

Special Issue of Second International Conference on Advancements in Research and Development (ICARD 2021)

# Sentimental Data Analysis-To Predict the User Emotions

Pranesh  $S^1$ , Jenita J  $R^2$ , Nisha  $D^3$ , Dr. Prabakar  $D^4$ 

<sup>1,2,3</sup>UG Scholar, Department of Computer Science and Engineering, SNS College of Technology, Coimbatore, India.

<sup>4</sup>Assistant Professor, Department of Computer Science and Engineering, SNS College of Technology, Coimbatore, India.

praneshsakthi@gmail.com<sup>1</sup>, jenita107@gmail.com<sup>2</sup>, nishadakshnamoorthy@gmail.com<sup>3</sup>, prabakaralam@gmail.com<sup>4</sup>

## Abstract

Creating a survey website in which the information will be updated automatically by analyzing the data in social media. In our project we are collecting the data from the user's through Application program interface and predicting the result of data whether is positive or negative emotions of text. The Neutral emotions are calculated through polarity score which represents the percentage of text emotions of a collected set of data. This way of predicting the data of user's helps to improve user's need and also understanding the user mindset, through that satisfying their needs and service. Biggest advantage of using the concept of sentiment analysis is to improve the user needs by directly collecting the data from the large set of users.

Keywords: Survey, Analysing, Polarity Score, Sentimental Result

## 1. Introduction

Sentimental data analysis is the process of predicting the user emotions such as positive or negative emotions from their text. Sentimental data analysis is text-based data mining that detects and removes subjective insights from user-generated content, allowing a company to better understand social feedback regarding its products, goods, and services when tracking online discourse. Sentimental analysis is a text classification technique that analyses a message and determines whether it is negative or positive.

# 1.1 Big Data Analysis

Big Data is a concept used to describe a large volume of organised and unstructured data that is difficult to access using traditional database and computational techniques. In most business cases, the amount of data is either too large, travels too quickly, or exceeds existing computing power. Emails, handheld devices, apps, directories, servers, and other sources are all used to gather information. When this data is collected, formatted, manipulated, saved, and evaluated, it will help a business obtain valuable insight into how to maximise sales, acquire or maintain clients, and enhance operations.

Using data that was historically unavailable or unusable, scientists, academics, and market customers can make smarter and quicker choices by analysing big data. [1-3]

#### **1.2 Data Mining**

The methods used to drive efficiencies and forecast consumer emotions are known as data mining. Data mining is the practise of looking through large data sets for similarities and associations in order to solve problems using data analysis. Enterprises will use data analysis software to forecast future developments.

#### Volume 03 Issue 03S March 2021

#### www.rspsciencehub.com

Association rules are generated in data mining by evaluating data for common trends and then using help and trust criteria to find the most relevant data relationships. Web mining is a form of data mining used in customer relationship management that combines data collected from the internet with data collected using conventional data mining methods and techniques. Website mining helps to better understand user behaviour and assess the effectiveness of a website. [4-6]

# **1.3 Data Preprocessing**

Data Preprocessing is one of the component way of web mining. Data preprocessing refers to any kind of processing performed on raw data to prepare it for further processing. Data preprocessing, which is commonly used as a preliminary data mining method, converts data into a medium that can be more conveniently and reliably interpreted for the user's purpose — for example, in a neural network.

Preprocessing web usage logs will retrieve useful collections of data called user transactions, which are classes of URL references. User sessions can be monitored to classify the user, the Web pages visited and the order in which they were visited, as well as the amount of time spent on each.

#### 2. Elastic Search

• Elastic search is a collaborative search and analytics engine based on Apache Lucene that is open-source and Restful. Elastic search has rapidly become the most popular search engine since its release in 2010, with applications

# 3. Block Diagram

including log analytics, full-text search, security intelligence, business analytics, and operational intelligence.

