IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

EMOTION RECOGNITION FROM TEXT USING MACHINE LEARNING

[1]Goru Swathi, [2]Behara Meghana Patnaik, [3]Arjala Janani, [4]Kanithi Karthik, [5]Janni Divya, M. Jayanthi Rao[1]*

Department of Computer Science and Engineering
Aditya Institute of Technology and Management, Tekkali, AP, INDIA-532201
[1]*e-mail: jayanth.mtech@gmail.com

ABSTRACT

Human emotion can be expressed across two channels which is via verbal and non-verbal. Verbal way of expression include through speech, sound and text whereas Non-Verbal way of expression include facial expressions and gestures. In today's technological world, a majority of population across the world prefer text as a common channel for sharing their opinions or emotions through social media as well as product reviews in e-commerce platforms. Emotion recognition is an affective computing development system utilized in the process of detecting, predicting human emotional state such as anger, sad, happiness etc. The main aim of this study is to develop an emotion recognition system for text-based content. All models were tested using criteria from the ISEAR (International Survey of Antecedents and Affective Reactions) dataset. In this study supervised machine learning algorithms were used to develop a model based on six basic emotions which are sadness, anger, love, fear, joy, surprise. Emotion prediction system for text-based content was successfully developed.

KEYWORDS: Text data, emotion detection, tokenization, SVM, Decision tree, Logistic Regression.

INTRODUCTION

Due to rapid progress in Information & Technology sector, people have access to any kind of information at the click of a button. Moreover, with the invent of smart-phones and 4G networks, even people from the remote areas are getting connected to Tier 1 and Tier 2 cities. With the growing population in countries like India, it has led to tremendous growth in the number of people using social networks. Communication via social media is done in the form of text, image, audio and video which contains information and consumes space, memory and Internet bandwidth. All these activities done on social media has resulted into vast amount of information being generated on a daily basis. Social media analysis has become an interesting field of research to understand the behavior and thoughts of people in response to social, economic, cultural, educational and all others activities happening around the world. This has created strong interest for research in emotion analysis or recognition. Emotion recognition has emerged as a popular solution to gain deeper insights such as knowing how happy our people are, predicting box-office revenues from movie reviews, predicting election results and stock market from Twitter data, gaining feedback before releasing a new product and visualizing of customers feel about product or service run by multinational companies and in some cases it is also used to reduce suicidal rates.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Recently, researchers have proposed various methods for text sentiment detection, including keyword-based, lexical similarity, learning-based, and hybrid models. At first, they introduced a rule-based approach consisting of two approaches: lexical affinity-based and keyword-based. Then came a new approach, the learning-based approach. This method was more accurate and gave better results. Learning-based approaches use different models for emotion recognition. Many researchers have also begun to combine and combine these approaches in search of greater accuracy. Studies have shown that no single approach provides a perfect solution for detecting sentiment from a text. Existing solutions had many limitations, such as not having a list of all emotions. Existing lists do not have enough vocabulary for words in dictionaries, ignore words, have semantic context, extract less contextual information from specific sentences, and fail to recognize specific emotions. Some models are not suitable for frequent emojis, weak semantic information extraction and sentence structure. It varies by model. This system had many limitations that were met by previous researchers. The proposed model has fulfilled many of the existing limitations. In 2001, W. Gerrod Parrot, wrote a book named "Emotions In Social Psychology", in which he explained the emotion system and formally classified the human emotions into six categories according to the emotional hierarchy: love, joy, anger, sadness, fear and surprise. In this paper, emotion recognition on textual data is analyzed by using supervised machine learning algorithms.

LITERATURE SURVEY:

