Opinion Mining using Machine Learning

R B Ravi Varma[1], Nagesh Y N[2], I M Umesh[3], Pradeep N C[4]

[1,2,3,4] *R V College of Engineering, Bengaluru, India*

Abstract: **Information contained in opinion can be either subjective or objective or both. Subjective form contains positive or negative opinions, while objective form contains the facts. Identifying the subjectivity and objectivity of information is the outcome of Opinion Mining and Sentiment Analysis. The result will be either positive or negative or a mix of both.**

Machine learning enables the computers to act without being explicitly programmed for a particular task. Applications of the machine learning include self-driving cars, effective web search, practical speech recognition etc.We use machine learning several times a day without knowing it. The developments in this area have resulted in human-level artificial intelligence.

Todesign innovative marketing strategies, opinion mining and sentiment analysis are being used in recent days. Using generated analog data, subjective content is extracted and prediction of subjectivity such as positive or negative is done. This information helps to build systems to understand customer's feedback and plan business strategies accordingly. This also helps in predicting the chances of product failure.In this paper, it is explained how machine learning can be used for opinion mining.

Keywords:**Machine Learning,Opinion Mining, Sentiment Analysis, Classification**

I. INTRODUCTION

Machine learning framework is an integrated system of programs. These programs learn from existing data and capable of predicting new observations. Machine learning deals with the systems study that learns from data, instead of following explicitly programmed instructions. This technique is used in a wide range of computing tasks.

Opinion originates from state of mind, when we experience something in our day to day life. The expression may be an appraisal or a negative comment. Some of the typical techniques to identify and predict the sentiments from the text are Lexicon, Natural Language Processing, Machine Learning based techniques.

In this study, we have used Machine learning based technique to extract opinions of customers and use it for business. The approach is quite straightforward; record customer's opinion, train and classify on selected key words. Similarly, opinion can be predicted by using a pre-populated list of positive and negative words. For example,in the sentence "performance of XYZ Laptop is not good", the word 'good' is a positive word but presence of word 'not' contradicts polar nature of the word. Simple negative and positive word combination creates a negative expression.

II. THE METHODOLOGY

A company plans to suggest a product to potential customers. A database of 10,000 customers exists; 2,000 purchases is the goal. Instead of contacting all the customers, only 1000 customers were contacted. The response is recorded. The subset is used to train a machine learning framework to tell which of the customers decide to buy the product. Then the remaining 9,000 customers are presented to the network which classifies 3,000 of them as potential buyers. The potential buyers are contacted and the goal is achieved.

The system architecture is shown in Fig. 1 which consists of important steps as explained below:

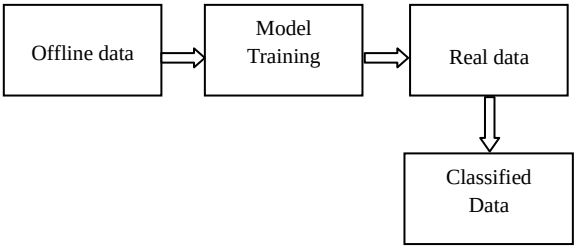

Fig. 1Methodology

The steps to achieve the goals include the following:

1) Speech to text conversion.
2) Pre-processing.
3) Machine Learning Training.
4) Testing and validation.

1. Speech to Text conversion

The conversations of all the customers are recorded and first converted to equivalent text using the speech to text conversion application. This is fed into pre-processing phase as input.

2. Pre-processing

In this step, after speech to text conversion, all the key words are extracted for machine learning training phase. To

increase the search performance, common words are removed from multiple word queries.

3. Machine Learning Training

Key words are used for machine learning training with proper selection of classification function. WEKA machine learning framework was used to build the model. Once the model is built, the opinions can be fed as inputs and the output of the classification gives potential customers.

