HUMAN CENTRIC CYBER SOCIAL COMPUTING

Enrollment no(s): 15103141, 15103144, 15103164
Name of Student(s): Bhavya Varshney, Jasleen Dhillon, Varn Gupta
Name of Supervisor: Dr. Adwitiya Sinha

March 2019

Submitted in partial fulfillment of the Degree of
Bachelor of Technology

in

Computer Science Engineering

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING & INFORMATION TECHNOLOGY

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY, NOIDA

TABLE OF CONTENTS

CHAPTER 1- INTRODUCTION

1.1 General Introduction

Sentiment Analysis is a Natural Language Processing and Information Extraction task that aims to obtain writer's feelings expressed in positive or negative comments, questions and requests, by analyzing a large numbers of documents. Generally speaking, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall tonality of a document. It is a process of identifying whether the opinion or reviews expressed in a piece of work is positive, negative or neutral. Sentiment analysis is useful in social media monitoring to automatically characterize the overall feeling or mood of consumers as replicated in social media toward a specific brand or political or social views that a group of people share and determine whether they are viewed positively or negatively on the web. In recent years, the exponential increase in the Internet usage and exchange of public opinion is the driving force behind Sentiment Analysis today. The Web is a huge repository of structured and unstructured data. The analysis of this data to extract latent public opinion and sentiment is a challenging task.

People use such platforms to express their opinions on current affairs, political campaigns, day to day activities, sports, and other services. One such platform is Twitter.Twitter is a social networking site that allows its users all over the world to post what they want to say in the form of tweets i.e. term used for chunk of data sent using twitter which was originally restricted to 140 characters but this limit was doubled on November 7, 2017. This public data can be taken as raw data mainly for opinion extraction, for analyzing customer satisfaction and for rating different government schemes and finally to do sentiment analysis. In general a person tweets without giving a second thought to what he or she wants to share. This can be a strong viewpoint they have about something happening currently. But usually the anonymity offered by the internet gives the power to people to say anything and this can lead to hatred content.

With this project we aim to use the machine learning-based approaches to sentiment analysis and brings out the salient features of techniques in place. The prominently used techniques and methods in machine learning- emoticon based sentiment analysis included in the project are SVC, LinearSVC classifier and SentiWordNet classifier. With the progression of the project we see how these machine-learning algorithms produce different results and accuracies.

1.2 Problem Statement

One of the biggest problem plaguing the internet today is hate-speech and cyber-bullying. Hidden in the anonymity of the internet, using feigned accounts, the perpetrator is sometimes unknown to the fact that his/her actions may have adverse effects on someone which in some cases leads to suicide. Social media is rife with hate-speech. A quick glance through facebook comments or at different tweets and various twitter threads demonstrates how pervasive the problem is. Not only do these propagate hatred but also create division in the society.

Humans have a spectrum of emotions –happy, sad, angry, disappointed, surprised, proud, scared etc. We are intuitive in sensing the tone with which a person wrote something. But to make a machine analyze all these emotions is not that easy as it will not be able to get the contextual meaning and rather breaks down the emotion of the text with the help of words underlying in it. So we resort to natural language processing, text analysis, computational linguistics and biometrics to implement this task, and hence gather the essence with which the text has been written with. The project methodology can be applied by various social media websites and various other websites the likes of twitter, facebook, youtube where the epidemic of hate speech has spread. The various algorithms also identify offensive speech which is not hate speech. This could be used by various conglomerates to track social media response of the products launched and the how are the companies and their products are received by social media users.

First of all, the ways in which people communicate things to each other have become increasingly asynchronous. Even if people may use texting, chats, direct messaging, emails, and even discussion forums, for communicating at a steady pace, or to respond to each other the very moment that they receive a message, the digital tools and platforms allow for delays. These possibilities for asynchronicity lead to a ‘conversational relaxation’. Emoticons are considered to be handy and reliable indicators of sentiment, and hence could be used either to automatically generate a training corpus or to act as evidence feature to enhance sentiment classification.
Emoticons are introduced as expressive, non-verbal components into the written language, mirroring the role played by facial expressions in speech. Their role is mainly pragmatic: emoticons give a positive or negative sense to written sentences by a visual expression. According to this consideration, there is a relationship between the sentiment orientation of emoticons and messages. Emoticons have been distinguished in two main categories, i.e. positive and negative.

Addition of emoticons to a text message can shed light on the anonymity of the emotion involved. Sometimes people use only emoticons to express their content, anger and so on. In the new era of online visual communication, when you have to be fast and clear, emojis are getting a strong relevance as the main language that allows us to communicate with anyone globally. Customers know it and brands too. Collecting and analyzing data on emojis give companies useful insights on how customers are feeling about a new product, a new campaign, or about the brand itself. Emojis can even help identify where there is a need to improve consumer engagement by picturing users’ moods, attitudes, and opinions.Thus, this creates the significance of our research.

1.3 Significance/Novelty of the problem

With the rise of social media, emoticons today have become an important means to communicate our feelings. But till now, there have not been much research and effort to analyse the emoticons and take them in consideration while determining the mood of the text/tweet.This leaves a very important deficit in the current mood-recognition softwares. To fill this deficit, we have implemented an algorithm which also uses the various emoticons to determine the mood of different texts.

1.4 Brief Description of the Solution Approach

The project is segregated into two major parts-

-The project aims at targeting the tweets shared on Twitter on the basis of emotion behind them. The underlying idea is to categorize tweets and to determine whether the tweet/text could be considered as hate speech or not, that is whether the content published is in accordance to accepted language by the society and not demeaning and denigrating. It identifies hate speech by employing various machine learning algorithms. We also derive the context in which various words are used and based on it identifies hate speech. It tries to keep the idea of freedom of speech intact and at the same time curtail hate speech which various other algorithms have failed to do.

-Emoji prediction is a fun variant of sentiment analysis. In this segment, we build what's called a classifier that learns to associate emojis with sentences. Although there are many technical details, the principle behind the classifier is very simple: we start with a large amount of sentences or tweets that contain emojis collected from Twitter messages. Then we look at features from those sentences (words, word pairs, etc.) and train our classifier to associate certain features with their (known) smileys. For example, if the classifier sees the word "happy" in many sentences that also has the smiley 😂, it will learn to classify such messages as 😂. On the other hand, the word "happy" could be preceded by "not" in which case we shouldn't rely on just single words to be associated with certain smileys. The classifier learns to look at the totality of many word sequences found in a sentence and figures out what class of smiley would best characterize that sentence.

CHAPTER 2- LITERATURE SURVEY

2.1 Summary of papers studied

Our day-to-day life has always been influenced by what people think. Ideas and opinions of others have always affected our own opinions. The explosion of Web 2.0 has led to increased activity in Podcasting, Blogging, Tagging, Contributing to RSS, Social Bookmarking, and Social Networking. As a result there has been an eruption of interest in people to mine these vast resources of data for opinions. Sentiment Analysis or Opinion Mining is the computational treatment of opinions, sentiments and subjectivity of text. In this report, we take a look at the various challenges and applications of Sentiment Analysis.

COMPARATIVE REVIEW ON APPROACHES TO SENTIMENT ANALYSIS

The machine learning method

It incorporates machine learning algorithms to deduce the sentiment by training on a known dataset. This approach to sentiment classification is supervised and allows effective text classification. Machine learning classification necessitates two different sets of documents, namely for training and testing. A training set is used by an automatic classifier to learn and differentiate attributes of documents, and a test set is used to check the performance of the automatic classifier. There are many machine learning techniques adopted to classify the reviews. Machine learning techniques like NB, ME, and SVM have achieved better performances in opinion mining.

Support Vector Machine (SVM)

It is also used for text classification based on a discriminative classifier. The approach is based on the principle of structural risk minimization. First the training data points are separated into two different classes based on a decided decision criteria or surface. The decision is based on the support vectors selected in the training set. Among the different variants of SVM, the multiclass SVM is used for sentiment analysis. The centroid classification algorithm first calculates the centroid vector for every training class. Then the similarities between a document and all the centroids are calculated and the document is assigned a class based on these similarities values.

2.2 Integrated summary of the literature studied

The following table contains integrated summary of the research papers:

PAPER 1

Title of the paper	Sentiment Analysis using Support Vector Machine based on Feature Selection and Semantic Analysis
Authors	Bhumika M. Jadav, Vimalkumar B. Vaghela
Year of Publication	2016
Publishing Details	International Journal of Computer Applications

Approach Used	Summary	Conclusion
Different features techniques like unigrams, bigrams, unigrams + POS tagging, Position and unigrams + Position are used and then machine learning techniques like Naïve bayes, Maximum Entropy and support vector machine classification algorithms are applied on preprocessed dataset. Classification algorithms perform better than human based classifier	The paper consists of preprocessed dataset to convert unstructured reviews into structured form. Then we have used lexicon based approach to convert structured review into numerical score value. In lexicon based approach we have preprocessed dataset using feature selection and semantic analysis. Stop word removal, stemming, POS tagging and calculating sentiment score with help of SentiWordNet dictionary have been done in preprocessing part. Then we have applied classification algorithm to classify opinion as either positive or negative. Support vector machine algorithm is used to classify reviews where RBF kernel SVM is modified by its hyper parameters which are soft margin constant C , Gamma γ. So optimized SVM gives good result than SVM and naïve bayes.	Sentiment analysis has been done for movie Review, Twitter and Gold dataset using optimized SVM. Here Comparison is made between Optimized Support Vector Machine towards Support Vector Machine and naïve bayes classifier. Modifying hyper parameter value of RBF kernel SVM gives better result compare to Support Vector Machine and Naïve Bayes algorithm. Hyper parameters are soft margin constant C and Gamma γ. Proposed approach has found optimal value for hyper parameter which classifies dataset with more accuracy than existing system. There are many SVM kernel functions available with many hyper parameters. These values can be modified to improve accuracy.