Several studies have used various techniques to detect emotions from text [1–3]. It will show that which is the best model and gives us a higher accuracy. Seal et al. [2] have performed emotion detection with a keyword-based approach mainly focused on phrasal verbs. They used ISEAR data, preprocessed the data, and then applied the keyword-based approach. They discovered several phrasal verbs that should have been associated with emotion terms but were not, and so they built their own database. They recognized phrasal verbs and keywords synonymous with various emotions and categorized them using their database. They did, however, achieve a much higher accuracy of 65%, but they were unable to address the researcher's existing issues, such as an insufficient list of emotion keywords and a lack of respect for word semantics in meaning. The work by Alotaibi [3] has worked on a learning-based approach. He has used the ISEAR database for emotion detection. Then, using classifiers like Logistic Regression, K-Nearest Neighbor (KNN), XG Boost, and Support Vector Machine (SVM), he preprocessed and trained the data. According to him, all other classifiers poorly performed as compared to Logistic Regression. Finally, he said that the deep learning technique would help to improve the model. J.G.Taylor [4] has worked on psychological literature for emotion detection. In his research he studied the relationship between emotions, facial expressions and body movements. They concluded by extracting artificial neural network architecture. S.Namratha [5] has analyzed the existing research of emotion recognition and their drawbacks. They also discussed about conversion of audio input into text using speech processing tool. Soumaya Chaffar [6] has discussed about a different machine learning approach in which they used heterogeneous dataset collected from blogs, fairy tales and many more. They concluded that SMO algorithm gave significant improvement over other classification algorithms. Saima Aman [7] described the task of annotating sentences in a blog corpus with information about emotion category and intensity, as well as emotion indicators. They planned to use a corpus-driven approach in building a lexicon of emotion words. They intended to start with the set of emotion indicators identified during the annotation process, and further extend that using similarity measures. Jamil Hussain [8] had made a comparison among SVM, NB and ME classifiers regarding sentence level sentiment analysis for depression measurement. They have adopted

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

voting model and feature selection technique. Their experiment indicates that SVM shows superior result as compare to Nave Bayes and Maximum Entropy classifiers. Several studies have used different approaches to detect emotions from text [9].

RELATED WORK

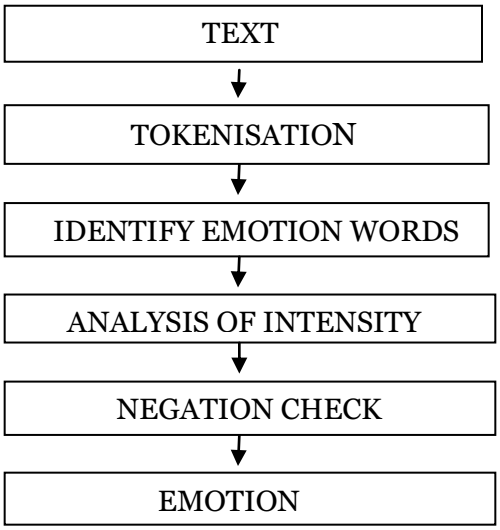
Emotion recognition has been performed over the past 30 years, and analyzing emotion is a relatively new field and has attracted researchers' attention due to the popularity of social media platforms that have created a new level of freedom of expression. Researchers have proposed various methods for sentiment analysis, such as lexical affinity approach, machine learning, or a combination of the two. Generally there are two types of writings/texts- Subjective and Objective. The subjective text consists of surface emotions and the objective text consists of hidden emotions. Subjective text is easy to work with than objective text because the latter is suggestive or completely emotionless.

Emotion is expressed as joy, sadness, anger, surprise, hate, fear and so on. Since there is not any standard emotion word hierarchy, focus is on the related research about emotion in cognitive psychology domain. In 2001, W. Gerrod Parrot, wrote a book named “Emotions In Social Psychology”, in which he explained the emotion system and formally classified the human emotions through an emotion hierarchy in six classes at primary level which are Love, Joy, Anger, Sadness, Fear and Surprise. Certain other words also fall in secondary and tertiary levels. Directions to improve the capabilities of current methods of text-based emotion detection are proposed in this paper.

Previous methods used for text based emotion recognition system are:

Keyword spotting technique:

A keyword pattern matching problem can be described as the problem of finding occurrences of a keyword in a given set as a substring in a given string. This problem has been studied in the past and algorithms have been proposed to solve it. In the context of emotion detection, this method is based on specific predefined keywords. These words fall into categories such as disgust, sadness, joy, anger, fear, and surprise.



ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

The emotion recognition keyword discovery technique consists of 5 steps that receive a text document as an input and generate the output as a sentiment class as shown in the figure. In the first step, text data is converted into tokens that identify and detect word-emotion**.** Initially this method takes some text as input and in the next step it performs tokenization of the input text. In the next step, after identifying words related to emotion, we analyze the strength of emotion words. Offer confirmed whether infidelity is involved or not**,** you will eventually find the kind of emotion you want with the desired outcome.

Method of lexical affinity:

Sentiment detection based on related keywords is a simple and easy to use method. Lexical Affinity approach is an extension of keyword arbitrary words apart from picking up emotional keywords. These possibilities are often part of the linguistic corpus, but there are also drawbacks. First, the probabilities assigned are biased toward specific body text genres. Second, it misses emotional content deeper than the word level at which this technique operates. For example the word accident, having been assigned a high probability of indicating a negative emotion, would not contribute correctly to the emotional assessment of phrases like I avoided an accident or I met my girlfriend by accident.

