A Mechanism Of Sentimental Analysis On Youtube Comments

Siddhant Thapliyal, Student Member, IEEE, Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun 248 002, India, sthapliyal37@gmail.com

Pranav Nautiyal, School of Computing, Graphic Era Hill University, Dehradun, Uttarakhand, pranavnautiyal013@gmail.com

Abstract

In today's era when we are generating, transmitting and using lot of digital data over various platforms by the help of internet. This data can be used for the various purpose like predictions, analysis and pattern matching. Data assist us in raising the standard of living for individuals we support: The primary reason why businesses should use data is to improve quality. A useful data framework may help any organization improve the quality of people's lives by allowing us to analyze and respond. So by inspiring from these applications. In this paper we have discussed about the Analysis of sentiments of the users who are commenting over various Youtube videos.

Index Terms Machine learning, sentiment analysis, bag of words, sentiments, simulation.

I. INTRODUCTION

Sentiment analysis is printed relevant mining that finds and concentrates abstract data from source material, helping organizations in figuring out the social sentiment of their image, item, or administration while checking on the web discussions. It is the method involved with looking at online substance to decide if it has a decent, terrible, or impartial close to home tone. Sentiment examination is the relevant mining of words that mirrors the social sentiment of a brand and furthermore helps organizations in deciding if the item they are driving will be in interest on the lookout or not [1], [2]. The motivation behind sentiment examination is to evaluate individuals' perspectives to help organizations develop [3]. It centers around extremity (positive, pessimistic, and nonpartisan), yet additionally on feelings (cheerful, miserable, furious, and so forth.). It utilizes an assortment of Natural Language Processing techniques, including Rule-based, Automatic, and Hybrid [4].

A. Data and its Uses

Data is basically the direct raw numbers accumulated over the span of a business' exercises. They might be utilized to quantify/record a wide scope of inside and outer organization tasks. While the actual information may not be extremely fascinating, it fills in as the establishment for all detailing
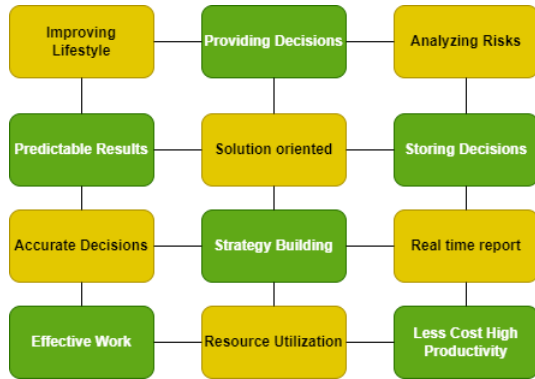
and is subsequently basic in business.

Fig. 1. Uses of data

1) Improving Lifestyle: Data will help you in working on the personal satisfaction for those you support: Improving quality is the main motivation behind why firms ought to utilize data. An effective data framework might assist your association with upgrading the nature of individuals' lives by permitting you to quantify and act.

2) Providing Decisions: Data approaches information. Great data gives unquestionable proof, but narrative proof, presumptions, or unique perception might bring about lost assets because of following up on a wrong end.

3) Analyzing Results: Data assists you with following along on the soundness of basic frameworks in   your association: Organizations might answer challenges before they become undeniable emergencies by involving data for quality checking. Powerful quality checking will empower your business to be proactive instead of receptive, and will assist it with supporting prescribed procedures after some time.

4) Predictable Results: Data empowers firms to evaluate the viability of a particular methodology: When strategies to handle a trouble are executed, assembling data will assist you with laying out how well your answer is doing and whether your system should be modified or changed over the long haul.

5) Solution Oriented: Data empowers firms to all the more successfully distinguish the main driver of issues. Data empowers organizations to imagine the associations between what's going on in different spots, offices, and frameworks. Is there an issue, for example, worker turnover or opportunity rates, that might show a reason assuming the occurrence of remedy botches has expanded? Contrasting these data focuses assists us with create more exact thoughts and carry out additional successful cures.

6) Storing Decisions: Data is a fundamental part of frameworks promotion. Involving data will help with presenting a convincing defense for framework change. Utilizing measurements to outline your contention will help you to make sense of why changes are required, whether you are requesting extra funding from public or confidential sources or presenting the defense for administrative changes.

7) Accurate Decisions: Data will help you in making sense of (both great and negative) activities to

partners. Whether your strategies and decisions give the ideal outcomes, you can be sure that you constructed your technique in light of dependable proof as opposed to taught mystery.

8) Strategy Building: Data helps proficiency. Data social affair and investigation that is viable will assist you with apportioning valuable assets where they are generally required. In the event that there is an expansion in extreme events in a specific help region, the data might be breaking down further to find assuming the ascent is wide or restricted to a solitary area. On the off chance that the issue is restricted, preparing, labor supply, or different assets might be given exactly where they are required, instead of fundamentally. Data will likewise assist organizations with concluding which regions ought to be focused on over others.