PAPER 2

Title of the paper	Combining Dependency Parsing with Shallow Semantic Analysis for Chinese Opinion-Element Relation Identification
Authors	Chen Mosha, Yao Tianfang
Year of Publication	2010
Publishing Details	IEEE- International Universal Communication Symposium

Approach Used	Summary	Conclusion
The paper presents a novel method to identify the opinion-element relation based on the dependency parsing analysis as well as shallow semantic analysis, using an ontology dictionary and a collocation database to take full consideration of the semantic behind the topic and sentiment.	The paper defines opinion-element relation to be the relation between a topic word and a sentiment word in a complete sentence (ended by period), and the sentiment word must directly (direct relation) or indirectly (indirect relation) modify the topic word, which has reflected the essence of opinion. First it is a relation pair between two elements, and then it is with opinion, so we call it opinion-element relation. The paper builds an ontology dictionary.	The method used starts from the analysis of opinion-element relation structure and according to the common feature we summarized, we have raised our idea for the task, we take the semantic information into account to some extent. Besides, we have given a simple but efficient algorithm to set up a collocation database for further use.

PAPER 3

Title of the paper	The Effects of Pre-Processing Strategies in Sentiment Analysis of Online Movie Reviews
Authors	Harnani Mat Zin, Norwati Mustapha, Masrah Azrifah Azmi Murad, Nurfadhlina Mohd

	Sharef
Year of Publication	2017
Publishing Details	in 11th IEEE International Conference on Data Mining

Approach Used	Summary	Conclusion
The support vector machine (SVM) with linear and non-linear kernel has been considered as classifier for the classification of the reviews. The performance of the classifier is critically examined based on the results of precision, recall, f-measure, and accuracy. Two different features representations were used which are term frequency and term frequency-inverse document frequency.	The main objective of this study was to examine the effect of several pre-processing strategies with the help of term weighting scheme in analyzing online reviews by using a variety of Support Vector Machine (SVM) classifiers. Specifically, the researchers used unigram as feature selection and term frequency (TF) and term frequency-inverse document frequency (TF-IDF) as the feature representation. SVM with linear and non-linear kernels were chosen to perform the classification process. This classifier was selected because it has been proven to be an effective machine learning technique for text categorization. Moreover, this classifier is also one of the common methods of classification for linear and non-linear problems.	The feature vectors of the reviews were pre-processed in several ways and the effects of these on the classifier's performance in term of precision, recall, f-measure, and accuracy were investigated. The results show that by removing the stop words, meaningless words, numbers, and word less than 3 characters, favorably affected the performance of the classification. Another important finding from this study was, the SVM with non-linear kernel achieved the best performance results for both features representation, TF and TFIDF.

PAPER 4

Title of the paper	Genetic Algorithm and Confusion Matrix for Document Clustering
Authors	A. K. Santra, C. Josephine Christy
Year of Publication	2012
Publishing Details	IJCSI International Journal of Computer Science Issues, Vol. 9

Approach Used	Summary	Conclusion
The confusion matrix is more commonly named contingency table in which the matrix could be arbitrarily large. The number of correctly classified instances is the sum of diagonals in the matrix; all others are incorrectly classified accurately. Improved Genetic algorithm starts with an initial population which is created consisting of randomly generated rules. Each rule can be represented by a string of bits. Based on the notion of survival of the fittest, a new population is formed to consist of the fittest rules in the current population, as well as offspring of these rules.	This paper presents an improved genetic algorithm which is used to evaluate the weights of the metrics such as F-measure, purity and accuracy. We apply improved genetic algorithm to find out and identify the potential informative features combinations for classification and then use the F-Measure to determine the fitness in genetic algorithm. The improved GA is general purpose search algorithm which provides rules inspired by natural genetic populations to evaluate solutions to problems. In our method, not as usual, an individual is joined together of the real-coded metrics' weight, and it's more natural to indicate the optimization problem in the continuous domain.	The improved Niche memetic algorithm and improved genetic algorithm have been designed and implemented by using confusion matrices. Our proposed method is applied to real data sets with an abundance of irrelevant or redundant features. Improved GA relies on confusion matrices and uses the F-measure as the fitness function. In this case, identifying a relevant subset that adequately captures the underlying structure in the data can be particularly useful. It is concluded by remarking that we consider the experimental results can further be improved through a fine-tuning of the GA parameters.

PAPER 5

Title of the paper	A Proposed Solution for Sentiment Analysis on Tweets to Extract Emotions from Ambiguous Statements
Authors	Dyuti Shukla, Mihika Shah, Prerna Parmeshwaran, Kiran Bhowmick
Year of Publication	2015
Publishing Details	International Journal of Engineering Research & Technology (IJERT)

Approach Used	Summary	Conclusion

In order to provide a more accurate picture of the user's opinions, it is simply not enough to classify them by polarity. More in-depth knowledge of their expression is needed, thus in this project, it is aimed to extract human emotions conveyed in the tweet with the help of existing emotion models from the domain of psychology.	As consumers move towards social media platforms like Twitter and Facebook to air their views about a variety of products, performing sentiment analysis on their responses becomes a desirable activity that can return a wealth of information about public perception. However, information posted in such networks is designed for human consumption instead of computers, and the nuances a person can catch from them are difficult for a machine to interpret. Thus most work in this field has always concentrated on polarity detection of the opinion into three broad fields of positive, negative or neutral. Paper aims to look at other techniques and emotion models that would aid us in helping computers understand the emotions attached to such ambiguous statements. It compares various techniques used for sentiment analysis to that end, and propose a solution for the same.	Performing sentiment analysis on data obtained from Twitter is a huge challenge because of the amount of ambiguity involved. Due to the widespread usage of slang, wrong spellings, emoticons etc. it becomes difficult for automatic detection of emotions from tweets. This project is a small step towards the efficient automation of sentiment analysis by focusing on ambiguous statements. The system proposed by us also attempts to extract actual emotions from tweets. However, in the future, the scope can be extended to accommodate the same. Finally, the project can be extended to work for natural languages other than English.

PAPER 6

Title of the paper	A Comparative Study Of Sentiment Analysis Techniques
Authors	MR. S. M. VOHRA, PROF. J. B. TERAIYA
Year of Publication	2013
Publishing Details	Journal Of Information, Knowledge And Research In Computer Engineering

Approach Used	Summary	Conclusion
There are two main approaches for sentiment analysis: machine learning based and lexicon based. Machine learning based approach uses classification technique to classify text. Lexicon based method uses sentiment dictionary with opinion words and match them with the data to determine polarity. They assigns sentiment scores to the opinion words describing how Positive, Negative and Objective the words contained in the dictionary are.	The growth of social web contributes vast amount of user generated content such as customer reviews, comments and opinions. This user generated content can be about products, people, events, etc. This information is very useful for businesses, governments and individuals. While this content meant to be helpful analyzing this bulk of user generated content is difficult and time consuming. So there is a need to develop an intelligent system which automatically mine such huge content and classify them into positive, negative and neutral category. Sentiment analysis is the automated mining of attitudes, opinions, and emotions from text, speech, and database sources through Natural Language Processing (NLP). The objective of this paper is to discover the concept of Sentiment Analysis in the field of Natural Language Processing, and presents a comparative study of its techniques in this field	Applying Sentiment analysis to mine the huge amount of unstructured data has become an important research problem. Now business organizations and academics are putting forward their efforts to find the best system for sentiment analysis. Although, some of the algorithms have been used in sentiment analysis gives good results, but still no technique can resolve all the challenges. Most of the researchers reported that Support Vector Machines (SVM) has high accuracy than other algorithms, but it also has limitations. More future work is needed on further improving the performance of the sentiment classification. There is a huge need in the industry for such applications because every company wants to know how consumers feel about their products and services and those of their competitors.

PAPER 7

Title of the paper	An Approach to Sentiment Analysis using Artificial Neural Network with Comparative Analysis of Different Techniques
Authors	Pranali Borele , Dilipkumar A. Borikar
Year of Publication	2016
Publishing Details	IOSR Journal of Computer Engineering

Approach Used	Summary	Conclusion
It incorporates machine learning algorithms to deduce the sentiment by training on a known dataset. This approach to sentiment classification is supervised and allows effective text classification. Machine learning techniques like NB, ME, and SVM have achieved better performances in text categorization.	This paper reviews the machine learning-based approaches to sentiment analysis and brings out the salient features of techniques in place. The prominently used techniques include - Naïve Bayes, Maximum Entropy and Support Vector Machine, K-nearest neighbour classification. NB has very simple representation but doesn't allow for rich hypotheses. Also the assumption of independence of attributes is too constraining. ME estimates the probability distribution from data, but it performs well with only dependent features. For SVM may provide the right kernel, but lacks the standardized way for dealing with multi-class problems. For improving the performance regarding correlation and dependencies between variables, an approach combining neural networks and fuzzy logic is often used.	Applying Sentiment analysis to mine the large amount of unstructured data has become an important research problem. Some of the algorithms have been used in sentiment analysis to gives good results, but no technique can resolve all the challenges. Our study suggests that the ANN implementations would result in improved classification, combining the best of artificial neural network with fuzzy logic.

