Automated Customer Feedback Analysis

Introduction

Why use ML/Deep Learning ?

Machine learning is an application of artificial intelligence that provides a system to learn automatically and enhance from experience without began programmed explicitly. The process of machine learning begins with observations of data, such as analysis, direct experiences and classifications, in order to look for patterns in data and make the better decision in the future based on the database we have. Ever wonder how human identify and classify beautiful shapes of clouds in the sky or different letters in papers and many kinds of objects that we came across, these kinds of actions are performed by our brain in just fraction of seconds. what if we able to build such machines which can perform this kind of tasks? the answer leads to evaluation of an interesting concept called deep learning. Working procedure of DL is basically based on the deep neural network. Deep learning is defined as the subset of machine learning, the Deep neural network is entirely based on the mechanism of hidden layers.

Why used NLP ?

One of the major characteristics of Machine intelligence, An optimal classification of input text based on emotions is called sentimental analysis, The input signal(audio or text) which will be provided to the machine that needs to be classified, is given as input to the NLP algorithm and processed to create emotional ontology for a better understanding of the semantics and relationships among input data. For an efficient sentiment analysis.

They can be real one day with the help of Natural Language Processing (NLP) and Deep Learning (DL). The aim of NLP is to help computers to understand human speech. NLP is a branch of artificial intelligence that deals with analyzing, understanding and generating the languages that humans use naturally to interface with computers in both written and spoken contexts using natural human languages instead of computer languages. In short, NLP is a technique where the machine can become more human thereby reducing the distance between them. We see everyday applications like how Facebook can automatically organize photos, identify faces, and suggest which friends to tag. Or how Google can programmatically translate 103 languages with extreme accuracy. [7] proposed the Natural Language Processing Toolkit (NLTK) is used to extract the audio and text features, for that the input provided to the machine is the word or a sentence or a document that highlight the user emotions. [3] this paper intends to explore the analytical aspects such as Natural Language Processing and Deep Learning that will derive meaningful insights from Customer Perception. In this paper, we analyze the specific question: Is it possible to solely understand and capture customer concerns and reasons, besides sentiment using DL/NLP

1. **Data Collection**

   [2, 6] proposed the dataset collection from Amazon, accessible through the Kaggle platform, to conduct sentiment analysis (SA). The dataset contains reviews spanning five diverse product categories: mobile electronics, furniture, camera, grocery, and watches. These reviews, originating from Amazon, were publicly accessible on Kaggle. The dataset is quite large, with more than 100

million reviews in total. However, we have chosen to focus on a subset of 80,000 reviews from each of these categories, resulting in a dataset of 400,000 records in total. Each record within the extracted Amazon reviews dataset is associated with a set of attributes, including marketplace information, customer-ID, review-ID, product-ID, product parent, product title, product category, star rating, helpful votes, total votes, verification status of the purchase, review headline, and review body. These attributes collectively provide a comprehensive foundation for conducting an in-depth analysis as part of this research endeavor.

[4, 10] this paper propose Modified Logistic Regression (MLR) tech unique to classify and analyze the movies review dataset, which shall help the viewers to decide whether the movies can be watched or not. The first step in classifying the movie reviews is to construct the movie review dataset from reliable sources Dataset: https://www.kaggle.com/nltkdata/ movie-review[Movie Reviews](https://www.kaggle.com/nltkdata/movie-review)

2. Preprocessing

2.1 Handling Missing Values

When addressing the issue of missing values within the dataset, our primary focus lies on managing the missing entries within the "review body" and "star rating" features. This emphasis is placed due to the significance of these features in relation to sentiments and their corresponding outputs. The specifics regarding the count of missing values for each feature in our dataset For features with an object data type, we have employed the fillna() method available in Python to populate the null values. Additionally, for the "star rating" feature, we have opted to utilize the Interpolate method. This method calculates an average based on the values that surround the empty cell, both above and below it, thereby aiding in the imputation of missing data.

2.2 Lowercasing

In this step, we convert all review words into lowercase. For example, "Computer" and "USBpin" are transformed into "comouter" and "usbpin". Lowercasing helps standardize the text and reduces the dimensionality of the data by treating words in a case-insensitive manner.
In the customer reviews, consumers enter material without following grammar norms, in that the entered text contains both lower and upper case characters. Many of the methods utilized in the study are case-sensitive. As a result, the classifier has difficulty determining the polarity of the provided text. Such an issue may be avoided simply by changing the entire text to a standard format. Conversely, if we wish to perform the same process manually, the lower (txt) statement is used. It transforms all upper case text to lower case while leaving the other characters untouched. The following example illustrates how to convert text data to lower case: "I Am A Senior Big Data Analyst in Islamabad" may be translated to lowercase as "I am a senior big data analyst in Islamabad".