9) Real time Report: Data empowers you to spread solid areas all through your firm. Data examination will help you in recognizing high-performing programs, administration regions, and people. Whenever you've recognized your superior workers, you might investigate them to assemble ways of aiding low-performing programs, administration regions, and individuals.

10) Effective Work: Great data empowers organizations to set baselines, benchmarks, and focuses to continue to push forward. You will actually want to construct baselines, find benchmarks, and set execution focuses since data permits you to quantify. A gauge is the condition of a district before a specific cure   is embraced. Benchmarks, like Personal Outcome Measures® public measurements, characterize where others are in a comparative segment. Data assortment will empower your firm to set execution targets and celebrate accomplishment when they are met.

11) Resource Utilization: Your business undoubtedly as of now has most of the data and information expected to start examination. Your HR office is probably previously following data on your representatives. You are doubtlessly previously submitting episode data to your state's oversight office. You probably have something like one representative in your business who knows all about Excel. Yet, there is still expectation on the off chance that you don't do any of these things! There are a few free assets accessible online to assist you with getting everything rolling. Search the expressions "how to dissect data" or "how to develop an Excel realistic."

12) Less Cost High Productivity: Financing is turning out to be more outcome and data-driven. With the move from funding in view of administrations given to cash in light of results accomplished, associations are turning out to be progressively significant in carrying out proof based practice and creating data assortment and examination frameworks.

II. LITERATURE REVIEW

The data which are using in this research is textual data by the help of various keywords and text score we will try to analyze the mood of user in this field various researchers has already worked a lot some of those recent researchers which are the reason behind completion of this paper. It has developed into a stage for scattering data, discussing thoughts, and giving editorial on worldwide occasions. Given the wealth of information, perspectives, and data posted by the two individuals and official sources, Twitter is likewise a significant wellspring of wellbeing related data. Recognizing fascinating and significant material from enormous text-streams in different dialects is troublesome, and hardly any distributions have concentrated on dialects other than English. Papers like, [5], [6],

[7], [8], [9] discussed the various aspects of sentiment analysis. They tried to solve the problem through some potential machine learning algorithms. Important results, like, accuracy, F1 score, precision and recall were estimated and compared.
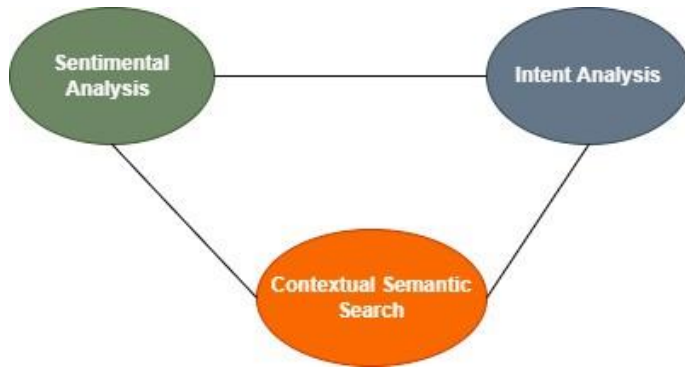
III. TYPES OF TEXT ANALYSIS

Fig. 2. Types of Text Analysis

A. Sentimental Analysis

Sentiment Analysis is the most successive message classification innovation that investigations an approaching message and decides whether the basic sentiment is positive, negative, or impartial.

B. Intent Analysis

Intent analysis raises the stakes by deciding whether a correspondence is connected with an assessment, news, promoting, grievance, proposal, appreciation, or enquiry.

C. Contextual Semantic Search

Contextual Semantic Search is an insightful brilliant hunt technique (a.k.a. CSS). CSS works by taking many messages and an idea (like Price) as info and separating those messages that intently match the gave thought. The diagram beneath shows how CSS offers a huge development over the business' ongoing practices.

IV. SEMANTIC ANALYSIS AND IT'S TYPES

A. Fine-grained sentiment analysis

Not set in stone by the extremity. This class may be exceptionally sure, extremely certain, impartial, negative, and exceptionally negative. The rating is given on a size of 1 to 5. A rating of 5 shows that it is very certain, a rating of 2 demonstrates that it is negative, and a rating of 3 shows that it is nonpartisan.
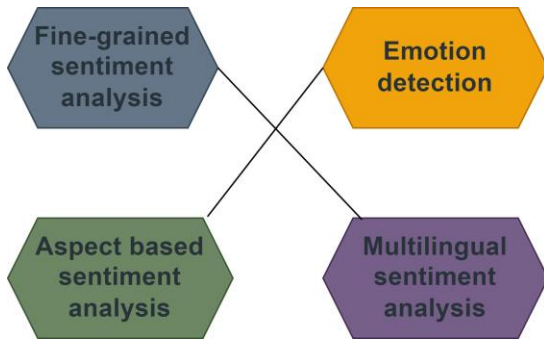
Fig. 3. Types of Semantic Analysis

B. Emotion detection

This incorporates the sentiments euphoric, miserable, furious, upset, lively, lovely, etc. It is otherwise called a sentiment analysis lexical methodology.