PAPER 8

Title of the paper	Sentiment Analysis Based Approaches for Understanding User Context in Web Content
Authors	M. SAKTHIVEL , G. HEMA
Year of Publication	2013
Publishing Details	International Conference on Communication Systems and Network Technologies

Approach Used	Summary	Conclusion
Current research offers several interesting approaches to the challenge posed by sentiment analysis . Abstracting away from specific implementations, these approaches can be classified into 3 categories: the lexical-phrasal approach, the compositional semantics approach, and the Machine Learning approach. Tf-Idf and ngrams is used for the analysis.	Sentiment analysis applies natural language processing techniques and computational linguistics to extract information about sentiments expressed by authors and readers about a particular subject, thus helping users in making sense of huge volume of unstructured Web data. Applications like review classification, product review mining and trend prediction benefit from sentiment analysis based techniques. This paper presents a study of different approaches in this field, the state of the art techniques and current research in Sentiment Analysis based approaches for understanding user's context. We show that information about social relationships can be used to improve user-level sentiment analysis. The main motivation behind our approach is that users that are somehow "connected"	Feature selection techniques have proved vital in the performance of several text categorization tasks, as they enhance the performance of the classification system considerably. This is on the line of recent approaches which face the challenge of increasing the user involvement in building the Semantic Web an alternative could be to integrate in ArsEmotica the use of automatic techniques. Finally, future possible uses include the development of emotion-aware search engines and of emotional tag clouds. This would open the way to a plethora of applications, including iOS and Android apps, not only with a cultural flavour (along the lines of the application in the previous section) but also

	may be more likely to hold similar opinions; therefore, relationship information can complement what we can extract about a user's viewpoints from their utterances.	more intrinsically related to leisure. The work can be considered as a building block for analyzing sentiment with minimal usage of linguistic resources and no complex patterns.

PAPER 9

Title of the paper	Sentiment polarity with SentiWordNet and machine learning Classifiers
Authors	Akshaya R. Garje, Karbhari V. Kale
Year of Publication	2015
Publishing Details	International Journal of Advanced Research in Computer Science

Approach Used	Summary	Conclusion
This research offers the results of making use of the SentiWordNet lexical resource to computerized sentiment classification on a labelled dataset.	Supervised machine mastering is the search for algorithms that cause from externally provided instances to provide widespread hypotheses, which then make predictions approximately further instances. The research uses Naïve Bayes algorithm and Support Vector Machine (SVM) to analyze the data.	In this research, three different website data reviews are used for training and testing the best between SentiWordNet and machine learning classifier methods. Using bigrams instead of unigrams is a trick for improving performance in text classification. The accuracies obtained using SentiWordNet are less than as compared to both machine learning classifiers. Hence, it can be said that machine learning

		classifiers with approximately 80-83% accuracies.

PAPER 10

Title of the paper	Enhanced Sentiment Analysis and Polarity Classification Using SentiWordNet-Vocabulary
Authors	Mr.Thavasi, Dr.P.Golda Jeyasheeli, P.Devisri
Year of Publication	2017
Publishing Details	International Journal of Advanced Research in Basic Engineering Sciences and Technology

Approach Used	Summary	Conclusion
The paper uses a SentiWordNet approach for sentiment analysis.	The proposed system consists of various steps: • Data Collection • Tag Identification POS Tagging POS Tagger to SWN-Tag • Preprocessing • Feature Extraction Sentiment Score Generate Index Term • Label Generation • Recommendation. The dataset consists of a total of 50,000 movie reviews in the large movie review dataset where 25,000 are positive and 25,000 are labeled negative.	SentiWordNet provides an efficient way for unsupervised text categorization. However, there is a need to improve the performance of SentiWordNet. The application of supervised learning has been the prime research focus for text classification. Labeled datasets are required in order to train supervised classifiers. This becomes the key concern as tagged datasets are not easily available and it takes huge effort and resources to build such datasets and this is where the SentiWordNet comes into play.

CHAPTER 3: REQUIREMENT ANALYSIS AND SOLUTION APPROACH

3.1 Overall description of the project

The project involves series of files from picking up tweets from a user timeline of politically active people and placing them under analysis. The tweets from different users are appended then in a single file and our doc2vec and word2vec algorithms are used to analyze hate or offensive nature of the tweet regarding the nation. Some words or phrases which may be objectionable are also noted and put under the required nature as hate, neutral or offensive. Then we used a dataset psychExp.txt for gathering the emotion and labeling the text or tweet based on the emoticon detected. Use of confusion matrix is done for depiction of these emoji labelled texts via the two algorithms- SVC and LinearSVC. SVC and LinearSVC classifiers are used to analyze the tone of the tweets regarding the emoticons used . Some words or phrases which may be objectionable are also noted.

Word2vec method: Word2vec is a two-layer neural net that processes text. Its input is a text corpus and its output is a set of vectors: feature vectors for words in that corpus. While Word2vec is not a deep neural network, it turns text into a numerical form that deep nets can understand. Deeplearning4j implements a distributed form of Word2vec for Java and Scala, which works on Spark with GPUs.
Word2vec's applications extend beyond parsing sentences in the wild. It can be applied just as well to genes, code, likes, playlists, social media graphs and other verbal or symbolic series in which patterns may be discerned.

The purpose and usefulness of Word2vec is to group the vectors of similar words together in vector space. That is, it detects similarities mathematically. Word2vec creates vectors that are distributed numerical representations of word features, features such as the context of individual words. It does so without human intervention.

Given enough data, usage and contexts, Word2vec can make highly accurate guesses about a word's meaning based on past appearances. Those clusters can form the basis of search, sentiment analysis and recommendations in such diverse fields as scientific research, legal discovery, e-commerce and customer relationship management.

The output of the Word2vec neural net is a vocabulary in which each item has a vector attached to it, which can be fed into a deep-learning net or simply queried to detect relationships between words.

Doc2vec method: Doc2vec(aka paragraph2vec, aka sentence embeddings) modifies the word2vec algorithm to unsupervised learning of continuous representations for larger blocks of text, such as sentences, paragraphs or entire documents. The algorithm then runs through the sentences iterator twice: once to build the vocabulary, and once to train the model on the input data, learning a vector representation for each word and for each label in the dataset.

SVC Classifier: In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVC in a probabilistic classification setting). An SVC model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

LinearSVC Classifier: Linear SVM is the newest extremely fast machine learning (data mining) algorithm for solving multiclass classification problems from ultra large data sets that implements an original proprietary version of a cutting plane algorithm for designing a linear support vector machine. LinearSVC is a linearly scalable routine meaning that it creates an SVC model in a CPU time which scales linearly with the size of the training data set. Our comparisons with other known SVC models clearly show its superior performance when high accuracy is required.

Data source and dataset:

Labeled_data- An excel file containing tweets with labels 1(neutral), 3(hate speech) and 2(offensive but not hate). This file has a further breakdown into two files named as label1 and label2 which will be used further for training and testing by the classifiers.

psychExp.txt (Our Labeled Dataset) - A text file containing tweets with labels 1 and 0 for the following emotions stored in an array : [joy, anger, fear, sad, disgust, shame, guilt]. This file has a labeled dataset including emoticons which will be used for training and testing by the classifiers.

Tools and Libraries used:

In the Python implementation of the code: collections, nltk.classify.util, nltk.metrics, NaiveBayesClassifier, MaxEntropy Classifier, SklearnClassifier, csv, cross_validation, LinearSVC, SVC, random, stopwords, itertools, BigramCollocationFinder, precision, Nltk.metrics, scores, spearman, paice, confusion matrix.

3.2 Requirement Analysis (Functional/Non-Functional/Logical Database requirements)

Functional Requirements:

Functional requirements are the functions or the features that must be included in any system to satisfy the business needs and be acceptable to the users. Based on this the functional requirements that the system must have are as follows:

1. System should be able to process new tweets stored in database after retrieval.
2. System should be able to retrieve tweets from the internet.
3. System should be able to analyze data.
4. System should be able to classify sentiments from each tweet.
5. System should be able to check polarity of each tweet.
6. System should be able to assign a score to the tweet.

Non Functional Requirements:

Non functional requirements is a description of features, characteristics and attributes of the system as well as any constraints that may limit the boundaries of the proposed system .

The non functional requirements are essentially based on the performance, information, economy, control and security efficiency and services.

Based on these the non functional requirements are as follows:

1. User Friendly
2. System should provide better accuracy
3. System should easily search the database for any query
4. System should perform with efficient throughput
5. To perform with efficient response time

Logical Database Requirements:

1. Concise - expose as few functions as possible.
2. Consistent - expose unified interfaces, no need to explore new interface for each task.
3. Flexible - helps to solve complex tasks.
4. Fast - maximize efficiency per single thread, transparently scale to multiple threads on multi-core machines.
5. Memory efficient - use streams and iterators not keep data in RAM if possible.

3.3 Solution Approach

Through our project, we have tried to implement the best way to differentiate hateful social presence of a person from an offensive one. But the task is not as easy as it sounds. We trained and tested a dataset(CrowdFlower) and implemented multiple machine learning sentiment analysis techniques to get the maximum accuracy.

For this task we tried to exploit the different techniques like Naive Bayes, SVM, Maximum Entropy, TF-IDF classifiers which are defined to work in different manner from each other, hence the results are different from one another. With implementation of every new technique we tried to evaluate what the previous algorithm failed to do and how to overcome the drawbacks of the earlier one. We compare the algorithms with our proposed algorithm for finding which one performed better.

After data files are created, we train and then test the data on the following algorithm which includes emoticon prediction -

Support Vector Machine : Support vector machine is non probabilistic algorithm which is used to separate data linearly and nonlinearly. We have used SVC and LinearSVC for further classification.

SVM classifier formula is defined as following:

Equation 1: $f(x) = aik\ x,\ xi + bn\ i = 1$

where $k(x, xi) = exp(-\,|\,x - x'\,|\ 2\ 2\sigma)$

In nonlinear data, Dimension is transformed to higher dimension so cost is increased due to multiplication of test tuple with every support vectors. So due to higher dimension space, training tuples are in form of φ(Xi)* φ (Xj) , which are replaced by kernel K(Xi , Xj) = φ (Xi) * φ (Xj). K(Xi , Xj) is kernel function and it is Mathematically equivalent to product of φ (Xi) and φ (Xj) So there is no need of nonlinear mapping. Further process is same as linear data case. SVM can able to handle linear separation on the high dimension nonlinear input data.