2.3 Stop word removal

Stop words are elements in a sentence that hold no significance across all sectors in the field of text mining. We have eliminated all stop words, punctuation marks, and HTML tags from the reviews within our corpus. This preprocessing step helps to reduce noise in the data and improve computational efficiency. [15] Many words in text files appear repeatedly. As a result, it is critical to delete the stop words. Indeed, stop words never provide significance to the written substance. These kinds of words often appear in large numbers in the text. Due to this, the text mining process has become difficult, and the classifiers have produced unexpected results. The stop words are deleted from the selected data in this stage. This strategy minimizes textual content facts while improving overall system efficiency. For example, after deleting the stop words, the preceding statement may read as “I data analyst Islamabad”.

2.4 Tokenization

we applied both sentence tokenization and word tokenization. Tokenization is a process where a sequence of text is broken down into individual components known as tokens. These tokens can encompass single words, phrases, or even entire sentences . These tokens then serve as inputs for various processes such as parsing and text mining. It helps models focus on the meaning of individual units rather than processing the entire text as a single sequence

2.5 Lemmatization/ Stemming

Part of speech tagging

M. Cataltas, S. Dogramaci, S. Yumusak, and K. Oztoprak [2] proposed a method that fully takes advantage of Part-of-Speech Tagging. They extracted features and opinions with predefined patterns. This method slightly outperforms M. Cataltas, S. Dogramaci, S. Yumusak, and K. Oztoprak method.

3. Text Representation

3.1 Bag of words

In the context of sentiment analysis, the Bag-of-Words (BoW) approach plays a pivotal role in transforming textual data into a format that can be understood by machine learning algorithms. By representing each review as a “bag” containing the count of words without considering their order, we simplify the complexity of language for computational analysis. This method allows us to convert diverse and unstructured customer reviews into numerical vectors, forming the foundation for our sentiment analysis model

$$P(w_t|\text{context}) = \text{softmax}\left(W_{\text{out}} \cdot \frac{1}{n}\sum_{i=1}^{n} W_{\text{in}}[w_{t-i}] + b_{\text{in}}\right)$$

3.2 TF-IDF

we also use TF–IDF to figure out which words are really important. TF–IDF looks at how often a word shows up in one review compared to how rare it is across all reviews. This helps us focus on words that truly matter for understanding feelings in each review.

So, TF–IDF is like a smart way of picking out the words that tell us the most about what people think in their reviews.

$$\text{TF–IDF}(t, d, D) = \text{tf}(t, d) \cdot \text{idf}(t, D)$$

TF-IDF is an efficient scoring method for calculating relative significance of a word to whole document in [5]. It is calculated by taking products of two terms: terms frequency (TF) and inverse document frequency (IDF). TF is a measurement of the number of occurrences of a word in a document which is a review in our case. It is normalized by dividing it into the total number of words in the document and scaled logarithmically. IDF is calculation of the importance of a word. It ranks frequent terms higher and infrequent ones lower to compute the significance of a term.

3.3 Word embedding

Every word is represented numerically and in vector form through word embedding. Word embedding refers to texts with exact representations of words with the same meaning. In particular, word embedding is unsupervised learning of word representation, which is relatively similar to semantic similarity. This refers to words in a coordinated scheme in which similar terms are put closer together, based on a set of relationships[15]

4. Rule-Based System

To extract syntactically actionable insights from customer feedbacks, it is necessary to understand the business intention for collecting them. Businesses run customer surveys or reviews for improvement or to proactively track customer churn, from customer complaints for products and service improvements; and in some cases, to capture the most desirable features from a product purchase or new product introduction. With better business understanding, it will be easier to create rule-based semantics relating to the problem in question[3]

[4] this paper proposed the classification methods as Supervised and Unsupervised Machine learning Approach, indicate the advantage is Sentence level analysis performs better than document and aspect level analysis and indicate the disadvantage is Performance and accuracy of the system is completely depending on the parameters considered for the classifier

4.1 Word-Sense Disambiguation-WSD

The anchor texts are annotated based on the business problem of car reviews analysis. Beyond rule-based annotation, word-sense disambiguation should be sorted out manually, or by the training sequences created out of relevant document scrapings[3]

4.2 Linguistic Variables

When deciding which linguistic variables to use, we came to the realization that in our case it was needed to be a combination of accuracy and flexibility, so that a fair comparison with other classification methods could be easily established, and a degree-of polarity of a given sentence could be established (poor, most, very, moderate, slightly and etc.)

4.3 Negation handling

A negation is any word matching the following regular expression as

never, no, nothing, nowhere, no one, none, not, haven't, hasn't, hadn't, cant, couldn't, shouldn't, wont, wouldn't, don't, doesn't, didn't, isn't, aren't, aint

According to the [10] notices that the so-called Weak (mild) words such as good and bad behave like their opposites when negated (bad = not good, good = not bad), whilst "Strong (intense) words like superb and terrible have very general meanings under negation.". According to Potts: "not superb is consistent with everything from horrible to just-shy-of-superb, and different lexical items favor different senses.

