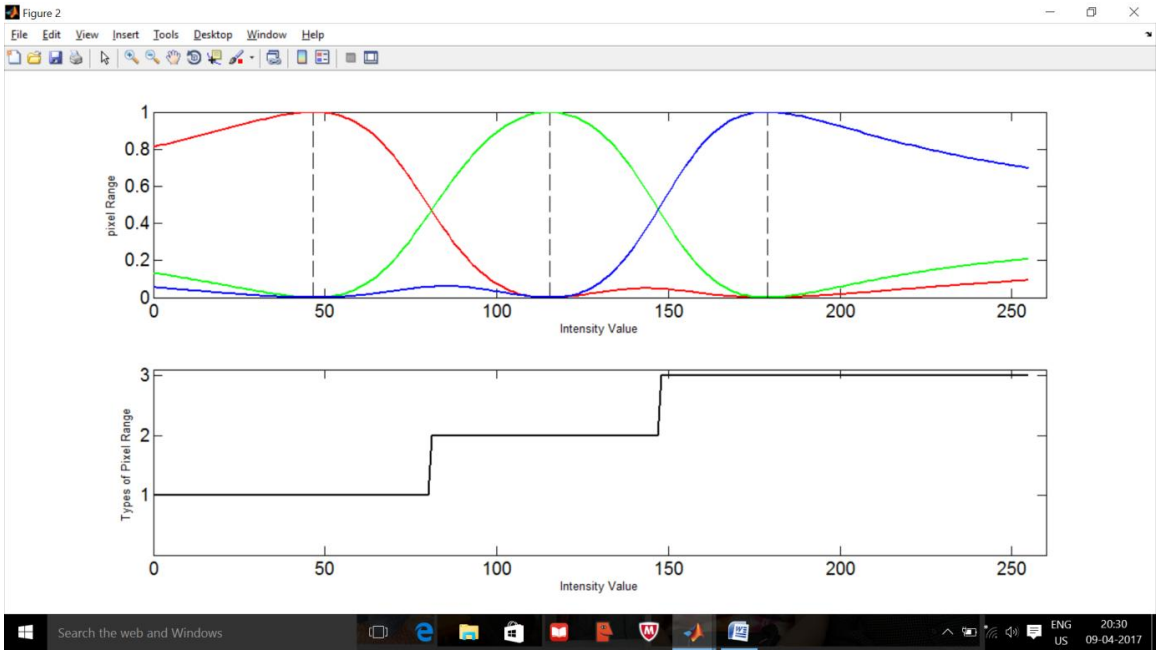

Fig 22: Pixel range & Intensity Value

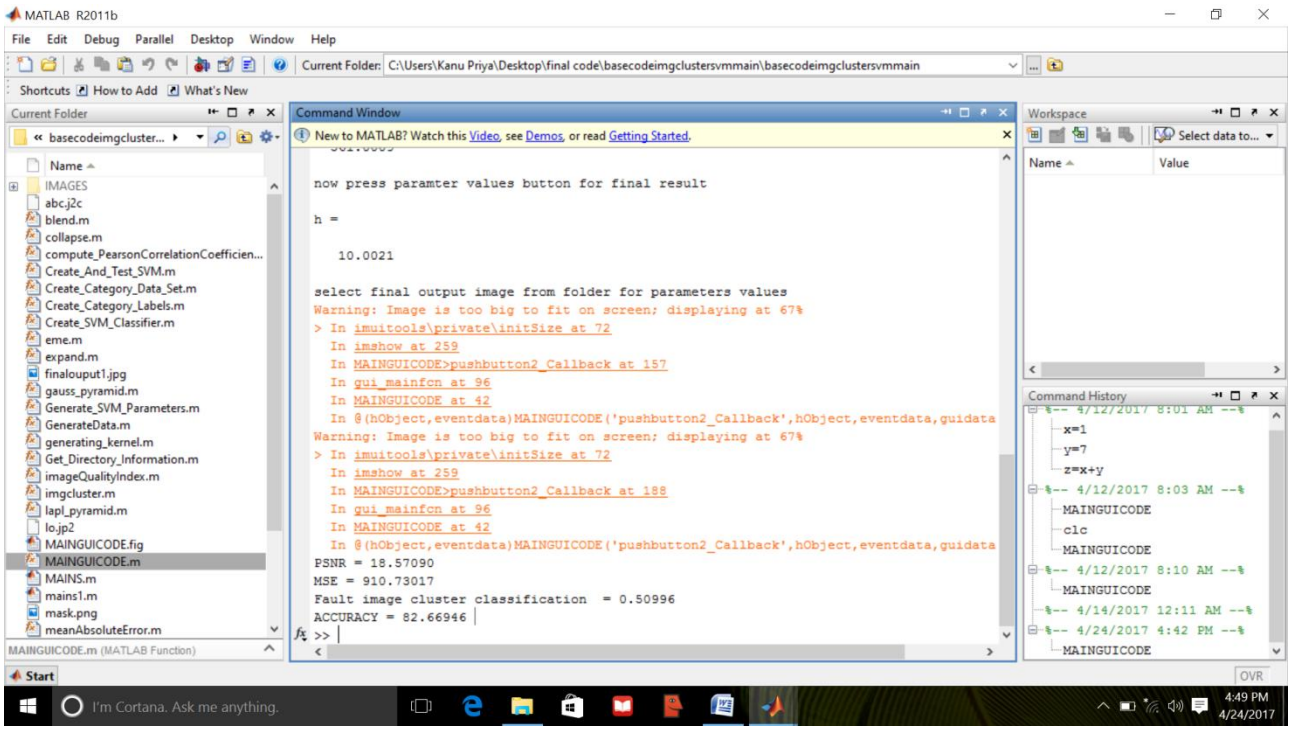

Fig 23:Base Paper Accuracy

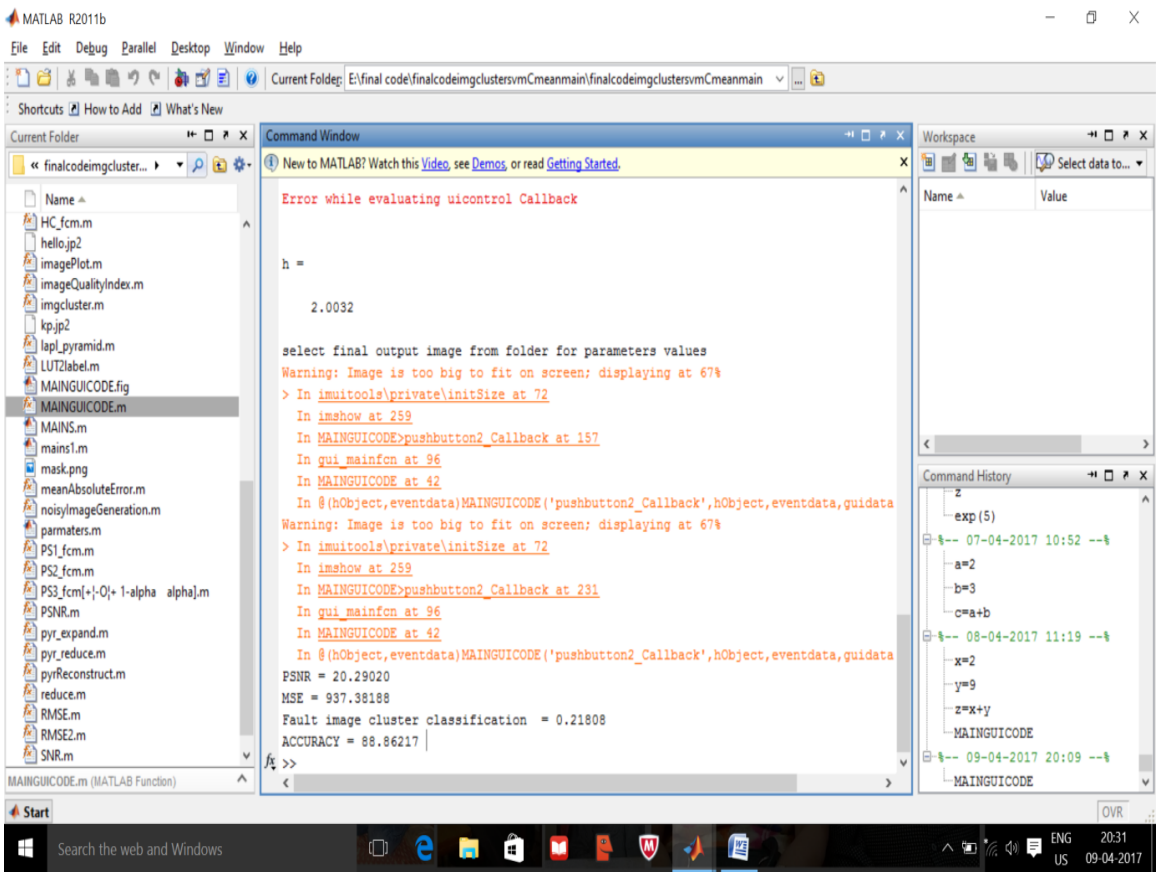

Fig 24: Final Output Image with accuracy

In the base code image cluster using SVM , the result is partially accurate which does not show the exact result of clustering of images as we explained in the above figures that are 82.66946 but in the final paper implementation image cluster using SVM and C-mean clustering as shown in Fig24, it gives the accurate result that are 88.86217 with histogram plotted in it which define the pixel intensity values and parameters values with accuracy.

CHAPTER 5

SUMMARY AND CONCLUSION

Summary and Conclusion

In this work, the paper are reviewed the research papers of social media focused on individuals subjective well-being. The trend of social media is very popular and people share their feelings on social media, provides large data related to subjective well-being. The researchers used that data to study the well-being of individuals. The higher accuracy of results and predicted outcomes was very impressive, thus gained a lot of attraction. More researchers are now attempting to get into new insights of social media analytics focusing on subjective well-being. As the data is changing rapidly, so many scientists faced different challenges such as high noise, real-time analysis of data, etc.