- Elastic search has its own database domainspecific language, which uses JSON to define the query. You can also nest other questions according to your requirements. Real-world ventures necessitate searching through several fields using various conditions, weights, recent records, values of predefined fields, and so on. All of this complication can be represented in a single question.
- All of this complication can be represented in a single question. The query DSL is robust, and it's designed to tackle real-world query complexities with only one query. Elastic search APIs are linked to Lucene and have the same name as Lucene operations. The Lucene Term Query is also used to run the Query DSL.
- A cluster is a group of one or more servers that store all of the data and provide federated indexing and search across all of them. DB Instance is the node for relational databases. There can be N nodes in a cluster of the same name.
- A Subset of index is called Shards. The collection of document in index is called Elastic Search. There are many shards in index. It identifies Unique Name such as Update, Search and Delete operation in elastic search. [5-9]



Fig.1. Sentimental Data Analysis



Fig 2.Architecture diagram

## 4. Proposed System

In today's web world, a wide variety of users use social media and social networks to search and read news and content related to it. Every day, a variety of issues arise, and social media has an impact on the news around these issues. As a result of previously poor statistics, this methodology's efficacy increases as the scale of the data grows. Social media facts and social news (i.e. text and emotions text) of a happening. As a result, we propose a new method for performing sentiment analysis on news results, focusing on social media information and social news (i.e. text and feelings text) about a particular event. The linguistics and emotions are classified using a Sentimental based algorithm, which serves as the basis for the sentiment analysis of the event. By employing a high-quality emotion wordbook, the word feeling computation algorithmic law aims to simplify the original word feelings. We'll deduce the feelings of each sentence using the term emotions. The proposed approach divides emotions into various groups using the Nave Bayes and Levenshtein algorithms from social media news info.

# Conclusion

Our project will analyse data from a broad dataset and present it in a visual format on a website. Centred on social media big data, we created a ground breaking method for sentiment computing for news events. The main goal is to analyse news data and determine what emotional responses this data elicits.

#### **Future work**

To analyze the data in our project, we use the APIs of various websites. In the future, we plan to incorporate the capability of storing people's search histories for the purpose of assisting people in better understanding the problem and obtaining real-time sentimental data from them. Creating a survey website that updates trending comments and suggests without the need for user interaction. **Application** 

1. This Survey website is user-friendly interaction with user's and easy to operate.

2. This can help the people to predict the result of the product through commands.

3. It is very helpful for business organization to get accurate feedback of their products from the user.

#### www.rspsciencehub.com

#### References

- [1]. J. Blitzer, M. Dredze, and F. Pereira. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In ACL, volume 45, page 440.
- [2].S. Bonzi. 1982. Characteristics of a literature as predictors of relatedness between cited and citing works. Journal of the American Society for Information Science, 33(4):208–216.
- [3].B. Pang, L. Lee, and S. Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In EMNLP, pages 79–86. Association for Computational Linguistics.
- [4].S. Piao, S. Ananiadou, Y. Tsuruoka, Y. Sasaki, and J. McNaught. 2007. Mining opinion polarity relations of citations. In International Workshop on Computational Semantics (IWCS), pages 366–371. Citeseer.
- [5].L. Polanyi and A. Zaenen. 2006. Contextual valence shifters. Computing attitude and affect in text: Theory and applications, pages 1–10. D.R. Radev, M.T. Joseph, B. Gibson, and P. Muthukrish
- [6].D. Hall, D. Jurafsky, and C.D. Manning. 2008. Studying the history of ideas using topic models. In EMNLP, pages 363–371.
- [7]. V. Hatzivassiloglou and K.R. McKeown. 1997.
  Predicting the semantic orientation of adjectives. In Proceedings of EACL, pages 174–181. Association for Computational Linguistics.
- [8].J.E. Hirsch. 2005. An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America, 102(46):16569.
- [9].K. Hyland. 1995. The Author in the Text: Hedging Scientific Writing. Hong Kong papers in linguistics and language teaching, 18:11.