International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 05 | May 2018 www.irjet.net p-ISSN: 2395-0072

An Integrated Framework to Identify Different Psychological Behaviour of a Human Being

Rajatha[1], Nirosha [2], Preemal Varsha Pinto[3], Pujitha J Shetty[4], Rithika M R[5]

[1]*Assistant Professor, Dept. of Information Science Engineering, Sahyadri College of Engineering and Management, Karnataka, India*
[2,3,4,5]*Student, Dept. of Information Science Engineering, Sahyadri College of Engineering and Management, Karnataka, India*

Abstract - *The project is about how a system is able to identify different psychological behaviours of a person such as whether the person is likely to be a positive minded or negative minded based on the feedback he gives in the form of text data. A person's way of dealing with the world differs as different people have different opinions on some certain events. Humans have tendency to take biased decisions based on their respective relationship with a person whereas machines have the capability to take unbiased decisions irrespective of their relationship with the person. The project is based on the concept of machine learning for the system to take decision on its own and text mining for classifying a pre-defined set of data. The project focuses towards designing an effective framework to achieve classification of text documents and make use of some pattern recognition methods to analyze a person's behaviour.*

Key Words: **Psychological behaviour, Pattern recognition, Machine learning.**

1.INTRODUCTION

Artificial Intelligence(AI) is the intelligence exhibited by machines. It is the machine that is being trained to function in an intelligent manner just like how humans think and take decisions. Alphago, a computer program which is one of the example for artificial intelligence was able to defeat humans in board games. There are various other examples for artificial intelligence like in gmail the automatic replies are generated based on the content of the mail. There are many sub domains AI such as machine learning, natural language processing etc. Machine learning field is broadly used and it has an ability to train machines in such a way that they can make judgement or take decisions on their own without the help of a human. Automatic text processing is a well known research theme in the domain of machine learning. Text Processing is a vast area which contains many branches like text classification, text clustering, text classification, text compression, information retrieval, document indexing, opinion mining, document summarization, plagiarism detection, psychological behaviors. Some of the applications of the text processing are : Text classifications for news agencies, Anti-spam filtering, Sentimental analysis, Indexing of scientific journal, Topic detection and tracking of journals. In earlier days it was all based on knowledge engineering i.e, the rules for the classifier which is used for classification of text files were written manually by the experts. But, nowadays automatic document management system is being introduced. Some problems such as managing unstructured data and handling large volumes of data still remain. Hence the machine learning and text processing algorithms can be combined together to solve such problems. One of the application of this combination is that it can be used to analyze a person's behaviour based on the text input a person gives. Analysing a person's behaviour has become a major task. A person's behaviour may change based on the respective situation and people he is dealing with. Hence it is difficult to understand the actual behaviour of a person. For this purpose a machine can trained using some of the machine learning and text mining techniques to identify the actual behaviour of a person which says whether he is positive minded or negative minded based on the feedback he gives for different situation in the form of text document. The machine is being trained by giving pre-defined set of data to get appropriate results.

2.PROBLEM STATEMENT

"An Integrated Framework to Identify Different Psychological Behaviour of a Human Being" focuses on developing a machine learning approach that identifies a human behaviour using SVM classifier.

3. ARCHITECTURE OF THE PROPOSED MODEL

The architecture diagram for the Behaviour Analyzer system is shown in Fig-1. The Behaviour Analyzer system does the processes such as pre-processing, representation, classification and evaluation. User gives the input(feedback) in the form of text files. These text files are given to the system for pre-processing. In pre-processing, unwanted data such as stop words are eliminated.

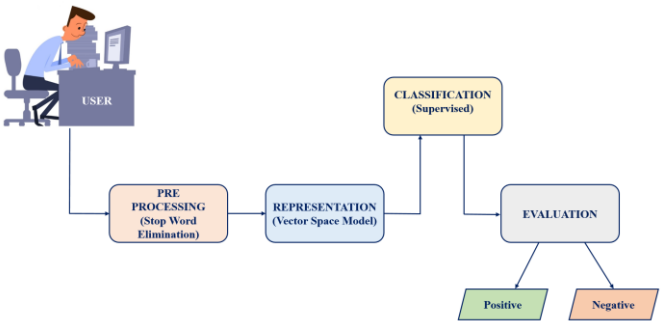

Fig-1: Architecture Diagram

Once the stop words are eliminated the unstructured text data is given for representation. In representation the unstructured data is represented in Vector Space Model in order to form structured data. This data is stored in storage. Then the data is given for classification where Support Vector Machine classifier is used to classify the data in order to find whether the data is positive or negative. The data obtained is then given for evaluation to get the desired results. Results indicate whether he is a positive minded person or negative minded person.

Data flow diagram indicates the flow of data which occurs in a system. The Data Flow Diagram of the Behaviour Analyzer is shown in Fig -2. Here the user and storage are two entities. The user gives feedback. The processing takes place once the user gives the feedback. Here the pre-processing is the first step. After the stop words are eliminated it goes for next step that is representation, where the text is represented in the form of matrix i.e, vector space model. It saves the data in storage. The stored data is then sent for classification. The data after classification is stored back again. The evaluation step is done based on the classified text. After evaluation the user gets the result. Result is either positive or negative.

