International Journal of Research Radicals in Multidisciplinary Fields (IJRRMF), ISSN: 2960-043X
Volume 1, Issue 2, July-December, 2022, Available online at: www.researchradicals.com

Customer Opinion Analysis Leveraging Natural Language Processing Algorithms

Amit Sengupta

Senior Delivery Manager, Software Engineering, Capgemini America

ABSTRACT

Customer Sentiment analysis has become crucial in today's digital age, enabling businesses to glean insights from vast amounts of textual data, including customer reviews, social media comments, and news articles. By utilizing natural language processing (NLP) techniques, sentiment analysis using NLP categorizes opinions as positive, negative, or neutral, providing valuable feedback on products, services, or brands. This analysis is powered by various algorithms such as Naive Bayes, Support Vector Machines (SVM), and Recurrent Neural Networks (RNN), which help in understanding the overall sentiment and emotional tone conveyed in the text, making it an indispensable tool for business intelligence and decision-making.

This document acquaints a methodology with assumption investigation which utilizes different content standardization methods in Natural Language Processing (NLP) for converting a text into vector and briefly explains the importance of standardization methods and how they are used in python with the help of its Natural Language Toolkit (NLTK) library.

Keywords: - Natural Language Processing, Cognitive Analysis, Sentiment Analysis, Opinion Analysis, Artificial Intelligence, Machine Learning, Historical Perspective, Emotion Detection, Targeted Customer Handling, Enhanced Customer Experience.

INTRODUCTION

Sentiment analysis using NLP is a method that identifies the emotional state or sentiment behind a situation, often using NLP to analyze text data. Language serves as a mediator for human communication, and each statement carries a sentiment, which canbe positive, negative, or neutral.

Types of Opinion Analysis: -
Sentiment analysis, also known as opinion mining, is a subfield of Natural Language Processing (NLP) that includes deciding and concentrating on the emotional data in an info text. This can be an assessment, an evaluation, or an inclination about a specific point or item. Here are the fundamental sorts of feeling examination:

- **Fine-grained Sentiment Analysis**: This goes beyond just positive, negative, or neutral. It involves very specific ratings, for example a 5-star rating.

- **Emotion detection**: This aims to detect emotions like happiness, frustration, anger, sadness, etc. The biggest challenge here is being able to accurately identify these emotions in text.

- **Aspect-based Sentiment Analysis**: This is generally used to understand specific aspects of a certain product or service. For example, in a review like "The battery life of this phone is great, but the screen is not very clear", the sentiment towards the battery life is positive, but it's negative towards the screen.

- **Multilingual sentiment analysis**: This can be particularly challenging because the same word can convey different sentiments in different languages.

- **Intent Analysis**: This goes a step further to understand the user's intention behind a certain statement. For example, a statement like "I would need a car" might indicate a purchasing intent.

Sentiment analysis using NLP is an interesting area because of the innate vagueness of human language.

Business Value for NLP: -
NLP for sentiment analysis is important for several reasons:

- **Business Intelligence:** It helps businesses understand how their customers feel about their products or services. This can guide improvements, address customer concerns, and enhance overall customer satisfaction.

- **Market Research:** By analyzing public sentiment towards products, services, or brand mentions on social media, companies can gain insights into market trends and competitors.

- **Customer Service:** Sentiment analysis can help identify negative reviews or feedback in real-time, allowing for quicker responses and problem resolution.

- **Product Analytics:** It can be used to understand user feedback on various aspects of a product, helping drive product strategy and development.

- **Public Relations:** Sentiment analysis can help monitor public sentiment towards a company or individual, enabling proactive management of public relations.

- **Politics and Public Policy:** In politics, sentiment analysis is used to gauge public opinion towards policies or political entities, which can inform strategy and messaging.

Keep in mind, the objective of sentiment analysis using NLP isn't simply to grasp opinion however to utilize that comprehension to accomplish explicit targets. It's a useful asset, yet like any device, its value comes with the fact of how it's utilized.