The real-time prediction of subjective well-being from social media may be the hot topic of future research. The real-time predicted happiness of people will be shared with friends and family. Hence, it may provide a better environment as people will adjust their mood levels according to the other individuals behavior. We have to concentrate on data, which may appears frequently different image text information, screenshot of images, screen shots of text conversations. The greater range of SWB which may suggest that the many people to share the content on social sites like text messages, images and different types of emotions over the insta-gram and twitter with greater percentage. The analysis relates on statistical formulation of a number of things into a cluster, with a socio economic indicators measures for a whole city such as average income, average housing price, and average education level which can statistics computed from hundreds of thousands of images shared in that city.

To extract the content based features from images which include the images of screen shots in more than one cities and other countries and we also want to interview of some twitter users who can use the image text to better understand their reasons of sharing images. We can do the statically analysis using socio economic

indicators that can measure the city with different attributes over the hundreds and thousands of images shares per city. We can compare the methods and results of our study with the use of historical images to learn about the social life and its details to investigate the strengths and weaknesses . Most method that can be used to extract bag of words features from texts in post and used some specific language database for measure the sentiments.

REFERENCES

[1] L. MANOVICH, "Trending: the promises and the challenges of big social data," *Debates in the Digital Humanities*, pp. 460–475, 2011.

[2] D. KAHNEMAN, A. B. Krueger, D. SCHKADE, N. Schwarz, and A. A. Stone, "Would you be happier if you were richer? a focusing illusion," *science*, vol. 312, no. 5782, pp. 1908–1910, 2006.

[3] H. S. Lett, J. A. Blumenthal, M. A. Babyak, A. Sherwood, T. Strauman, C. Robins, and M. F. Newman, "Depression as a risk factor for coronary artery disease: evidence, mechanisms, and treatment," *Psychosomatic medicine*, vol. 66, no. 3, pp. 305–315, 2004.

[4] L. Mitchell, M. R. Frank, K. D. Harris, P. S. DODDS, and C. M. Dan Forth, "The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place," 2013.

[5] A. J. Oswald and S. Wu, "Objective confirmation of subjective measures of human well-being: Evidence from the USA," *Science*, vol. 327, no. 5965, pp. 576–579, 2010.

[6] L. MANOVICH, A. Tifentale, M. Yazdani, and J. Chow, "The exceptional and the everyday: 144 hours in kiev," in *Big Data (Big Data), 2014 IEEE International Conference on*. IEEE, 2014, pp. 72–79.

[7] N. Hochman and L. Manovich, "Zooming into an instagram city: Reading the local through social media," *First Monday*, vol. 18, no. 7, 2013.

[8] S. A. Golder and M. W. Macy, "Diurnal and seasonal mood vary with work, sleep, and day length across diverse cultures," *Science*, vol. 333, no. 6051, pp. 1878–1881, 2011.

[9] M. R. Frank, L. Mitchell, P. S. Dodds, and C. M. Danforth, "Happiness and the patterns of life: A study of geo-located tweets," *Scientific reports*, vol. 3, 2013.