A **confusion matrix** is a matrix (table) that can be used to measure the performance of an machine learning algorithm, usually a supervised learning one. Each row of the confusion matrix represents the instances of an actual class and each column represents the instances of a predicted class. Example of confusion matrix usage to evaluate the quality of the output of a classifier on the iris data set. The diagonal elements represent the number of points for which the predicted label is equal to the true label, while off-diagonal elements are those that are mislabeled by the classifier. The higher the diagonal values of the confusion matrix the better, indicating many correct predictions.

CHAPTER-4 MODELING AND IMPLEMENTATION DETAILS

4.1 Design Diagrams

4.1.1 Use Case diagrams

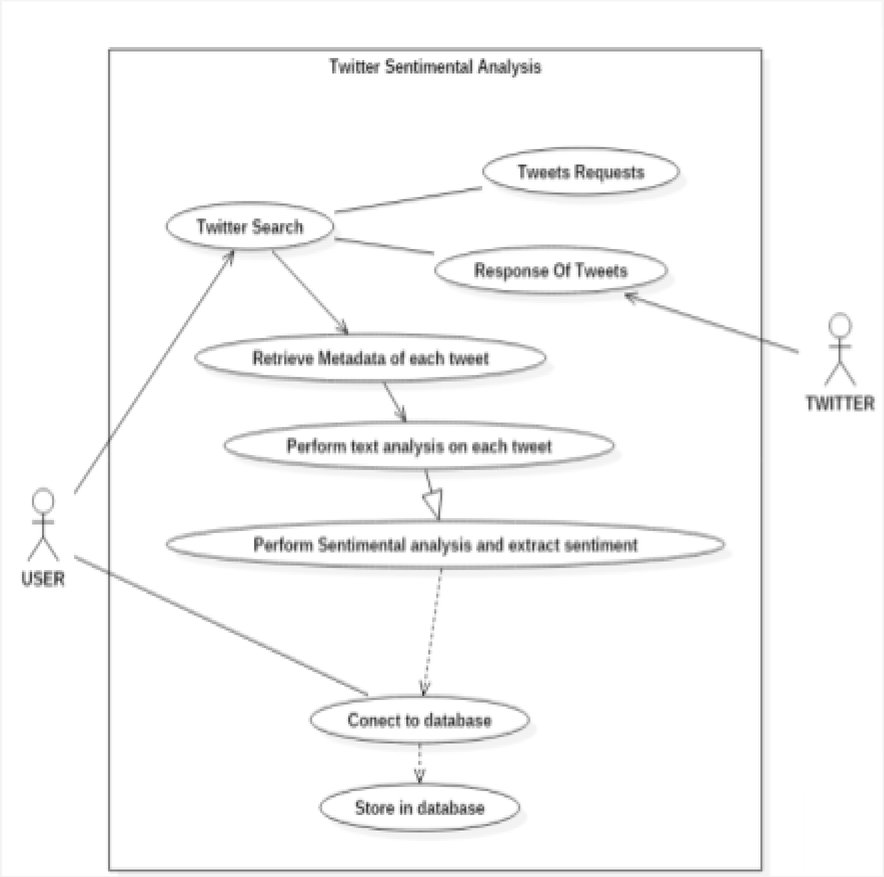

Twitter Sentimental Analysis
Tweets Requests
Twitter Search
Response Of Tweets
Retrieve Metadata of each tweet
TWITTER
Perform text analysis on each tweet
Perform Sentimental analysis and extract sentiment
USER
Conect to database
Store in database

Fig 1.1(a) : A use case diagram depicting the actors and actions involved in the sentimental analysis of data

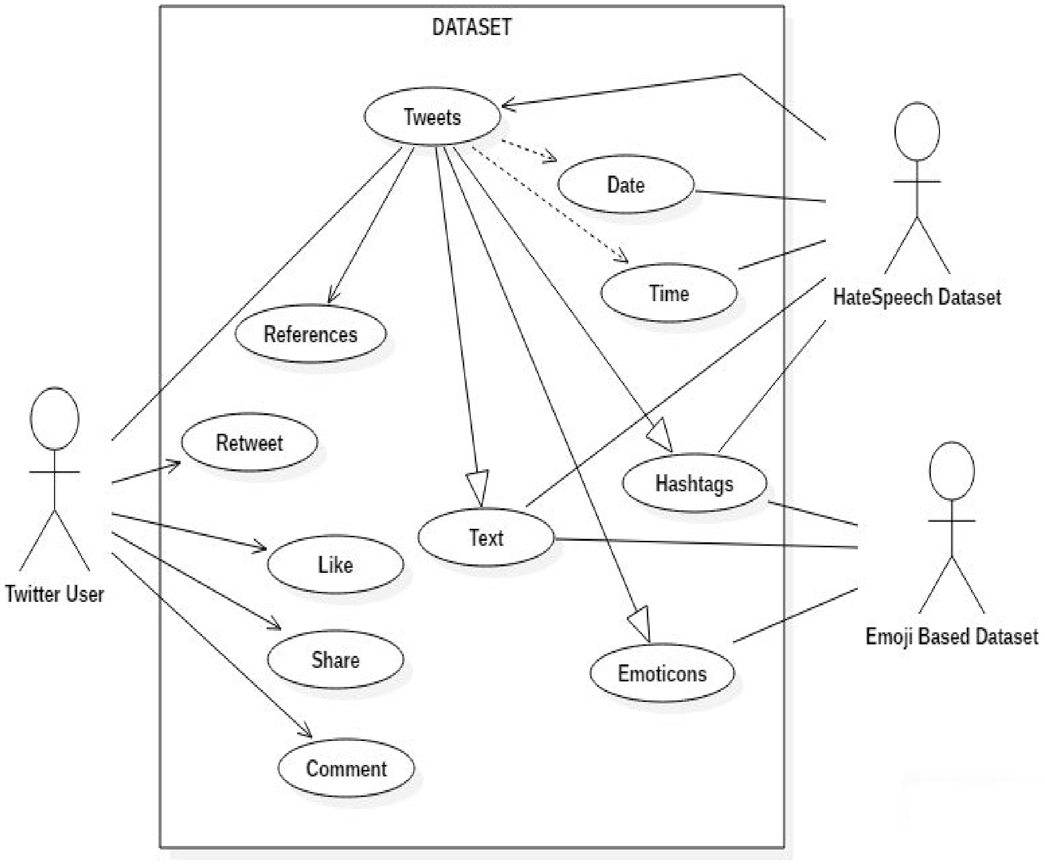


Fig 1.1(b) : A use case diagram depicting the actors and actions presenting the dataset used.

4.1.2 Class diagrams / Control Flow Diagrams

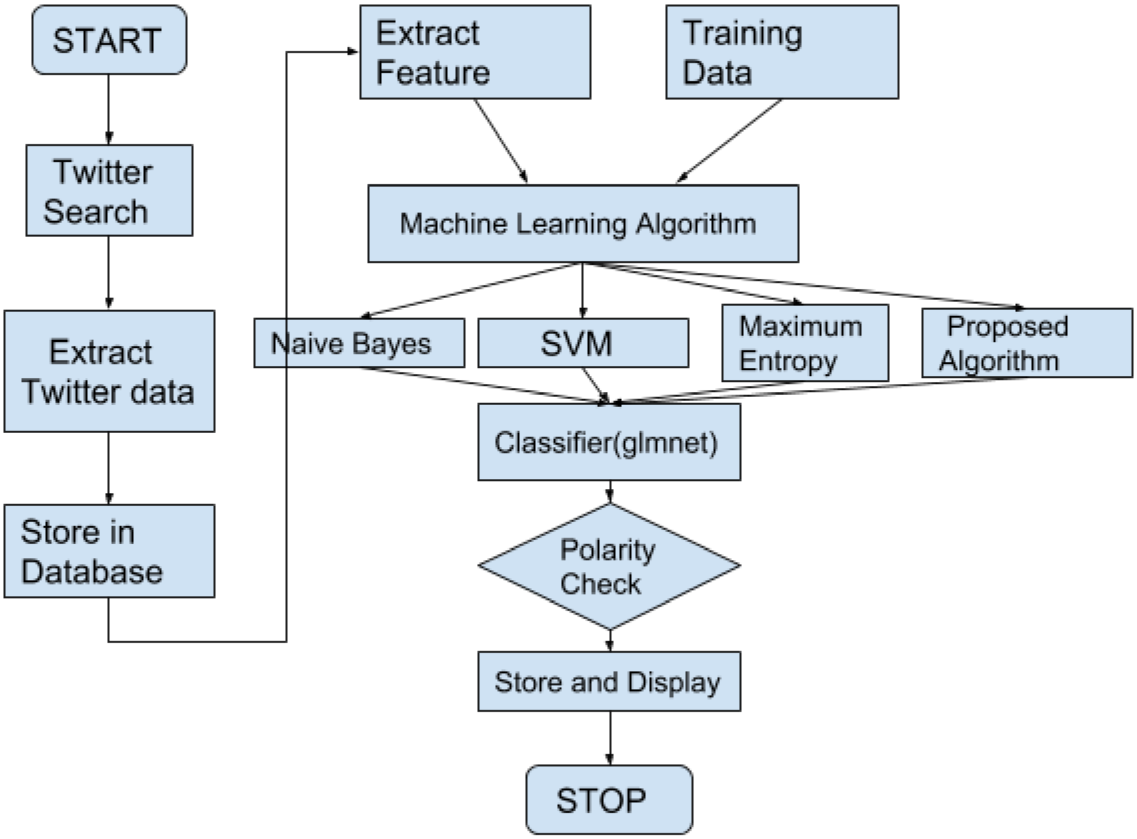
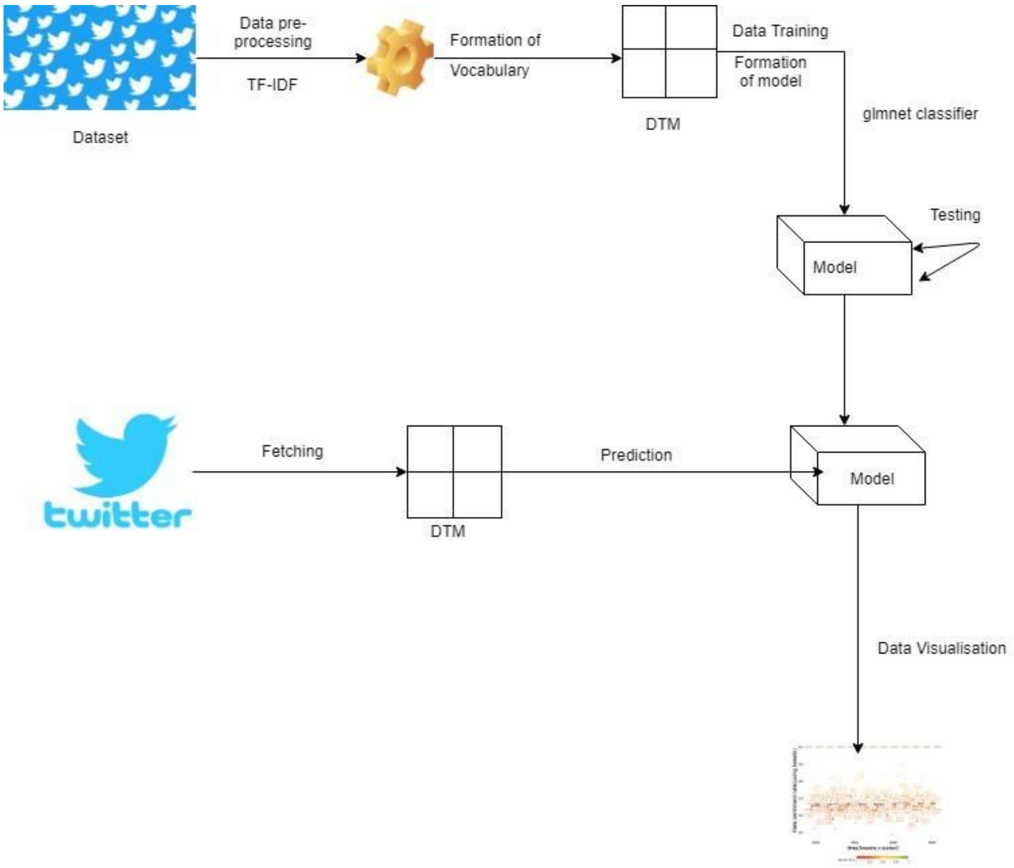