Common Classification Algorithms in NLP: -
In machine learning and statistics, classification is a supervised learning approach in which the computer program learns from the input data and then uses this learning to classify new observations. Few types of classification algorithms in machine learning are:

- **Logistic Regression -**It is a statistical method for analyzing a data set in which there are one or more independent variables that determine an outcome.One should consider utilizing logistic regression when the Y variable takes on just two qualities . Such a variable is alludedto as binary or dichotomous. Dichotomous essentially implies two classifications, for example - yes/no, deficient/non-blemished,achievement/disappointment, etc. Binary refers to 0's and 1's.

- **Naive Bayes Classifier -**The Naive Bayes classifier is a simple classifier that classifies based on probabilities of events. It is an arrangement strategydependent on Bayes' Theorem with the supposition of freedom among indicators

$$P(A\backslash B) = P(B\backslash A)P(A) / P(B)$$

Where, P(A\B) is the probability of A given B. Here, A will be dependent variable (which is to be predicted) and B will beindependent variable (Features).

Logistic Regression Model v/s Naive Bayes Model: -
There are prominent models to wade in various ML issues. Two of them Logistic regression and Naive Bayes are popularly recognized, these two provide results which are analogous but still are contrasting, the reason is they take into account different viewpoints. Naive Bayes is based on the concept that it extracts probability of the situation that a component vector is related with a label. It is based on the Bayes's hypothesis. The algorithm has certain assumptions which is unaffected from the realistic facts which says the component vectors are independent which is totally a hypothetical condition and is not always feasible. Whereas Logistic regression is an algorithm which calculates the probability that a feature is associated to a class. This algorithm is preferred because it can give satisfactory results even if the some of the features are found to be associated to a certain class. We have used the following features to do the comparison.

- Both algorithms (Naive Bayes and Logistic Regression) are used for classification Cases Though both algorithms are mainly used for solving problems that include classification tasks , the main difference is , Logistic Regression is limited for only binary classification e.g. Predicting whether a person is infected with a disease or not , mail is

spam or not, given sentiment is positive or negative , etc. whereas Naive Bayes Algorithm can be applied on multiclass classification problems also.

- Working for both the algorithms are very different. Naive Bayes algorithm is a probabilistic approach whereas, in Logistic Regression we use sigmoid function as activation function to map the result between 0 and 1. Naive Bayes classification calculation includes P(y—x) (which means probability of y given x).So, when there are multiple features then Naive Bayes classifier assumes that all features are independent of each other which does not hold true while dealing with real world problems.Therefore, if binary classification is considered, then it is generally observed that logistic regression gives better results as compared to Naive bayes classifier.

- Naive Bayes assumes all the features are different from each other. It is dependent upon each other a lot hence accuracy is less. Logistic regression splits it linearly; hence it is important to watch the accuracy according to it.

- The outcome for both the models depends on the data set used. While with Naive Bayes, when the preparation information size is a bit unequal with the data and information on earlier probabilities this algorithm helps in improving the outcomes. With Logistic regression, when the training data size is small compared to various features which is already supported with it.

Challenges with Sentimental Analysis in NLP: -
Sentiment analysis, while powerful, comes with its own set of challenges:

- **Sarcasm and Irony:** These linguistic features can completely reverse the sentiment of a statement. Detecting sarcasm and irony is a complex task even for humans, and it's even more challenging for AI systems.

- **Contextual Understanding:** The sentiment of certain words can change based on the context in which they're used.

- **Negations and Double Negatives:** Phrases like "not bad" or "not unimpressive" can be difficult to interpret correctly because they require understanding of double negatives and other linguistic nuances.

- **Emojis and Slang:** Text data, especially from social media, often contains emojis and slang. The sentiment of these can be hard to determine as their meanings can be subjective and vary across different cultures and communities.

- **Multilingual Sentiment Analysis:** Sentiment analysis becomes significantly more difficult when applied to multiple languages. Direct translation might not carry the same sentiment, and cultural differences can further complicate the analysis.