Learning-based Method:

Learning-based methods[10-12] are used to frame problems in a variety of ways**.** The original problem was to identify sentiment in the input text, but now the problem is to classify the input text into different sentiments**.** Unlike keyword-based detection methods, learning-based methods attempt emotion detection based on pre-trained classifiers that apply various machine learning theories, such as support vector machines and conditional random fields, to determine which emotion categories an input should belong to.

Hybrid Methods:

Keyword-based methods using naive and thesaurus-based methods did not yield satisfactory results, so some systems use hybrid approaches that combine keyword spotting techniques with learning-based methods to improve accuracy. The most important hybrid system to date is the work of Wu, Chuang, and Lin, which uses a rule-based approach to extract semantics associated with specific emotions and a Chinese dictionary ontology to extract features. These semantics and attributes are associated with emotion in the form of association rules. Consequently, these emotion-related rules that replace the original emotion keywords serve as training features for learning modules based on separable hybrid models. This method outperforms previous approaches, but categories of emotions are still limited.

PROPOSED METHODOLOGY:

In the proposed methodology we will be using Supervised machine learning classification algorithms such as Naïve Bayes, Support Vector Machine, Decision Tree ,KNN and Logistic Regression. After the classifiers are trained, text data can be fed into them to determine the emotion type. The documents were tested for emotion detection and calculated on the basis of the confusion matrix shown in table which contains the classifier's decisions in the rows,

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

and the actual decision about classification in the class in the columns. The four fields of the table contain number of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) classified document.

		ACTUAL	
		YES	NO
PREDICTED	YES	TP	FP
	NO	FN	TN

Accuracy is expressed as the proportion of correctly classified cases over all cases and is calculated according to formula:

$$\text{Accuracy} = (TP+TN) /(TP+TN+FP+FN)$$

(2) Precision is expressed as the proportion of positive cases that are correctly recognized as positive over all cases classified as positive and is calculated according to the formula:

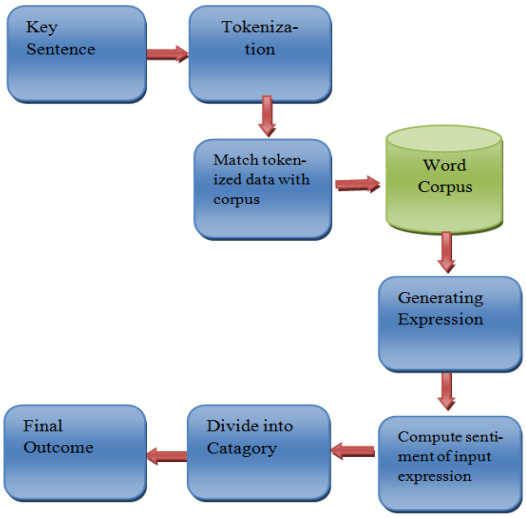
$$precision = TP /(TP+FP)$$

(3) Recall is expressed as the proportion of positive cases that are correctly recognized as positive over all actual positive cases and is calculated according to the formula:

$$\text{Recall} = TP / (TP+FN)$$

(4) A measure that combines precision and recall is called the F1-measure and represents their weighted harmonic mean. It is calculated according to the formula:

$$F1 = (2\times precision\times recall) / (precision+recall)$$



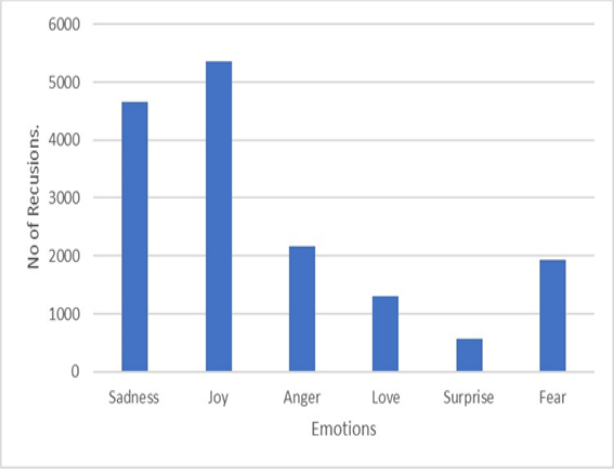
ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

DATASET:

The project used various resources such as the ISEAR (International Survey of Emotional Antecedents and Responses) dataset and the NLTK (Natural Language Toolkit) corpus.The ISEAR dataset is a sentiment labeled dataset where the data is cleaned and lightly normalized. It contains six main emotions: joy, fear, anger, sadness, love and surprise. Since ISEAR is a clean and constructed database,it has been tested to find the best classification algorithm.Then ISEAR is also evaluated further.Meanwhile, the NLTK (Natural Language Toolkits) corpus is used as a text processing library because the actual input texts often contain stop words, misspellings, etc. Therefore, unwanted text should be removed. All data sources used in this study are described in the following table

Emotion	No of Records
Sadness	4666
Fear	1937
Joy	5362
Anger	2159
Love	1304
Surprise	572



ALGORITHMS WORKED:

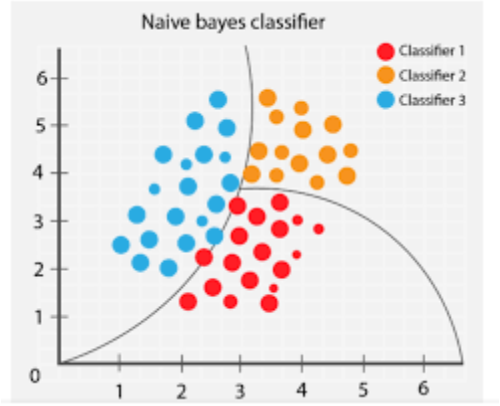
Naïve Bayes:

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in *text classification* that includes a high-dimensional training dataset. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

$$P(A/B) = \frac{P(B/A)\,P(A)}{P(B)}$$

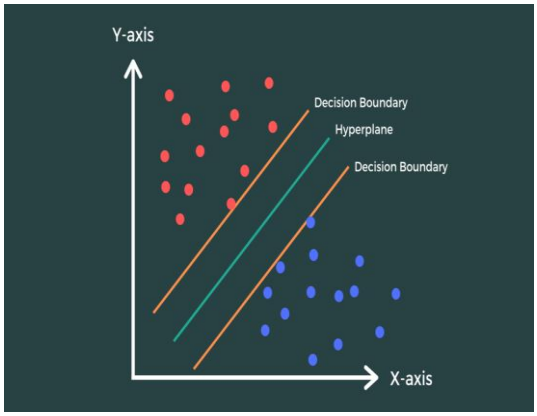
ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022



SVM:

SVM stands for Support Vector Machine and is one of the most popular supervised learning algorithms used for classification and regression. The goal of SVM is to create optimal lines or decision boundaries that can separate the n-dimensional space into classes so that new data can be easily classified into the correct category. This decision boundary is called a hyperplane. SVM algorithms can be used for face recognition, image classification, text classification, and more.

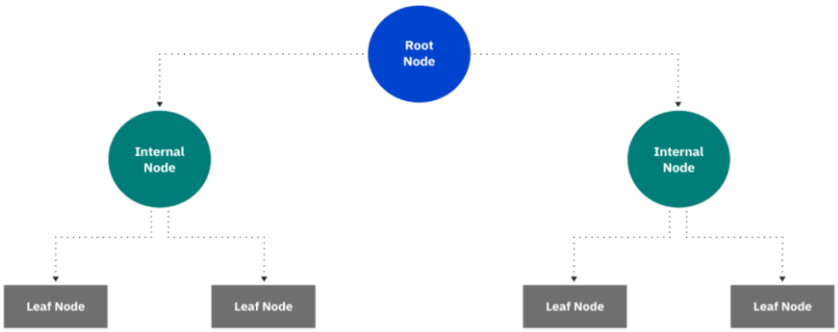


DECISION TREE:

Decision trees are a supervised learning technique that can be used for classification and regression problems. It is a tree-structured classifier where internal nodes represent dataset features, branches represent decision rules, and each leaf node represents an outcome. Decision trees are usually easier to understand because they mimic the human thinking capacity when making decisions. The logic behind decision trees is easy to understand because it shows a tree-like structure.