5. ML-Based Sentiment Analysis

Sentiment analysis describe the type and amount of emotion expressed in product reviews/text.Discrete emotions and cognitive characteristics are important determinants and demonstrated significant role in several classification problems using product review. In addition, discrete positive and discrete negative emotion dimensions are explored from product reviews and their effectiveness is investigated for helpfulness prediction such as angry and anxiety discrete negative emotions.

[9, 13] Most previous studies on emotion recognition focused on use of single sensor modality, features and classifiers, which are ineffective to dis criminate complex emotion classes. Fusion of multiple modalities at improving classification accuracy by exploiting the complementarity of different modalities. For effective emotion recognition, data, features and classifiers may need to be fused with appropriate strategies as follows.

Analysis of user reviews or opinions are becoming one of the significant aspects of sentiment analysis. It involves finding the polarity of each review created by the user on social net working through opinion mining. The three review polarity indicators are positive, negative and neutral. User's sentiments are expressed in specific emotions, numbers, ratings and words for classification[4, 11]

5.1 Model selection

This paper proposed classification methods for Supervised and Unsupervised Machine learning Approach, indicate the advantages are usage of dictionary is not required, high classification accuracy and indicate the disadvantage is trained classifier algorithm for one domain may or may not work for other domains.[4]

5.2 Model training

To build a robust sentiment analysis model, the dataset should first be split into training, validation and test sets. [4, 7, 11] The training set is used to train the model, while the validation set helps tune the model's hyperparameters, such as learning rate, batch size, and regularization, through cross-validation. This ensures that the model generalizes well to unseen data. Finally, the test set is used to evaluate the model's performance after tuning.[9, 11] The process of constructing the emotion recognition model includes data collection, emotion-related feature extraction, feature reduction, and classifier model building. After the first three steps are completed, the final task is to design an effective emotion classifier model based on certain classification performance criteria. The accuracy of the emotion recognition depends largely on the classifier developed.

Additionally, it's important to check if the dataset is **balanced or imbalanced** in terms of sentiment labels (Ex: positive, negative, neutral). If the dataset is **imbalanced**, where one sentiment class is more frequent than others, techniques like oversampling the minority class, under sampling the majority class or using techniques [7] such as **SMOTE** (Synthetic Minority Over-sampling Technique) can be applied. Alternatively, class weighting can be adjusted in the model to ensure it doesn't favor the majority class, leading to more accurate and fair predictions across all sentiment categories.

6. Hybrid Integration

6.1 Combining Rule-based and ML models

This process is performed by using rule base approach and ML based sentiment analysis as follows every intermediate step has as an outcome a list of features as specified run tokenization and tagging process, takes the output from the tokenization, mis spelling errors cleaning, PoS tagging and smart parsing processes, applying some of the semantic rules then extract essential particles convey sentiment/ opinion(adjective, nouns, verbs and adverbs) the sematic orientation or each sentence is calculated some of the information indicate in the [10, 11]

5.3 Sentiment indication table

Sample from one paper

Table 2 List of psychological indicators

Indicators	Description	Examples	# Words
Positive emotions	% of positive emotion words in the review text	Love, nice, sweet	620
Negative emotions	% of negative emotion words in the review text	Hurt, ugly, nasty	744
Differ	% of words in the review text that show differentiation	Hasn't, but, else	81
Drive words	% of words in the review text that are drive words	Friend, social, win	979
Reward	% of words in the review text that are reward words	Take, prize, benefit	120
Risk	% of words in the review text that are risk words	Danger, doubt	103
Words focusing future tense	% of words in the review text that focus future tense	May, will, soon	97
Affect	% of words in the review text that show affective processes	Happy, cried	1393
Auxiliary verbs	% of auxiliary verbs in the review text	Am, will, have	141

7. **Evaluation & Testing**

The implementation of the proposed work is carried out on Anaconda platform using python programming language. The platform is embedded with Natural Language Toolkit (NLTK) and scikit libraries. Various parameters are considered for the evaluation of Modified Logistic Regression (MLR) technique such as classification accuracy, precision, recall, and F-Measure for movie reviews.