[10] J. C. Eichstaedt, H. A. Schwartz, M. L. Kern, G. Park, D. R. Labarthe, R. M. Merchant, S. Jha, M. Agrawal, L. A. Dziurzynski, M. Sap *et al.*, "Psychological language on twitter predicts county-level heart disease mortality," *Psychological science*, vol. 26, no. 2, pp. 159–169, 2015.

[11] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: Liwc and computerized text analysis methods," *Journal of language and social psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[12] Jyoti Islam, Yanqing Zhang," Visual Sentiment Analysis for Social Images Using Transfer Learning Approach", 2016, IEEE, 978-1-5090-3936

[13] Eunjeong Ko, Chanhee Yoon, Eun Yi Kim," Discovering Visual Features for Recognizing User's Sentiments in Social Images", 2016, IEEE, 978-1-4673-8796

[14] Sara Ted mori, Rashed AI- Lahaseh," Towards a Selfie Social Network with Automatically Generated Sentiment-Bearing Hashtags", 2016, IEEE, 978-1-4673-8914

[15] F. Ciullo, C. Zucco, B. Calabrese, G. Agapito, P. H. Guzzi, M. Cannataro," Computational Challenges for Sentiment Analysis in Life Sciences", 2016, IEEE, 978-1-5090-2088

[16] Marie Katsurai, Shin'ichi Satoh," IMAGE SENTIMENT ANALYSIS USING LATENT CORRELATIONS AMONG VISUAL, TEXTUAL, AND SENTIMENT VIEWS", 2016, IEEE, 927y592-3453909-768-67

[17] Eun Jeong Ko and Eun Yi Kim," Recognizing the Sentiments of Web Images using Hand-designed Features", 2015, IEEE, 918-1-4613-1290-9

[18] C. Xu, S. Cetintas, K.-C. Lee, and L.-J. Li, "Visual sentiment prediction with deep convolutional neural networks," arXiv preprint arXiv:1411.5731, 2014.

[19] Dua'a Al-Hajjar, Afraz Z. Syed," Applying Sentiment and Emotion Analysis on Brand Tweets for Digital Marketing", 2015, IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)

[20] Parinya Sanguansat,” Paragraph2Vec-Based Sentiment Analysis on Social Media for Business in Thailand”, 2016, IEEE, 348765342-43565-65765-555

[21] Yilin Wang, Baoxin Li,” Sentiment Analysis for Social Media Images”, 2015, IEEE, 978-1-4673-8493-3

[22] P. S. Smita1, "Use of Data Mining in Various Field: A Survey Paper," IOSR Journal of Computer Engineering, May-June, 2014.

[23] Mehrdad Yazdani, Lev Manovich,” Predicting Social Trends from Non-photographic Images on Twitter”, 2015, IEEE, 978-1-4799-9926

[24] R. S. Manjunath T.N, "Realistic analysis of data warehousing and data mining application in education domiain," International journal of machine learning and computing, August,2012

[25] S. Asur and B. A. Huberman 2010 Predicting the Future with Social Media arXiv:1003.5699v1

APPENDIX

SVM: Support Vector Machine

SWB: Subjective Well Being

KDD: Knowledge Discovery in Database

GDP: Gross Domestic Product

FCM: Fuzzy Centeroid Mean

K-mean clustering Algorithm

PSNR: Peak signal-to-noise ratio

MSE: Mean squared error

MSD: Mean squared deviation

OLAP: Online analytical processing

GUI: Graphical User Interface

Checklist for Dissertation-II Supervisor

Name: _____ UID: _____ Domain: _____
Registration No: _____Name of student:_____

Title of Dissertation:
_____

- ☐ Front pages are as per the format.
- ☐ Topic on the PAC form and title page are same.
- ☐ Front page numbers are in roman and for report, it is like 1, 2, 3…….
- ☐ TOC, List of Figures, etc. are matching with the actual page numbers in the report.
- ☐ Font, Font Size, Margins, line Spacing, Alignment, etc. are as per the guidelines.
- ☐ Color prints are used for images and implementation snapshots.
- ☐ Captions and citations are provided for all the figures, tables etc. and are numbered and center aligned.
- ☐ All the equations used in the report are numbered.
- ☐ Citations are provided for all the references.
- ☐ **Objectives are clearly defined.**
- ☐ Minimum total number of pages of report is 50.
- ☐ Minimum references in report are 30.

Here by, I declare that I had verified the above mentioned points in the final dissertation report.

Signature of Supervisor with UID