C. Aspect based sentiment analysis

It focuses on a specific component, for example, if an individual need to check the element of a cell phone, it checks the perspective like battery, screen, camera quality, etc.

D. Multilingual sentiment analysis

Multilingual alludes to numerous dialects that should be named great, negative, or impartial. This is a truly requesting and testing task.

V. IMPLEMENTED WORK

By the help of sentimental analysis, we are analyzing comments on a particular youtube video by the help of bag of word model. We are performing certain operation over the data by creating a python data- frame after that we are calculating the individual scores of each comment and providing them a tag whether it is a positive, negative or a neutral intent. after combining all the scores by the help of average scores we are demonstrating the final result and top comments.

TABLE I DEVICE SPECIFICATIONS

Name	Specification
Device name	DESKTOP-KRNS9VD
Processor	Intel(R) Core(TM) i3
Installed RAM	4.00 GB
Device ID	5AC360CB-DF96-4AD5-9794-1BBF54FCC795

Product ID	00331-10000-00001-AA437
System type	64-bit operating system, x64-based processor
Pen and touch	No pen or touch input is available for this display

In this paper we have used pafy library of python to extract the comments of the given youtube video and then analyzed and processed it with the semantic analyzer and lemmatization methodologies. We have used the python3 for this work.

VI. CONCLUSION AND FUTURE SCOPE

In this paper, we achieved good analysis of the comments in the language English. But this research has limitation to English only. So in future there is scope of further improvement and enhancement in other languages so that we can perform same operations on other languages also.

```
lst=function_analyzer("https://www.youtube.com/watch?v=3j7Ypfs7RAM")
colname = ["Title","Author","Duration","Likes","Dislikes","ViewCount","Rating","Sentiment of Comments","Top Comments"]
i=0
for item in lst:
  if(i<len(colname)):
    print(colname[i],item,sep=" : ")
  else:
    print(item)
  i+=1
```

```
Title : Dash Cam Owners Australia - Biggest Crashes Compilation
Author : Dash Cam Owners Australia
Duration : 00:08:28
Likes : 4474
Dislikes : None
ViewCount : 322892
Rating : None
Sentiment of Comments : Neutral
Top Comments : 65.0
0    If turn in front a truck like some these guys,...
1    <a href="https://www.youtube.com/watch?v=3j7Yp...
2              Now that&#39;s entertainment 🤣 🤣🤣 👌
6    You can always tell the Aussie dashcam footage...
9    <a href="https://www.youtube.com/watch?v=3j7Yp...
Name: comments, dtype: object
```

Fig. 4. Neutral comment result

REFERENCES

[1]L. Wang, J. Niu, and S. Yu, "SentiDiff: Combining Textual Information and Sentiment Diffusion Patterns for Twitter Sentiment Analysis," IEEE Transactions on Knowledge and Data Engineering, vol. 32, no. 10, pp. 2026–2039, 2020.

[2]G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng, and X. Wu, "Chinese Text Sentiment Analysis Based on Extended Sentiment Dictionary," IEEE Access, vol. 7, pp. 43 749–43 762, 2019.

[3]G. Zhai, Y. Yang, H. Wang, and S. Du, "Multi-attention fusion modeling for sentiment analysis of educational big data," Big Data Mining and Analytics, vol. 3, no. 4, pp. 311–319, 2020.

[4]Z. Li, R. Li, and G. Jin, "Sentiment Analysis of Danmaku Videos Based on Na ̈ıve Bayes and Sentiment Dictionary," IEEE Access, vol. 8, pp. 75 073–75 084, 2020.

[5]G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng, and X. Wu, "Chinese Text Sentiment Analysis Based on Extended Sentiment Dictionary," IEEE Access, vol. 7, pp. 43 749–43 762, 2019.

[6]Z. Li, R. Li, and G. Jin, "Sentiment Analysis of Danmaku Videos Based on Na ̈ıve Bayes and Sentiment Dictionary," IEEE Access, vol. 8, pp. 75 073–75 084, 2020.

[7]K. Schouten and F. Frasincar, "Survey on Aspect-Level Sentiment Analysis," IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 3, pp.
813–830, 2016.

[8]X. Fu, J. Yang, J. Li, M. Fang, and H. Wang, "Lexicon-Enhanced LSTM With Attention for General Sentiment Analysis," IEEE Access, vol. 6, pp.
71 884–71 891, 2018.

[9]H. T. Phan, V. C. Tran, N. T. Nguyen, and D. Hwang, "Improving the Performance of Sentiment Analysis of Tweets Containing Fuzzy Sentiment Using the Feature Ensemble Model," IEEE Access, vol. 8, pp. 14 630–14 641, 2020.