We have used TP (True Positives), TN(True Negatives), FP (False Positives) and FN(False Negatives) labels for the same.
[3, 5, 6,10, 14]

Accuracy : It is defend as the rate of correctly classified reviews as positive

$$(TP + TN)/(TP + FP + FN + TN)$$

Recall (R): It is considered to determine the number of true positive function and can be expressed as follow

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

Precision (P): It is defined as the ratio of properly classified over total number of classifications, expressed as follows

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

F1-Score : The harmonic mean of precision and recall, providing a balanced evaluation

$$\text{F1-Score} = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$$

References:

[1] Y. Piris and A.-C. Gay, “Customer satisfaction and natural language processing,” *Journal of Business Research*, vol. 124, pp. 264–271, Jan. 2021, doi: 10.1016/j.jbusres.2020.11.065.

[2] M. Cataltas, S. Dogramaci, S. Yumusak, and K. Oztoprak, “Extraction of product defects and opinions from customer reviews by using text clustering and sentiment analysis,” *DOI: 10.1109*, Dec. 2020, doi: 10.1109/bigdata50022.2020.9377851.

[3] S. Ramaswamy and N. DeClerck, “Customer Perception Analysis using deep learning and NLP,” *Procedia Computer Science*, vol. 140, pp. 170–178, Jan. 2018, doi: 10.1016/j.procs.2018.10.326.

[4] R. Reddy and U. M. A. Kumar, “Classification of user’s review using modified logistic regression technique,” *International Journal of Systems Assurance Engineering and Management*, vol. 15, no. 1, pp. 279–286, Jul. 2022, doi: 10.1007/s13198-022-01711-4.

[5] Y. Abbas and M. S. I. Malik, “Defective products identification framework using online reviews,” *Electronic Commerce Research*, vol. 23, no. 2, pp. 899–920, Jun. 2021, doi: 10.1007/s10660-021-09495-8.

[6] H. Ali, E. Hashmi, S. Y. Yildirim, and S. Shaikh, “Analyzing Amazon Products Sentiment: A comparative study of machine and deep learning, and Transformer-Based techniques,” *Electronics*, vol. 13, no. 7, p. 1305, Mar. 2024, doi: 10.3390/electronics13071305.

[7] M. R. Kounte, P. K. Tripathy, P. P, and H. Bajpai, “Analysis of Intelligent Machines using Deep learning and Natural Language Processing,” *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, Jun. 2020, doi: 10.1109/icoei48184.2020.9142886.

[8] V. W. De Vargas, J. A. S. Aranda, R. D. S. Costa, P. R. Da Silva Pereira, and J. L. V. Barbosa, “Imbalanced data preprocessing techniques for machine learning: a systematic mapping study,” *Knowledge and Information Systems*, vol. 65, no. 1, pp. 31–57, Nov. 2022, doi: 10.1007/s10115-022-01772-8.

[9] J. Zhang, Z. Yin, P. Chen, and S. Nichele, “Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review,” *Information Fusion*, vol. 59, pp. 103–126, Jul. 2020, doi: 10.1016/j.inffus.2020.01.011.

[10] O. Appel, F. Chiclana, J. Carter, and H. Fujita, “A Hybrid Approach to Sentiment Analysis with Benchmarking Results,” in *Lecture notes in computer science*, 2016, pp. 242–254. doi: 10.1007/978-3-319-42007-3_21.

[11] P. Nandwani and R. Verma, “A review on sentiment analysis and emotion detection from text,” *Social Network Analysis and Mining*, vol. 11, no. 1, Aug. 2021, doi: 10.1007/s13278-021-00776-6.

[12] F. M. Plaza-Del-Arco, S. Halat, S. Padó, and R. Klinger, “Multi-Task Learning with Sentiment, Emotion, and Target Detection to Recognize Hate Speech and Offensive Language,” *arXiv.org*, Sep. 21, 2021. https://arxiv.org/abs/2109.10255v4

[13] Tian, X., Vertommen, I., Tsiami, L., Van Thienen, P., & Paraskevopoulos, S. (2022). Automated Customer complaint Processing for water utilities Based on Natural Language Processing—Case study of a Dutch water utility. *Water*, *14*(4), 674. https://doi.org/10.3390/w14040674

[14] Merizig, A., Belouaar, H., Bakhouche, M. M., & Kazar, O. (2024). Empowering customer satisfaction chatbot using deep learning and sentiment analysis. *Bulletin of Electrical Engineering and Informatics*, *13*(3), 1752–1761. https://doi.org/10.11591/eei.v13i3.6966

[15] Iqbal, A., Amin, R., Iqbal, J., Alroobaea, R., Binmahfoudh, A., & Hussain, M. (2022). Sentiment analysis of consumer reviews using deep learning. *Sustainability*, *14*(17), 10844. https://doi.org/10.3390/su141710844

[16] Kaur, G., & Sharma, A. (2024). Automatic customer review summarization using deep learning-based hybrid sentiment analysis. *International Journal of Power Electronics and Drive Systems/International Journal of Electrical and Computer Engineering*, *14*(2), 2110. https://doi.org/10.11591/ijece.v14i2.pp2110-2125