- **Aspect-Based Sentiment Analysis:** Determining sentiment towards specific aspects within a text can be challenging. For instance, a restaurant review might have a positive sentiment towards the food, but a negative sentiment towards the service.

These challenges highlight the complexity of human language and communication. Overcoming them requires advanced NLP techniques, deep learning models, and a large amount of diverse and well-labelled training data. Despite these challenges, sentiment analysis continues to be a rapidly evolving field with vast potential in today's customer first industry.

Use Cases for NLP Adoption: -
Natural Language Processing has a wide range of applications across various domains. Here are some key applications:

- **Customer Feedback**: Businesses use sentiment analysis to process customer feedback and reviews. This helps them understand customer satisfaction and preferences and make data-driven decisions.
- **Social Media Monitoring**: Brands monitor social media platforms to understand public sentiment about their products or services. This can help in reputation management and in identifying potential crises before they escalate.
- **Market Research**: Sentiment analysis can be used to understand public opinion about a product or a political event. This can provide valuable insights for market research.

- **Product Analytics**: Companies use sentiment analysis to gather insights from product reviews. This can guide product enhancements and innovations.
- **Healthcare**: In healthcare, sentiment analysis can be used to understand patient experiences and feedback about treatments, doctors, or hospitals.
- **Finance**: In the financial sector, sentiment analysis is used to gauge market sentiment. Traders and investors use this information to make informed decisions.
- **Politics**: In politics, NLP for sentiment analysis is used to understand public opinion about certain policies or politicians. This can guide political campaigns and strategies.
- **Human Resources**: HR departments use sentiment analysis using NLP to understand employee feedback and improve workplace culture.

The potential applications of NLP and Supervised Learnings are vast and continue to grow with advancements in AI and Machine Learning technologies.

CONCLUSION

NLP stands as a powerful tool in deciphering the complex landscape of human emotions embedded within textual data. By leveraging various techniques and methodologies such as text analysis and lexicon-based approaches, analysts can extract valuable insights, ranging from consumer preferences to political sentiment, thereby informing decision-making processes across diverse domains. The polarity of sentiments identified helps in evaluating brand reputation and other significant use cases, offering capabilities through which we can better comprehend, analyze and navigate the digital realm.

REFERENCES

[1]. Stanford NLP Group. CoreNLP. Stanford University, Stanford USA. https://stanfordnlp.github.io/CoreNLP/index.html.
[2]. S. Vijayarani and R.Janani. Text mining: Open-source tokenization tools – an analysis. Advanced Computational Intelligence: An International Journal (ACII), 3(1):37–47, January 2016.
[3]. M. Erwig and R. Gopinath. Explanations for regular expressions. In A. Zisman J. de Lara, editor, Fundamental Approaches to Software Engineering, volume 7212 of Lecture Notes in Computer Science, pages 394–408, Berlin, 2012. Springer.
[4]. Goswami, MaloyJyoti. "Leveraging AI for Cost Efficiency and Optimized Cloud Resource Management." International Journal of New Media Studies: International Peer Reviewed Scholarly Indexed Journal 7.1 (2020): 21-27.
[5]. Pala, Sravan Kumar. "Databricks Analytics: Empowering Data Processing, Machine Learning and Real-Time Analytics." Machine Learning 10.1 (2021).
[6]. Hitali Shah.(2017). Built-in Testing for Component-Based Software Development. International Journal of New Media Studies: International Peer Reviewed Scholarly Indexed Journal, 4(2), 104–107. Retrieved from https://ijnms.com/index.php/ijnms/article/view/259
[7]. Palak Raina, Hitali Shah. (2017). A New Transmission Scheme for MIMO - OFDM using V Blast Architecture.Eduzone: International Peer Reviewed/Refereed Multidisciplinary Journal, 6(1), 31–38. Retrieved from https://www.eduzonejournal.com/index.php/eiprmj/article/view/628
[8]. Raina, Palak, and Hitali Shah."Security in Networks." International Journal of Business Management and Visuals, ISSN: 3006-2705 1.2 (2018): 30-48.
[9]. Leveraging Inflection Tablesfor Stemming and Lemmatization, volume 1, Berlin, January 2016. Association for Computational Linguistics.
[10]. Amri Samir and ZenkouarLahbib. Stemming and lemmatization for information retrieval systems in amazigh language. In M. Al Achhab N. Enneya Y. Tabii, M. Lazaar, editor, Big Data, Cloud and Applications. BDCA 2018, volume 872 of Communications in Computer and Information Science, pages 222–233, Kenitra, Morocco, August 2018. Springer, Cham.
[11]. Raina, Palak, and Hitali Shah."Data-Intensive Computing on Grid Computing Environment." International Journal of Open Publication and Exploration (IJOPE), ISSN: 3006-2853, Volume 6, Issue 1, January-June, 2018.
[12]. Hitali Shah."Millimeter-Wave Mobile Communication for 5G". International Journal of Transcontinental Discoveries, ISSN: 3006-628X, vol. 5, no. 1, July 2018, pp. 68-74, https://internationaljournals.org/index.php/ijtd/article/view/102.
[13]. Chintala, S. "Evaluating the Impact of AI on Mental Health Assessments and Therapies." EDUZONE: International Peer Reviewed/Refereed Multidisciplinary Journal (EIPRMJ) 7.2 (2018): 120-128.