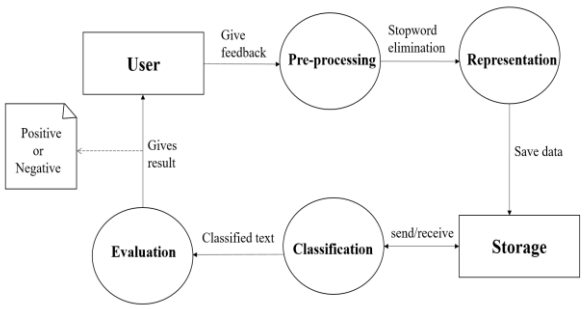

Fig -2: Data Flow Diagram

4.METHODOLOGY

This section tells about the implementation of the proposed system. The proposed system, based on the feedback given by the user analyses the users behaviour by identifying whether the feedback is positive or negative. The input dataset is basically the comments and reviews collected from the various sources such as Flipkart, Amazon, Twitter, Facebook and from various other news websites. Once the data is collected it is automatically labelled using TextBlob(). It calculates the polarity, if it is greater than 0 then it is labelled as 0 which indicates it is positive, if it is equal to 0 then it is labelled as 1 which indicates it is neutral and if it is less than 0 then it is labelled as 2 which indicates it is negative. Initially the data is being pre-processed. The following are the main steps that is implemented in the system:

Step 1: Preprocessing
Step 2: Representation
Step 3: Classification
Step 4: Evaluation

4.1.PREPROCESSING

In preprocessing step the punctuations, numbers, HTML tags and other unwanted data are removed. Once it is removed the strings are split into words and all the stopwords are eliminated.

The following shows the pseudocode:

1. Begin
2. Eliminate HTML
3. Eliminate non-letters i.e, punctuations and numbers
4. Turn to lower case and split into words
5. Load stopword list and eliminate if found in input data
6. End

4.2. REPRESENTATION

In representation, the count of occurrence of each word is given. All this data will be stored in a CSV file. The following shows the pseudocode:
1. Begin
2. Initialize the"CountVectorizer" object
3. Apply fit transform() which does two functions: First, it fits the model and learns the vocabulary; Second, it transforms our training data into feature vectors
4. Convert the result to an array
5. Sum up the counts of each vocabulary word and store
6. End

4.3. CLASSIFICATION

After the preprocessing and representation the data needs to be classified. Here the Support Vector Machine(SVM) classifier is being applied. SVM is used basically for classification and for regression analysis. It is an algorithm which is used to examine the data and it is a supervised learning model. The following shows the pseudocode:
1. Begin
2. Split input data into test data and train data using train test split( ) function
3. Train the SVM classifier using training data
4. Store the trained model to disk
5. Predict target variables for testing data
6. Calculate the accuracy
7. End

4.4. EVALUATION

Lastly, the evaluation step takes place. In Evaluation step, the result is given as positive, negative or neutral and the accuracy is also calculated and compared with other classifiers. A confusion matrix is also constructed in order to measure the performance. Once the result is obtained, a graph is being plotted. The following is the pseudocode:
1. Begin
2. Calculate accuracy by comparing predicted labels with actual labels using accuracy score ( ) function
3. Construct a confusion matrix for the predicted data using confusion matrix ( ) function
4. Plot graphs for the results
5. Display results
6. End

5. RESULTS AND ANALYSIS

When a user enters a feedback the time taken to process and the time taken to display the result is measured in seconds. The accuracy is displayed where SVM gives more accuracy when compared to the rest of the classifier.
RF accuracy:59%
SVM accuracy:62%
GNB accuracy:59%
KNNaccuracy:47%

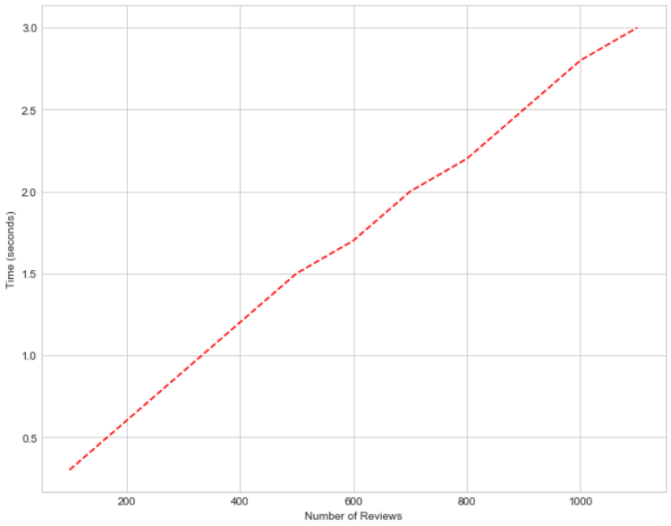

Fig -3: Preprocessing Time Graph

The Fig -3 shows the preprocessing time graph. It can be observed that the graph is a linear graph. The time taken for preprocessing is just in few seconds. It can be observed that around 1100 input reviews or comments got preprocessed in around 3.0 seconds only.