Fig 1.2: A flow diagram showing the necessary steps for completion of the proposed idea

4.1.3 Sequence Diagram/Activity diagrams



TF-IDF - Term frequency- Inverse Document Frequency
DTM - Document Term Matrix

Fig 1.3: A descriptive flow of information via flow diagram

4.2 Implementation details and issues

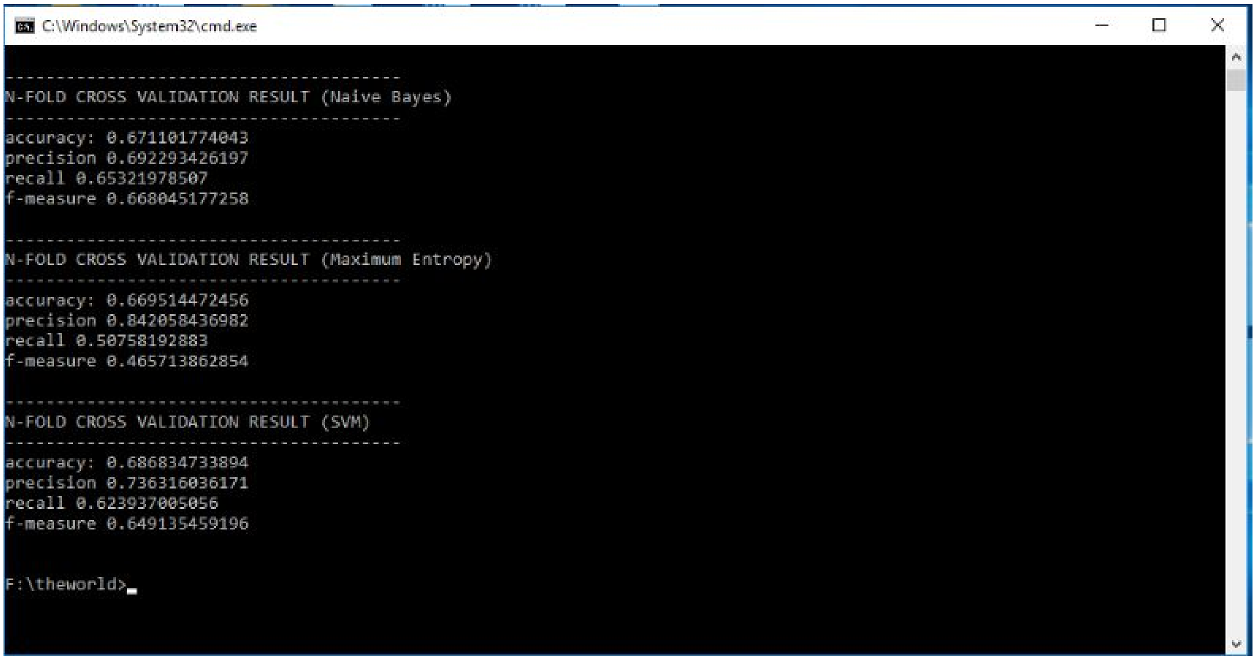
Our experiment is banked on two datasets. The source of the first dataset is CrowdFlower that contains the tweets to be extracted for analysis. This dataset is made up of 20,000 classified tweets; the hate sentiment is marked by 1, while the negative sentiment (offensive but not hate) is marked by 3 and the neutral sentiment is marked by 2. The second dataset is used for the hate sentiment extraction and it is crawled from twitter.

We have used Naive Bayes, SVM and Maximum Entropy as Single fold to analyse the given dataset. The results varied according to the available tweets and the information present in them. We used 2 data sets, one for tweets containing Hate-speech and one for neutral tweets.

This time, to improve our research we have improved Naive Bayes, SVM and Maximum Entropy and used them as Multi-fold to analyse the data. The multi fold results are better than the single fold results but are still not comparable to our proposed algorithm. This has improved our research output.

To further improve our research, this time we have also analysed(used) the emoticons used in the text or tweets to analyse the sentiments of the tweets. We have used 2 algorithms SVC and linear- SVC to analyse the text(along with the emoticons) to determine the sentiments. We have used final_tweets.csv as dataset which sourced from crowdflower.

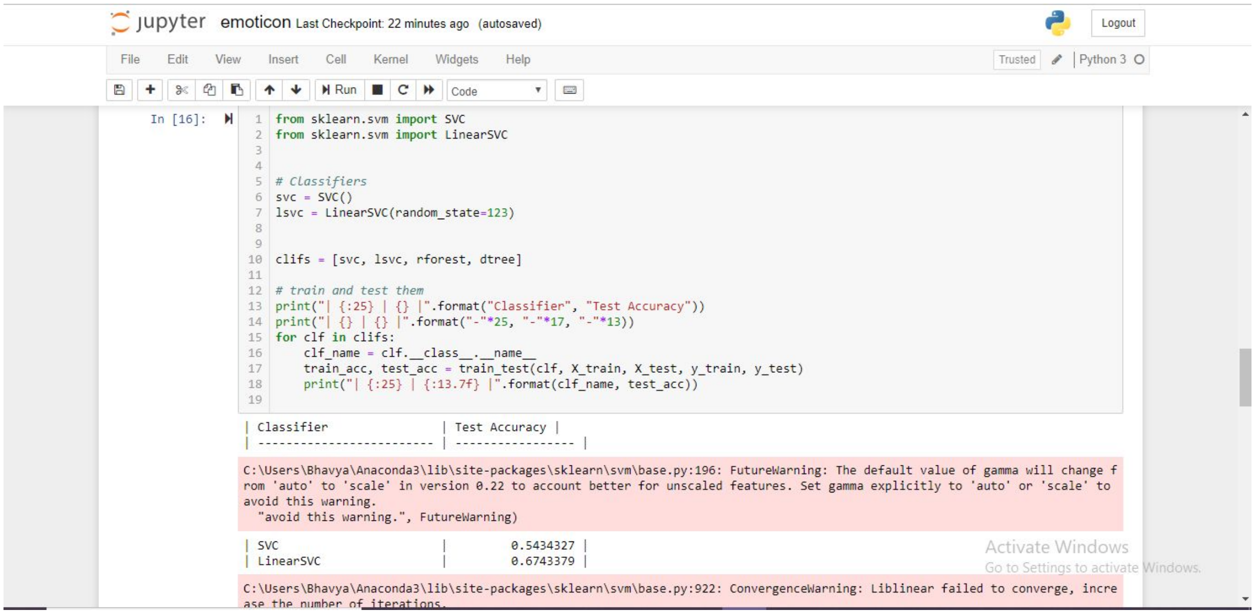
Pictorial Representation of Naive Bayes, SVM and Maximum Entropy Classifier



```
---------------------------------------
N-FOLD CROSS VALIDATION RESULT (Naive Bayes)
---------------------------------------
accuracy: 0.671101774043
precision 0.692293426197
recall 0.65321978507
f-measure 0.668045177258

---------------------------------------
N-FOLD CROSS VALIDATION RESULT (Maximum Entropy)
---------------------------------------
accuracy: 0.669514472456
precision 0.842058436982
recall 0.50758192883
f-measure 0.465713862854

---------------------------------------
N-FOLD CROSS VALIDATION RESULT (SVM)
---------------------------------------
accuracy: 0.686834733894
precision 0.736316036171
recall 0.623937005056
f-measure 0.649135459196


F:\theworld>_
```