[14]. Kumar, Bharath. "Machine Learning Models for Predicting Neurological Disorders from Brain Imaging Data." EDUZONE: International Peer Reviewed/Refereed Multidisciplinary Journal (EIPRMJ), ISSN: 2319-5045, Volume 10, Issue 2, July-December, 2021.
[15]. Chintala, S. "IoT and Cloud Computing: Enhancing Connectivity." International Journal of New Media Studies (IJNMS) 6.1 (2019): 18-25.
[16]. Anjali Ganesh Jivani. A comparative study of stemming algorithms. International Journal of Comp. Tech. Appl, 2(6):1930– 1938, 2011.
[17]. Anna Stavrianou, Caroline Brun, Tomi Silander, and Claude Roux. Nlp-based feature extraction for automated tweet classification. In Proceedings of the 1st International Conference on Interactions between Data Mining and Natural Language Processing, volume 1202, pages 145–146, September 2014.
[18]. Goswami, MaloyJyoti. "Challenges and Solutions in Integrating AI with Multi-Cloud Architectures." International Journal of Enhanced Research in Management & Computer Applications ISSN: 2319-7471, Vol. 10 Issue 10, October, 2021.
[19]. Sravan Kumar Pala, Improving Customer Experience in Banking using Big Data Insights, International Journal of Enhanced Research in Educational Development (IJERED), ISSN: 2319-7463, Vol. 8 Issue 5, September-October 2020.
[20]. Sravan Kumar Pala, Use and Applications of Data Analytics in Human Resource Management and Talent Acquisition, International Journal of Enhanced Research in Management & Computer Applications ISSN: 2319-7463, Vol. 10 Issue 6, June-2021.
[21]. Goswami, MaloyJyoti. "Utilizing AI for Automated Vulnerability Assessment and Patch Management." EDUZONE,Volume 8, Issue 2, July-December 2019, Available online at: www.eduzonejournal.com
[22]. Ammar Ismael Kadhim. Term weighting for feature extraction on twitter: A comparison between bm25 and tf-idf. In International Conference on Advanced Science and Engineering, pages 1–6, Iraq, April 2019.IEEE.
[23]. Chao-Ying Joanne Peng, Kuk Lida Lee, and Gary M. Ingersoll. An introduction to logistic regression analysis and reporting. The Journal of Educational Research, 96(1):3–14, 220.
[24]. I. Rish. An empirical study of the naive bayes classifier. Technical report, T.J.Watson Research Center, Hawthorne, NY, 2001. [14] Kaggle. Sentiment analysis on movie reviews. https://www.kaggle.com/c/sentiment-analysis-on-movie-reviews/data, March 2014.