ISSN PRINT 2319 1775 Online 2320 7876

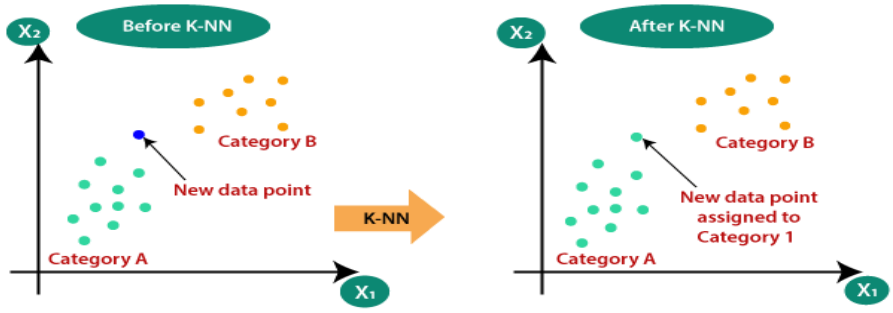
Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022



KNN:

K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.By calculating the Euclidean distance we get the nearest neighbors, and thereby classify the data to which category it belongs.

Euclidean distance= $\sqrt{(x2-x1)^2 + (y2-y1)^2}$

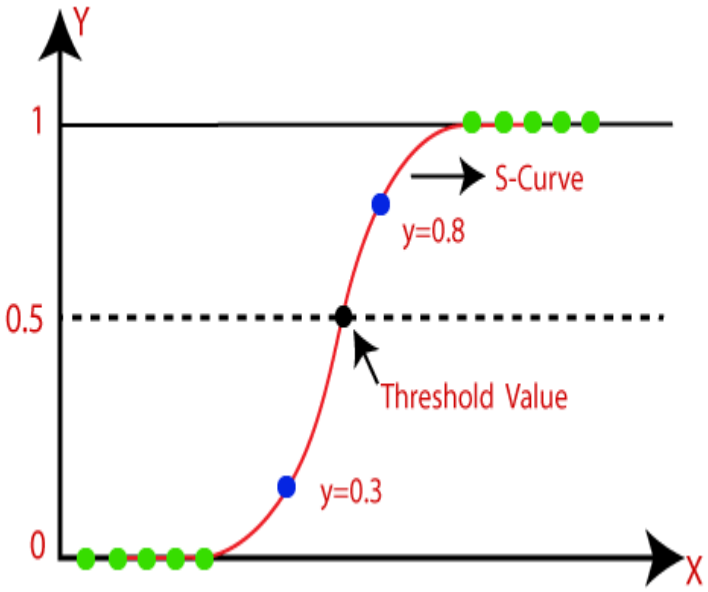


ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

LOGISTIC REGRESSION:

It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).The curve from the logistic function indicates the likelihood of something.



RESULTS AND DISCUSSION:

Over the past decade, several scientific experiments on emotion recognition have been conducted using multim odal mixtures of multimodal cues such as facial and audio gestures, audio and written language, physiological cues, and different variations of these modalities.

In the proposed multimodal emotion recognition, after reviewing previous studies, they all use more than one modality to recognize emotions by combining deep learning with different methods and techniques. Among th-em, machine learning also has a variety of algorithms, methods and architectures in terms of classification and feature extraction.

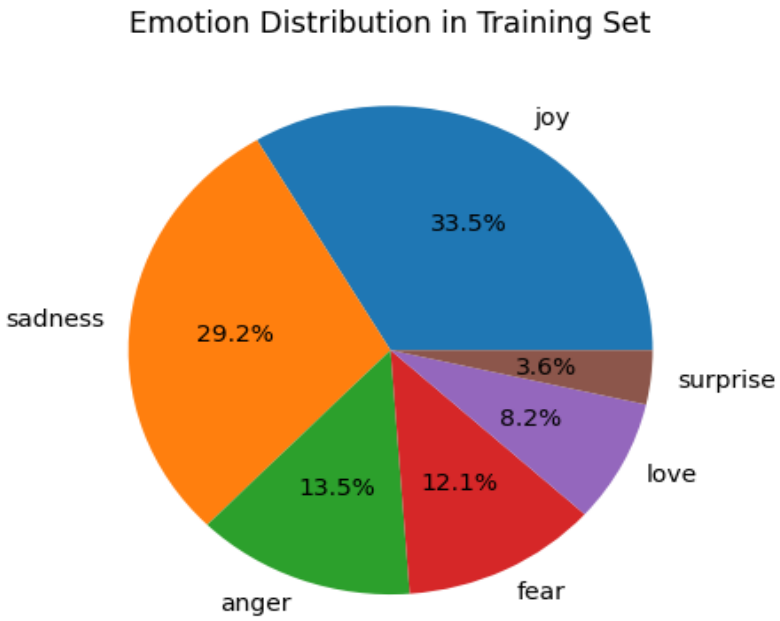
We conducted a number of experiments using different methods and used a machine learning approach to obtain the highest emotion classification accuracy for the proposed model. Here in our model we would be giving text as input where the text is processed and for training our model labels for emotions were used which