Figure 2.1 Result of the classifiers with n-fold cross validation technique

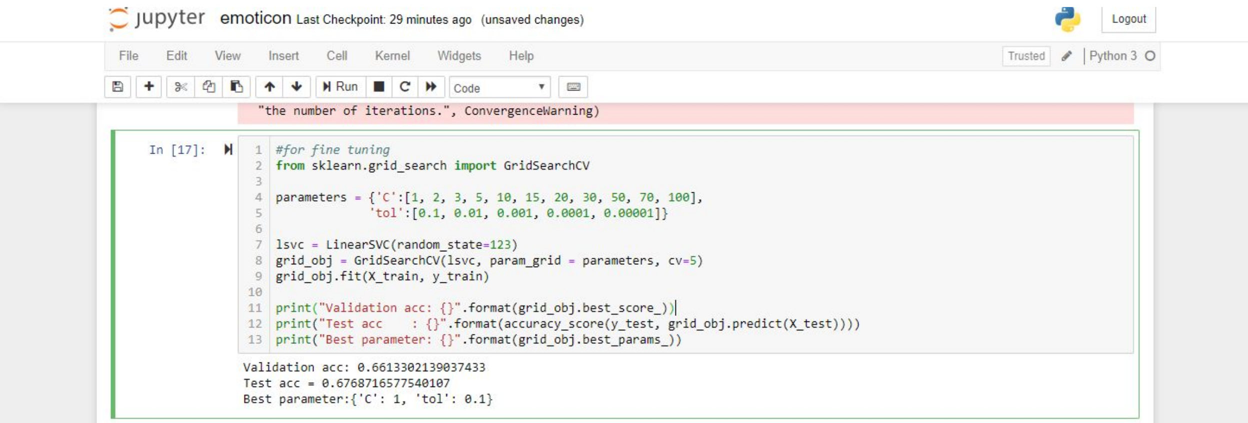
Pictorial Representation Of Emoji Based Classification Results



```python
from sklearn.svm import SVC
from sklearn.svm import LinearSVC


# Classifiers
svc = SVC()
lsvc = LinearSVC(random_state=123)


clifs = [svc, lsvc, rforest, dtree]

# train and test them
print("| {:25} | {} |".format("Classifier", "Test Accuracy"))
print("| {} | {} |".format("-"*25, "-"*17, "-"*13))
for clf in clifs:
    clf_name = clf.__class__.__name__
    train_acc, test_acc = train_test(clf, X_train, X_test, y_train, y_test)
    print("| {:25} | {:13.7f} |".format(clf_name, test_acc))

```

```
| Classifier                | Test Accuracy |
| ------------------------- | ----------------- |
C:\Users\Bhavya\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma will change f
rom 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly to 'auto' or 'scale' to
avoid this warning.
  "avoid this warning.", FutureWarning)
| SVC                       |     0.5434327 |
| LinearSVC                 |     0.6743379 |
C:\Users\Bhavya\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Liblinear failed to converge, incre
ase the number of iterations.
```

Figure 2.2 Result Of SVC and LinearSVC classifiers



```
  "the number of iterations.", ConvergenceWarning)
```

```python
#for fine tuning
from sklearn.grid_search import GridSearchCV

parameters = {'C':[1, 2, 3, 5, 10, 15, 20, 30, 50, 70, 100],
              'tol':[0.1, 0.01, 0.001, 0.0001, 0.00001]}

lsvc = LinearSVC(random_state=123)
grid_obj = GridSearchCV(lsvc, param_grid = parameters, cv=5)
grid_obj.fit(X_train, y_train)

print("Validation acc: {}".format(grid_obj.best_score_))
print("Test acc    : {}".format(accuracy_score(y_test, grid_obj.predict(X_test))))
print("Best parameter: {}".format(grid_obj.best_params_))
```

```
Validation acc: 0.6613302139037433
Test acc = 0.6768716577540107
Best parameter:{'C': 1, 'tol': 0.1}
```

Figure 2.3 Fine Tuning Of Accuracy

4.3 Risk Analysis and Mitigation

Table 1

RISK ID	Classification Of Risk	Description Of Risk	Risk Area	Probability	Impact	RE (P*I)
1	User Authorizati	User may not be	Tweet Extraction	0.5	3	0.15

	on	authorized for extracting tweets				
2	Security	Personal tweets extraction	Tweet Extraction	0.5	3	0.15
3	Network Connectivity	User may not be connected to the network	Network	0.5	1	0.5
4	Configuration Risk	Wrong query is extracted	Tweet Analysis	0.9	5	0.45
5	Data Format Risk	Tweets in invalid format are extracted	Sentiment Analysis	0.5	0.5	0.25

Table 2

SNO	Risk Area	No of Risk Statements	Weights(In+ Out)	Total Weight	Priority
1	Configuration Risk	1	3+9+1	13	1
2	Security	1	3+9	12	2
3	User Authorization	1	1+9	10	3
4	Data Format Risk	1	9+1	10	4
5	Network Connectivity	1	1+3+3	7	5

Table 3

Risk Id	Risk Statement	Risk Area	Priority Of Risk Area In IG
1	Risk of writing query	Configuration risk	1
2	Risk of feeding tweets with many emoticons	Data Format Risk	4

CHAPTER-5 TESTING

5.1 Testing Plan

Type of Test	Will test be performed?	Comment/ Explanation	Software Component
Requirement Testing	No	No prerequisites to be checked for the two platforms	N.A
Unit	Yes	Needed to perform for testing modules against detailed design	Python
Integration	Yes	To allow us to quickly identify breaking changes in the code and help with bug fixation.	Python
Performance	Yes	done to test the run-time performance of the software within the context of integrated system	All components
System	Yes	performed to identify all possible issues/bugs before releasing the product to everyday users or public	All components
Stress	Yes	Will be required	All components
Security	Yes	Will be required	Tweet Extraction

5.2 Component decomposition and type of testing required

UNIT TESTING:

Unit testing is performed for testing modules against detailed design. Inputs to the process are usually compiled modules from the coding process. Each module is assembled into a larger unit during the unit testing process. Testing has been performed on each phase of project design and coding carried out to ensure the proper flow of information into and out of the program unit while testing.

SYSTEM TESTING:

Alpha testing is a type of acceptance testing; performed to identify all possible issues/bugs before releasing the product to everyday users or public. The focus of this testing is to simulate real users by using blackbox and whitebox techniques. The aim is to carry out the tasks that a typical user might perform. The project is alpha tested and the accuracy is checked. To put it as simple as possible, this kind of testing is called alpha only because it is done early on, near the end of the development of the software, and before beta testing.

PERFORMANCE TESTING:

It is done to test the run-time performance of the software within the context of integrated system. These tests out throughout the testing process.

TEST SCHEDULE

Software Items:
Python
R Studio
JUnit Testing

Hardware Items:
Laptop
Internet Connection

Sno	List of Components that require Testing	Type of Testing	Technique for writing Test Cases
1	Dataset cleaning	Requirement	Black Box-Equivalent Partition and Boundary Value

2	Tweet Extraction in R	Requirement	Black Box-Equivalent Partition and Boundary Value
3	Sentiment Analysis	Performance	White Box
4	Emoji Extraction	Performance	White Box

5.3 List all test cases in prescribed format

TestCaseID	Input	Expected Output	Status
1	Wrong twitter handle	Error	Fail
2	Right twitter handle	Analysis	Pass
3	Right keyword	Analysis	Pass
4	Wrong keyword	Error	Fail
5	Right emoji	Analysis	Pass

5.4 Error and Exception Handling

In this age of modern development the Integrated Development Environment (IDE) has become an essential tool of productive developer. There are many advantages to using an IDE and one of the biggest is an integrated debugger. An integrated debugger allows a developer to execute their code and then inspect it as it executes.

Sometimes you can’t use an IDE to debug your code. In these instances it is very useful to know about the pdb module that is built into the standard python library.

TestCaseID	Test Case	Debugging Technique
1.	Wrong twitter handle	Print debugging
2.	Wrong keyword	Print debugging

5.5 Limitations of the solution

The system we designed is used to find out the negativity in the tweets by the public on the general trends. The results are not much accurate and the classification algorithms are not improvised.

CHAPTER-6 FINDINGS, CONCLUSION AND FUTURE WORK

6.1 Experimental Outcomes

GRAPHICAL REPRESENTATION OF HATE SPEECH:

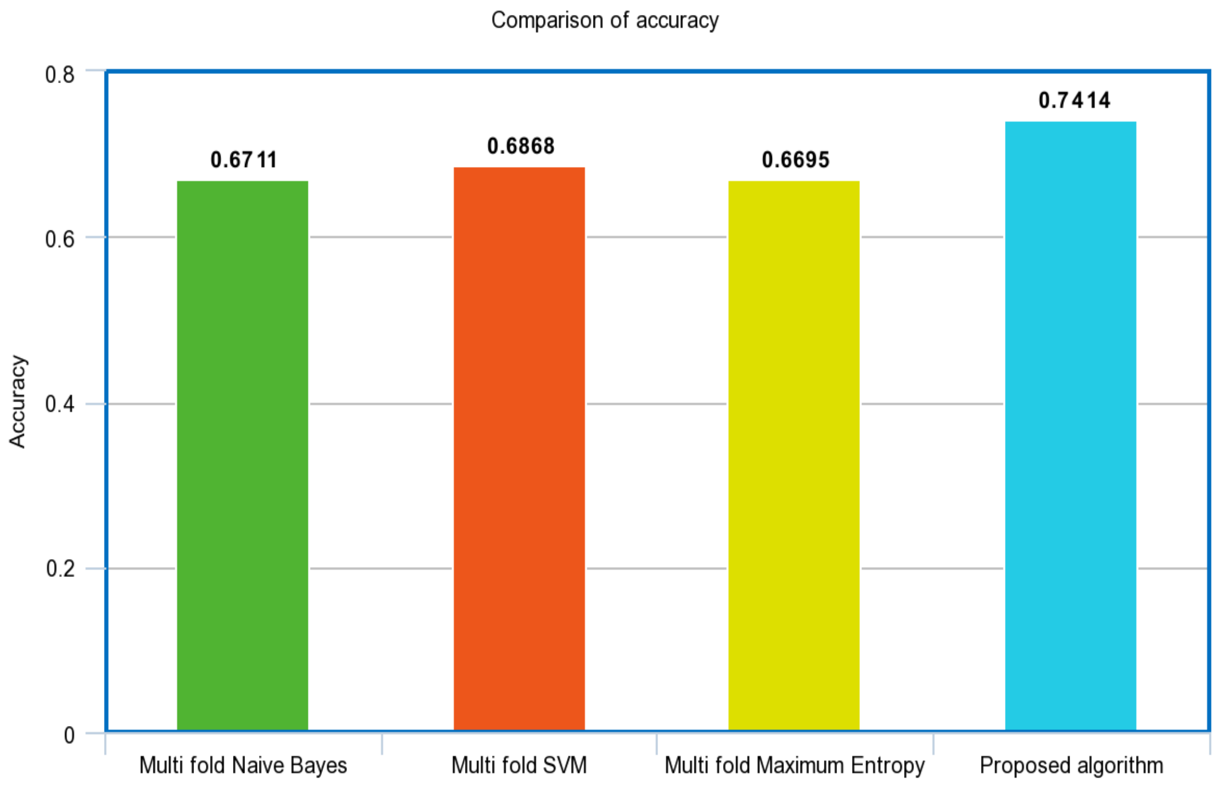


Figure 6.1.1 Bar Graph Chart

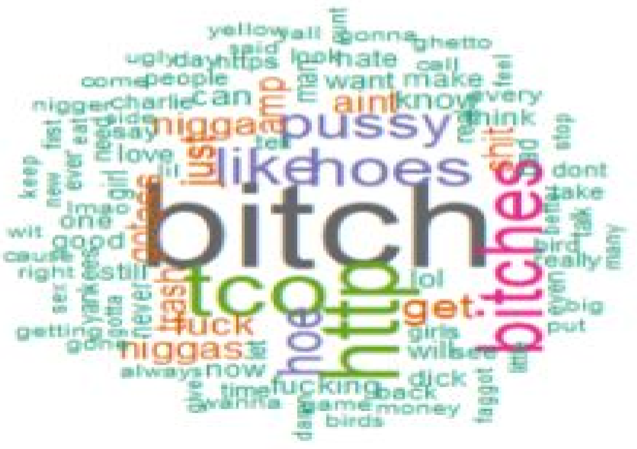


Figure 6.1.2 Word Cloud Of Hate Tweets of the dataset (1000 tweets) providing us the words with maximum frequency.

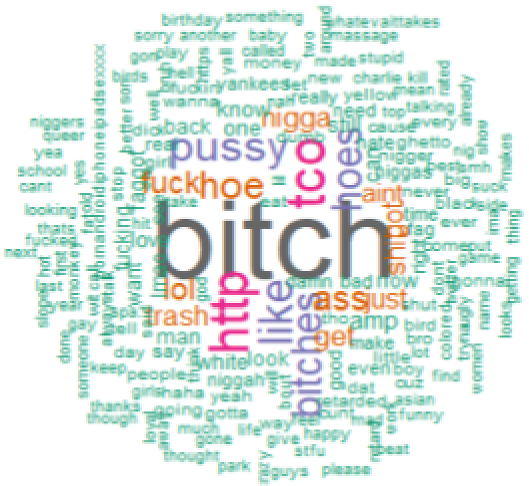


Figure 6.1.3 WordCloud Of Hate Tweets of the dataset (2000 tweets) providing us the words with maximum frequency.

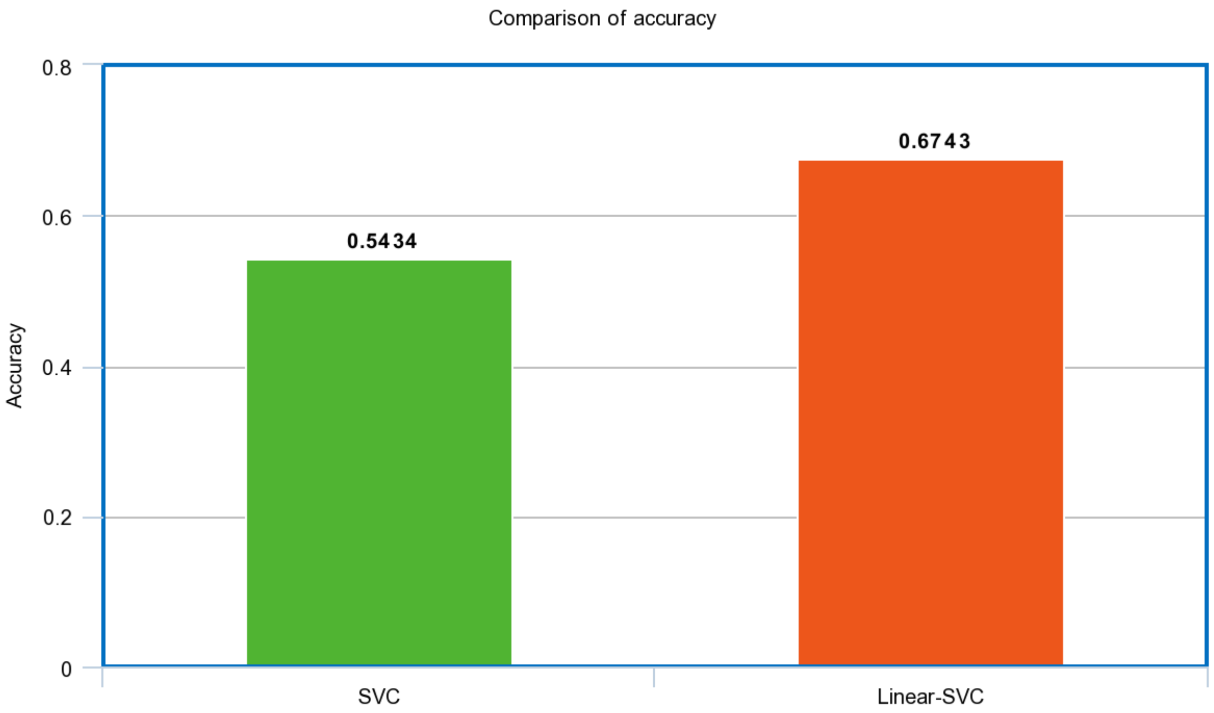
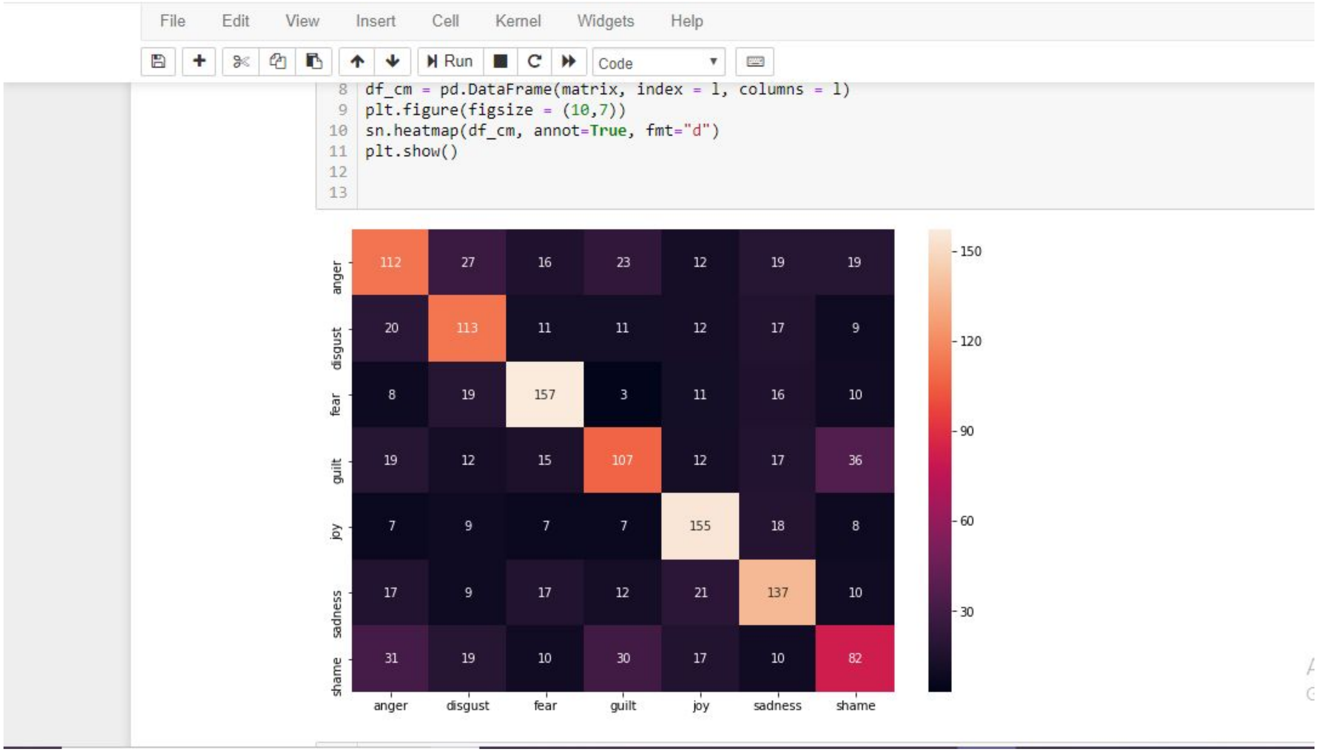


Figure 6.1.4 Bar Graph Chart



```
df_cm = pd.DataFrame(matrix, index = l, columns = l)
plt.figure(figsize = (10,7))
sn.heatmap(df_cm, annot=True, fmt="d")
plt.show()
```

	anger	disgust	fear	guilt	joy	sadness	shame
anger	112	27	16	23	12	19	19
disgust	20	113	11	11	12	17	9
fear	8	19	157	3	11	16	10
guilt	19	12	15	107	12	17	36
joy	7	9	7	7	155	18	8
sadness	17	9	17	12	21	137	10
shame	31	19	10	30	17	10	82

Figure 6.1.5 Pictorial Representation Of Confusion Matrix made using LinearSVC

6.2 Findings

Sentiment computing brings some new application opportunities and technique challenges in artificial intelligence of the next generation, and it has become a fascinating research field. Sentiment analysis is a type of text research aka mining. It applies a mix of statistics, natural language processing (NLP), and machine learning to identify and extract subjective information from text files, for instance, a reviewer's feelings, thoughts, judgments, or assessments about a particular topic, event, or a company and its activities as mentioned above. This analysis type is also known as opinion mining (with a focus on extraction) or affective rating. Some specialists use the terms sentiment classification and extraction as well. Regardless of the name, the goal of sentiment analysis is the same: to know a user or audience opinion on a target object by analyzing a vast amount of text from various sources.

