<u>Introduction</u> <u>To</u> <u>Machine</u> <u>Learning</u> - <u>Project</u>
<u>Task-3</u>

<u>Report</u>: - Regarding the critical analysis and the review of the Paper.
<u>Paper</u>: -Thumbs up? Sentiment Classification using Machine Learning Techniques

<u>Student</u>: -
Addepalli Bharat Sai (B19BB002)

We humans are able to classify between any sentiment that is being expressed in the sentences, so by using the responses of each individual we want to try our best to design a model, using which we will be able to classify the responses according to the sentiments, like positive, negative or neutral. Sentiment classification would also be helpful in business intelligence application and recommender systems, where user input and feedback can be quickly summarized.Here, we would like to examine the effectiveness of Machine learning techniques. Thus, sentiment seems to want more understanding than the same old topic-based classification. For our work we wish to choose the dataset of a movie review, as these contain the machine-readable scale i.e., rating indicator like number of stars.

A challenging aspect of this problem seems to distinguish it from the traditional topic based classification is that while topics are often identifiable by the keywords alone, sentiment can be expressed in a more subtle manner, in some thwarted manner which conveys the crisp of the sentiment but it will be difficult for us to tell by selecting some words from sentences. For example, the sentence **“How could anyone sit through this movie?”** contains no single word that is obviously negative.

By using sentiment analysis, you gauge how customers feel about different areas of your business without having to read thousands of customer comments at once.
If you have thousands of feedback per month, it is impossible for one person to read all of these responses. By using sentiment analysis and automating this process, you can easily drill down into different customer segments of your business and get a better understanding of sentiment in these segments.
While sentiment analysis is useful, it is not a complete replacement for reading survey responses. Often, there are useful nuances in the comments themselves. Where sentiment analysis can help you further is by identifying which of these comments you should read.

Thus, sentiment seems to require more understanding than the usual topic-based classification. So, apart from presenting our results obtained via machine learning techniques, we also analyze the problem before by traditional human approach, to gain a better understanding of how difficult it is. Also this requires some tremendous human effort to put forward this request.

From the initial unigram results we obtained from these above methods we see that it beats the human-selected-unigram’s, although here improvement in case of SVM is not large. Also by relying on frequency information we could account for the higher accuracies of Naïve Bayes and SVMs, we binarized the document vectors, setting $n_i(d)$ to 1 if and only feature fi appears in d. As we can see that better performance is achieved from feature presence not from feature frequency. Here bigrams are not effective at our requirements. Now by including parts of speech (POS), the accuracy improves slightly for Naive Bayes but declines for SVMs, and the performance of ME is unchanged. By applying explicit feature-selection algorithms on unigrams could improve performance. Also, unigrams alone didn’t differ much, so by including more defined notions like position in our features we may become successful.

The results produced by using machine learning techniques are quite good as compared to the human-generated baselines as discussed within the presentation. In terms of the relative performance, Naive Bayes tends to try and do the worst and SVMs tend to try and do the most effective, although the differences don't seem to be very large. On the other hand, we weren't able to achieve accuracies on the sentiment classification problem resembling those reported for traditional topic-based categorization, despite the several differing types of features we tried. Unigram presence information turned out to be the foremost effective, in fact, none of the choice features we employed provided consistently better performance once unigram presence was incorporated. Interestingly, though, the superiority of presence information as compared to frequency information in our setting contradicts the previously made observations in the topic based classification work by McCallum and Nigam in 1998.
For example,
“This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can’t hold up” or “I hate the Spice Girls. ...[3 things the author hates about them**]... Why I saw this movie is a really, really, really long story, but I did, and one would think I’d despise every minute of it. But... Okay, I’m really ashamed of it, but I enjoyed it. I mean, I admit it’s a really awful movie ...the ninth floor of hell...The plot is such a mess that it’s terrible. But I loved it.”**

In these examples, a personality would easily detect actual sentiment of the review, but bag of features classifiers would presumably find these instances difficult as there are many words which indicate the other sentiment to it of the whole review. So, it seems that some sort of discourse analysis is important (using some more sophisticated techniques than our positional feature discussed in the presentation), or a minimum of a way of determining the main focus of every sentence in order that one can decide when the author is talking about the film itself. Turney (2002) makes the same point, noting that for reviews, “the whole is not necessarily the sum of the parts”.

Furthermore, it seems likely that this thwarted-expectations device will appear in many varieties of texts dedicated to expressing an overall opinion about some topic. Hence, we believe that a very important next step could be an identification of features indicating whether sentences are on-topic (which is a reasonably co-reference problem); we are looking forward to addressing this challenge in the future work.

References :-

[1] [McCallum and Nigam1998] Andrew McCallum and Kamal Nigam. 1998. A comparison of event models for Naive Bayes text classification. In Proc. of the AAAI-98 Workshop on Learning for Text Categorization, pages 41–48.

[2] [Turney and Littman2002] Peter D. Turney and Michael L. Littman. 2002. Unsupervised learning of semantic orientation from a hundred-billion word corpus. Technical Report EGB-1094, National Research Council Canada.

[3] https://getthematic.com/insights/sentiment-analysis/

[4] https://www.irjet.net/archives/V7/i7/IRJET-V7I7767.pdf