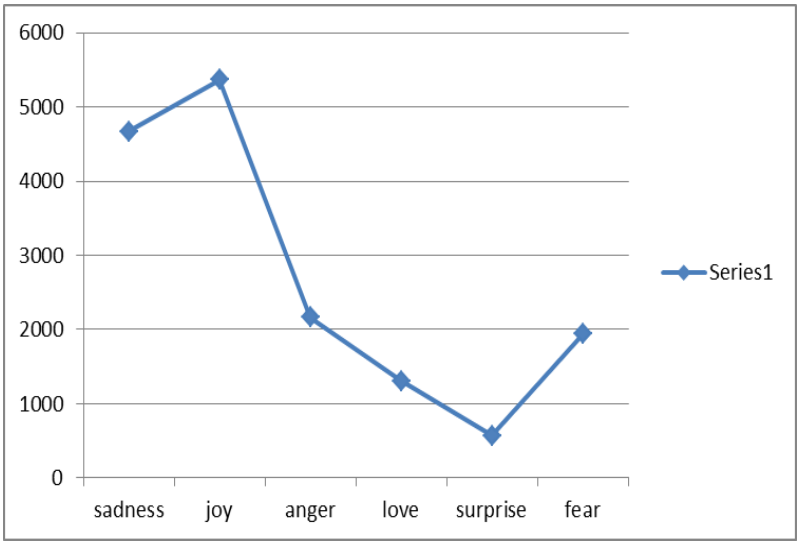
ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

would be helpful for predicting the output. In order to train our model we have used several machine learning algorithms such as SVM, NAÏVE BAYES, KNN Classifier, DECISION TREE, LOGISTIC REGRESSION. Out of all the algorithms we got one best accuracy model for our prediction.

The dataset being used in this project is from ISEAR. The graphical data representation of our dataset is as follows:

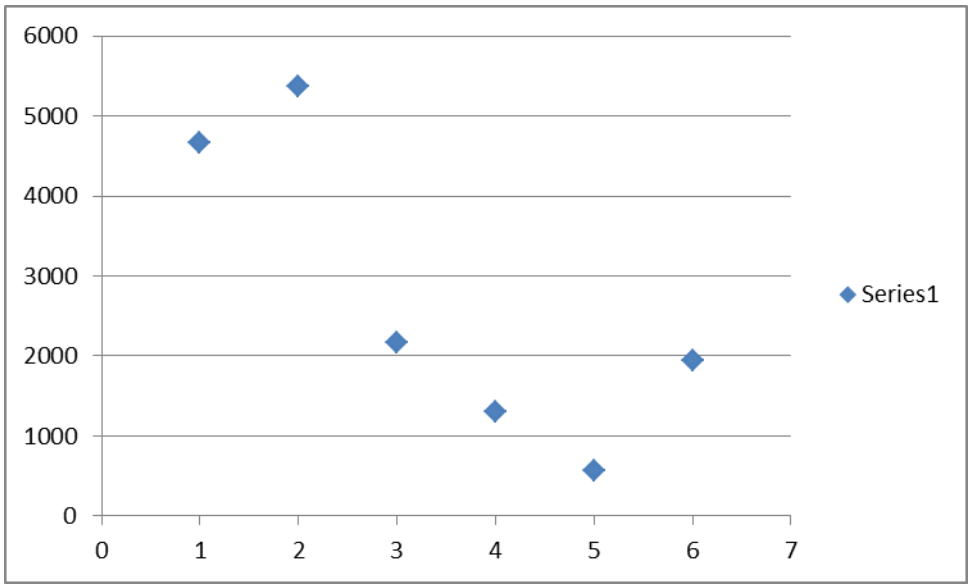




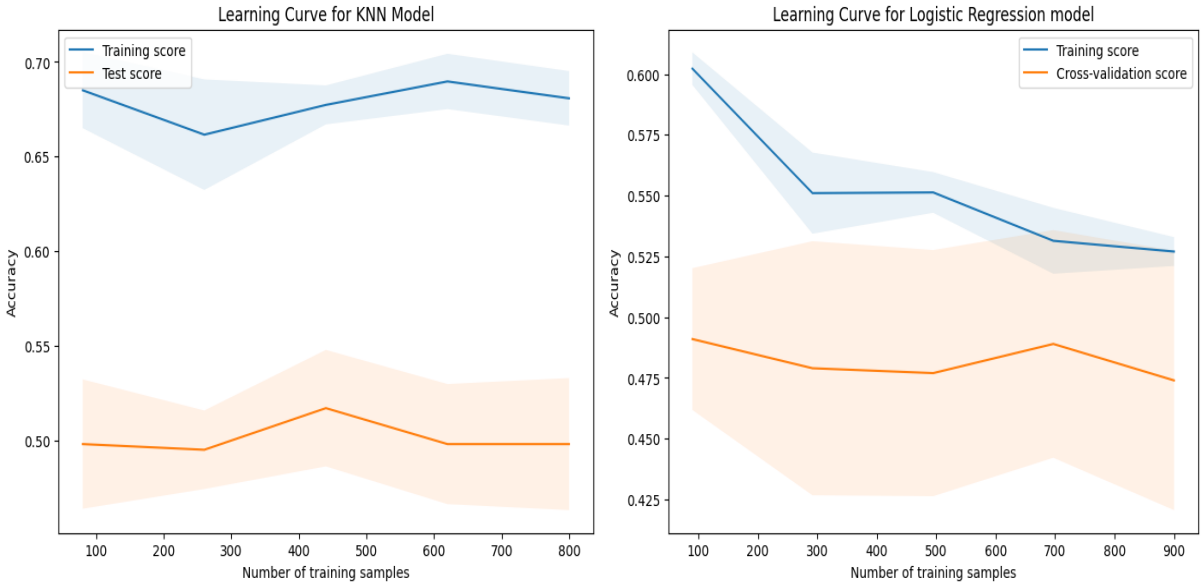
ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

As we have 6 emotions for our model these emotions are labeled numerically and scatter plot representation as follows:

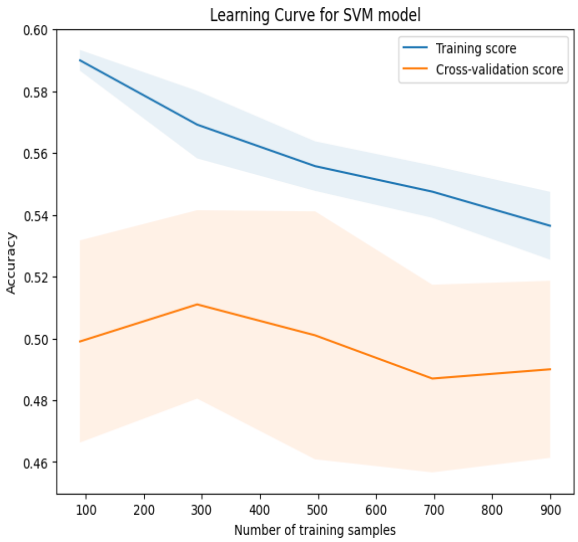


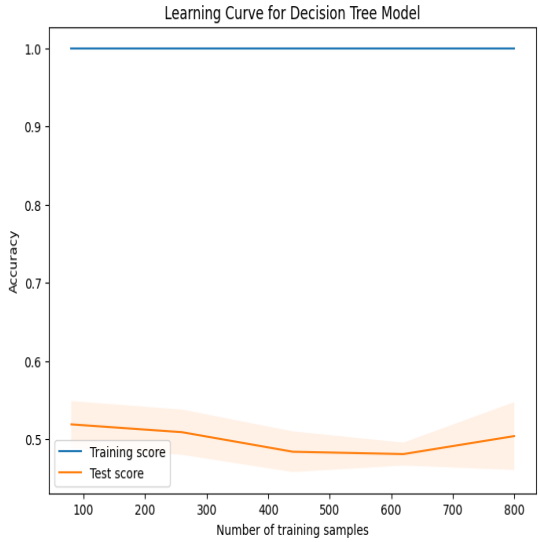
The learning curves for the algorithms used in this project are as follows:

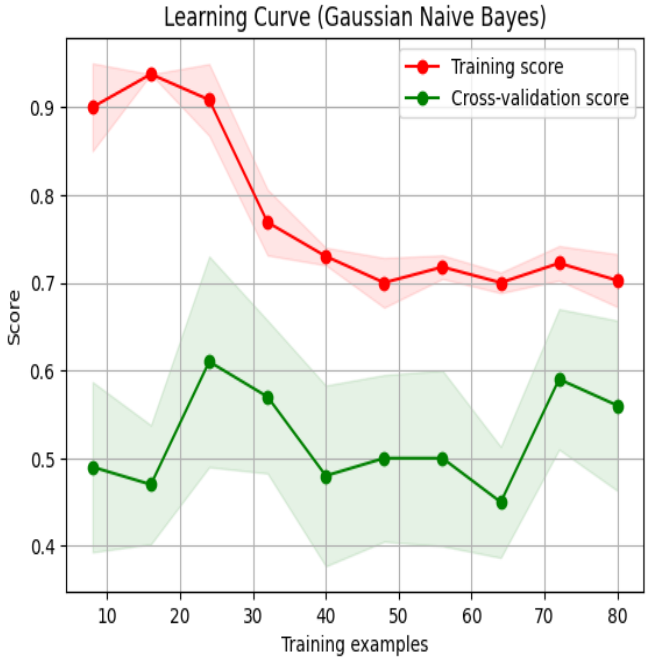


ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022







CONCLUSION:

In this paper, proposed a text-based emotion recognition model. The proposed model is a combination of machine learning approaches. According to machine learning classifier, Logistic Regression gives highest accuracy of 84% as compared to SVM, Naïve Bayes, KNN, Decision tree. Moreover, in this digital world,

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

people's usage of sending text messages, uploading tweets, and writing online reviews of products have been in great use and demand. Therefore, by having a lot of data, we can make a real-time text-based emotion recognition model to find the emotions or moods of the people.

REFERENCES:

1. C. R. Chopade, "Text based emotion recognition: a survey," International Journal of Science and Research, vol. 2, no. 6, pp. 409–414, 2015.
2. D. Seal, U. K. Roy, and R. Basak, "Sentence-level emotion detection from text based on semantic rules," Information and Communication Technology for Sustainable Development, Springer, Singapore, pp. 423–430, 2020.
3. Text-Based Emotion Recognition Using Deep Learning Approach Santosh Kumar Bharti,1 S Varadhaganapathy,2 Rajeev Kumar Gupta,3 Prashant Kumar Shukla,4 Mohamed Bouye,5 Simon Karanja Hingaa , 6 and Amena Mahmoud7.
4. N. Fragopanagos, J.G. Taylor, "Emotion recognition in human– computer interaction", Department of Mathematics, King's College, Strand, London WC2 R2LS, UK Neural Networks 18 (2005) 389–405 march 2005.
5. C. Yang, K. H.-Y. Lin and H.-H. Chen, "Emotion classification using web blog corpora," Proc. IEEEWICACM International Conference on Web Intelligence. IEEE Computer Society, Nov. 2007, pp. 275-278, doi 10.1109WI.2007.50.
6. Soumaya Chaffar and Diana Inkpen (2011) Using a Heterogeneous Dataset for Emotion Analysis in Text, Proceeding Canadian AI11 Proceedings of the 24th Canadian Conference on Advances in Artificial Intelligence (Pg. 62-76).
7. Saima Aman and Stan Szpakowicz (2007), Identifying Expressions of Emotion in Text, V. Matouek and P. Mautner (Eds.): TSD 2007, LNAI 4629, pp. 196205, 2007. Springer Verlag Berlin Heidelberg 2007.
8. Sentiment Analysis of Social Networking Sites Data using Machine Learning Approach for the Measurement of Depression, Anees Ul Hassan, Jamil Hussain, Sungyoung Lee, ICTC 2017, (Pg.-138-140).
9. C. R. Chopade, "Text based emotion recognition: a survey," International Journal of Science and Research, vol. 2, no. 6, pp. 409–414, 2015.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

10. M. Balakrishna, M. Ramanaiah, B. Ramakrishna, M. Jayanthi Rao and R. Neeraja,Inductively Coupled Plasma-Mass Spectroscopy: Machine Learning Screening Technique for Trace Elemental Concentrations in Hemidesmus Indicus. 65(1), 4431-4445, 2022.

11. M. Jayanthi Rao, P. Prasanthi, P. Suresh Patnaik, M. Divya, J. Sureshkumar and M.Ramanaiah, Forecasting systems for heart disease using advanced machine learning algorithms. Int. J. Food and Nut. Sci., 11(7), 1257-1268, 2022.

12. M. Jayanthi Rao, M. Divya, M. Ratnan Mohitha, P. Prasanthi, S. Paparao, M.Ramanaiah, Analyzing the effectiveness of convolutional neural networks and recurrent neural networks for recognizing facial expression. Int. J. Food and Nut. Sci., 11(7),1269-1282, 2022.