Information contained in Social media, which became one of the major types of communication, make them an attractive source of data for sentiment analysis. Not only texts but also emoticons, which represent linguistic elements typically used on social media to elicit a given message, can be used to boost sentiment analysis.

The precision of recognizing emotions can increase and improve with the analysis of emojis. They provide a crucial piece of information and this is essential for companies in order to better understand their customers feelings. The way people communicate is constantly changing thanks to technology; in order to understand what they are saying, it is important that companies adjust the way they are listening to suit these changes, and this means taking also emojis into account.

6.3 Conclusion

The project can be further taken on a large scale basis for social sites like facebook or instagram to keep a track of their users who constantly posts antagonistic stuff on the site. The users who demean certain community is a national level concern and posting on social sites can be detrimental and so the site can block or deactivate such user accounts.

6.4 Future Work

From future perspective, we would like to extend this project by implementing some machine learning algorithms for applications like election results, product ratings, movies' outcomes and running the project on clusters to expand its functionalities. Moreover, we would like to make a web application for users to input keywords and get analyzed results instantly. In this project, we have worked only with unigram models, but we would like to extend it to bigram and further which will increase linkage between the data and provide accurate sentiment analysis results.

Computation of overall tweet score is done for a single keyword which can provide an overall sentiment of public regarding a topic which will be a great asset for cyber cell and government officials to cover the reportson a large database.

On the other hand, this information can be used by the customers as testimonials by extracting the strengths and weaknesses of the distinguishable features of each product, as well as finding the satisfaction levels of other users of those products. Besides the benefits in entrepreneurship, an analysis of political pages provides information to political parties regarding people’s view of their programmes. Social organisations may seek people’s opinion on current debates or on matters like the next presidential candidate. This information can be obtained by analysing the sentiment orientation of comments, the number of likes, shares or comments on posted topics.

REFERENCES

[1] http://www.methodsandtools.com/tools/staruml.php
[2] https://www.meta-chart.com/
[3] A. Pak, P. Paroubek, Twitter as a Corpus for Sentiment Analysis and Opinion Mining,
“LREC” 2010, Vol. 10.
[4] Boiy, E., Hens, P., Deschacht, K. &amp;Moens, M.-F. (2007), “Automatic Sentiment Analysis in OnLine Text”. In Proceedings of the Conference on Electronic Publishing (ELPUB-2007), p.349-360.
[5] Jerome Friedman ,Trevor Hastie, Rob Tibshirani [Department of Statistics, Stanford University] (2009), “Regularization Paths for Generalized Linear Models via Coordinate Descent” , p. 4-9
[6] L. Barbosa, J. Feng, Robust Sentiment Detection on Twitter from Biased and Noisy
Data [in:] Proceedings of the 23rd International Conference on Computational Linguistics:
Posters, Association for Computational Linguistics, Beijing 2010, pp. 36-44.
[7] Pak A, Paroubek P. Twitter as a corpus for sentiment analysis and opinion mining. Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10); 2010 May 19-21.
[8] Rao NP, Srinivas SN, Prashanth CM. Real time opinion mining of Twitter Data. International Journal of Computer Science and Information Technologies. 2015; 6(3):2923–7.
[9] S. Kim, E. Hovy, Determining the Sentiment of Opinions [in:] COLING '04 Proceedings
of the 20th international conference on Computational Linguistics, Geneva 2004.
Xing Fang, Justin Zhan, “Sentiment analysis using product review data , Journal of Big Data, 16 June 2015.
[10] Sharma Y, Mangat V, Mandeep K. Sentiment analysis and opinion mining. International Journal of Soft Computing and Artificial Intelligence. 2015 May; 3(1).
[11] Shruti Kohli,HimaniSingal, “Data Analysis with R”, 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing T. R. Foundation, "R: What is R?,".
[12] Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. Automated Hate Speech Detection and the Problem of Offensive Language.&quot; Proceedings of the 11th International Conference on Web and Social Media (ICWSM).
[13] Trevor Hastie , Junyang Qian (2014), “GLMNET VIGNETTE”, In Proceedings of the conference on Computer Science publishing Stanford University.

BIO-DATA/ RESUME OF THE STUDENTS

JASLEEN DHILLON

668/12, Srinagar, Rani Bagh, Delhi-110034
+91- 8750407770 | dhillonjasleen97@gmail.com

Objective sochna h

EDUCATION

QUALIFICATION	SCHOOL/COLLEGE	UNIVERSITY/BOARD	SCORE	YEAR OF PASSING
B.tech(Computer Science and Engineering)	Jaypee Institute of Information Technology, Noida	Jaypee University	7.7(10)	2019
12th	D.L.D.A.V Model School	CBSE	92.8%	2015
10th	D.L.D.A.V Model School	CBSE	10(10)	2013

WORK EXPERIENCE

- NEXTAGE SKILLSETTERS (Intern, Art and Crafts Department) – June 2018- Present
- Centre for Railway Information Systems, New Delhi (Summer Intern) – May 2018-July 2018
- Worked on developer’s end on two modules of their ongoing TDMS project
- CRUX-Coding Blocks (Student Training) - January 2017- May 2017
- Learned basic and Intermediate Java concepts and implementation

PROJECTS

- PlayMoods - Built a mood recognition music app and handled the machine learning part of the project and the database for the functioning.
- Course of Literature- Now and Then (Wrote this paper in 2016 and presented in my college Literature Seminar).

SKILLS

- Technical Skills- Java, MySQL, HTML, CSS, Javascript, C++
- Behavioral Skills- Attention to Detail, Keen Listening, Asking Questions, Creative thinking
- Managerial Skills- Team Management, Importance of Deadlines, Goal Setting and meeting, Multi Tasking

EXTRA-CIRRUCULAR

- Organizing Team Member of Students Seminar on Forms & Experimentations in Contemporary Literature – November 2016
- It’s Your Earth Welfare Association (JIIT NGO) – September 2015 - October 2016
- Event management of International Conference on “Peaceful and Prosperous South Asia-Opportunities and Challenges (ICSA)-March 2017
- Technical Management of Ninth International Conference on Contemporary Computing (IC3)-August 2016

- Member of KHAPPIS(Bhangra Troupe of JIIT-62)

INTERESTS

- Eye for art- Do hand lettering in my free time
- Run an Art Account for hand lettering- Helps with client interaction and time management
- Great love for football

Varn Gupta

varngupta@gmail.com | +91-8700325452
Jaypee Institute of Information Technology, Noida, India

EDUCATION

Bachelor of Technology | 2015 - 2019
Jaypee Institute of Information Technology, Noida, India
Computer Science and Engineering | CGPA - 8.0/10.0

12th Grade CBSE | 2014 - 2015
Abhinav Public School, New Delhi, India | Percentage - 89.6

10th Grade CBSE | 2012 - 2013
St. Joseph’s Convent School, Bathinda, India | CGPA - 10/10

TRAINING AND INTERNSHIPS

1. **National Fertilizer Ltd. (NFL)** | Jun’ 18 - Jul’ 18
Web Developer | Bathinda, Punjab
Tech stack: HTML5, CSS3, Bootstrap 3, JavaScript, JQuery

2. **Sun-Soft Technologies** | Jun’ 17 - Jul’ 17
Java Developer | Bathinda, Punjab
Tech stack: J2SE, JavaFX

3. **Sun-Soft Technologies** | Jun’ 16 - Jul’ 16
Web Developer | Bathinda, Punjab
Tech stack: PHP, SQL, HTML5, CSS3, JQuery

4. **Hindupost.in**
Published Writer | Gurugram
Writes about politics, human rights, and Indian history

SKILLS

- **Technical:**
  R Java SQL
  PHP JavaScript CSS
  HTML C++ C

AWARDS/ACHIEVEMENTS

- **Adjudicated** 12th annual parliamentary debate at IIT Delhi
- **Adjudicated** socio - technical debate at JIIT
- **Won** Manthan organised by University Grants Commission (UGC)

LEADERSHIP EXPERIENCE

- **Coordinator** at Jaypee Debsoc, Jul’ 17 - Jun’ 18
- **Coordinator** at Google Developers Group, Jul’ 16 - Jun’ 17
- **Manager** at It’s Your Earth, Oct’ 15 - Jun’ 16

PROJECTS

1. **Playmoods** | Aug’ 17 - Dec’ 17
Developed an application for recognizing and mapping the genre of a song and the mood of a person for a good musical experience.
Tech stack: Python, HTML5, CSS3

2. **Blood Donation Website** | Jan’ 17 - May’ 16
Designed and developed a web application for bringing together blood donors and people in need.
Tech stack: PHP, SQL, HTML5, CSS3, JQuery

CO-CURRICULAR ACTIVITIES

- **Debates** - 8 Parliamentary debates, 2 Model United Nation Conferences, 4 Conventional Debates
- **National Seminar Paper** on customer relationship and fake news