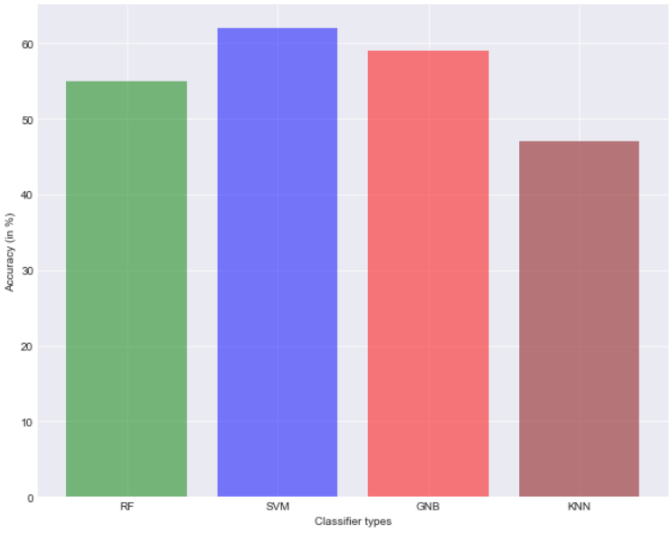

Fig -4: Pie chart for Training data

The Fig -4 shows a bar graph that shows accuracy of various classifiers. It can be observed that RF classifier gave an accuracy of 55%, SVM classifier gave an accuracy of 62%,

GNB classifier gave an accuracy of 59% and KNN classifier gave an accuracy of 47%.

6. CONCLUSION

This work has investigated the user's feedback and has given the particular prediction for that input. It identifies whether the feedback given by the user is positive, negative or neutral. It also displays the accuracy. Compared with other classifier. SVM classifier gives the better accuracy and efficient enough in execution.

ACKNOWLEDGEMENT

It is with great satisfaction and euphoria that we are submitting the paper on "An Integrated Framework to Identify Different Psychological Behaviour of a Human Being". We are profoundly indebted to our guide, Miss Rajatha, Assistant Professor, Department of Information Science & Engineering, for innumerable acts of timely advice, encouragement and we sincerely express our gratitude. We also thank him for his constant encouragement and support extended throughout. Finally, yet importantly, we express our heartfelt thanks to our family & friends for their wishes and encouragement throughout our work.

REFERENCES

[1] Yung-Shen Lin, Jung-Yi Jiang, and Shie-Jue Lee, "A Similarity Measure for Text Classification and Clustering", IEEE transactions on knowledge and data engineering, Vol. 26, No. 7, July 2014

[2] Fabrizio Sebastiani ,"Machine Learning in Automated Text Categorization", ACM Computing Surveys, Vol. 34, No. 1, March 2002.

[3] Anil K. Jain, Robert P.W. Duin and Jianchang Mao, "Statistical Pattern Recognition: A Review", IEEE Transactions on pattern analysis and machine intelligence, Vol. 22, No. 1, January 2000.

[4] Ahmed Elwhishi, Pin-Han Ho, K. Naik, and Basem Shihada, "Text Classification using Support Vector Machine", International Journal of Engineering Science Invention, ISSN (Online):2319 – 6734, ISSN (Print): 2319 – 6726 www.ijesi.org, Volume 4 Issue 11, November 2015, pp.33-37.

[5] Ming Li, Xin Chen, Xin Li, Bin Ma, and Paul M. B. Vita´nyi, "The Similarity Metric", IEEE transactions on information theory, Vol. 50, No. 12, December 2004.

[6] Sang-Bum Kim, Kyoung-Soo Han, Hae-Chang Rim, and Sung Hyon Myaeng, "Some Effective Techniques for Naive Bayes Text Classification", IEEE transactions on knowledge and data engineering, Vol. 18, No. 11, November 2006.

[7] Ana Carolina. E. S. Lima and Leandro N. de Castro, "Predicting Temperament from Twitter Data", 5th IIAI International Congress on Advanced Applied Informatics, 2016.

[8] Laleh Asadzadeh and Shahram Rahimi, "Analyzing Facebook Activities for Personality Recognition", 16th IEEE International Conference on Machine Learning and Applications, 2017

[9] Rashid Kamal, Munam Ali Shah, Asad Hanif, J Ahmad, "Real-time Opinion Mining of Twitter Data using Spring XD and Hadoop",Proceedings of the 23rd International Conference on Automation and Computing, University of Huddersfield, Huddersfield, UK, 7-8 September 2017.

[10] Michael Biaha and James Rumbaugh, "Object-Oriented Modeling and Design with UML", Second Edition, Pearson Publication.