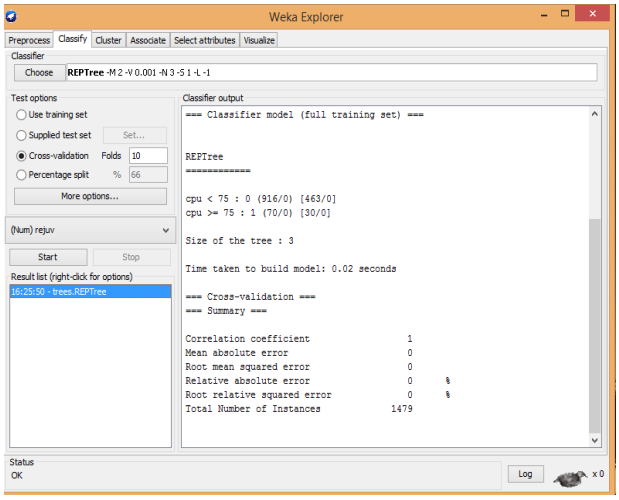

Fig.2WEKA Training Phase

To order the objects in data collection, the classification uses class labels and concept is called supervised classification. Classification approach always uses training set, where all objects are already associated with known class labels. The classification algorithm learns from the training set and builds a model. The customers may fall into any one of the groups called '*Interested*', '*Not Interested*', and '*Decide Later*'.

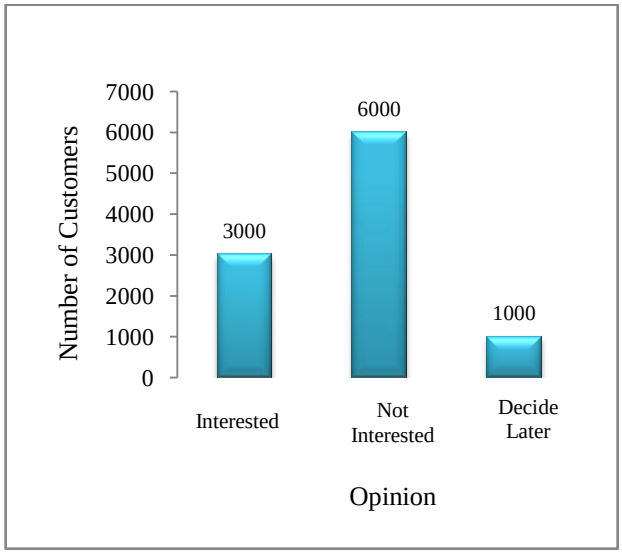

Fig. 3 Graphical representations of customer responses

It is clearly evident through the graphical representation that the out of the 10000 customers 3000 customers are interested, 6000 are not interested and remaining 1000 customers could not decide immediately. Fig. 3 is the visualization of the data mining results which helps better understand the customer needs.

III. Advantages

This pilot project is one of the examples to exploit the machine learning framework and this can be implemented to achieve similar type of goals. Implementation of this type of innovative methods to understand and evaluate customer's opinion regarding particular product will definitely help the top management in a big way to devise innovative marketing strategies leading to better profitability.

The usage of technique is not limited to consumer centric applications but also can be deployed to get clear mandate of the people during election campaigns and accordingly adopt the newer strategies. In the academic circles, the same technique can be used to understand the needs of the student community, accordingly plan, design and disseminate the courseware and embrace new teaching methodologies.

IV. Conclusion

The paper throws a light on application of Artificial Neural Networks in one of latest types of data mining areas i.e., Opinion Mining. Opinion Mining and Sentiment Analysis is the fast emerging area in the field of data mining. The work can be taken forward by implementing the same technique to non-business domains such as politics to find the people's opinion regarding the functioning of the government. This can also be used to find the student's opinion on various key aspects of the academic institution.

References

[1]. Jagannadh G, "Opinion mining and sentiment analysis", an article in CSI communication, May 2012.
[2]. B. Liu, "Sentiment Analysis and Subjectivity."A Chapter in *Handbook of Natural Language Processing*, 2nd Edition, 2010.
[3]. Pang, Bo and Lee, L. (2008). "Opinion Mining and Sentiment Analysis",Foundations and Trendsin Information Retrieval, Vol. 2, Nos. 1–2 (2008)1–135,eBook
[4]. Wiebe, J. Cardie, C. and Riloff, E.(2007). "Manual and Automatic Subjectivity and Sentiment Analysis", Center for Extraction and Summarization of Events and Opinions in Text.University of Utah.
[5]. Almas, Y. and Ahmad, K. (2007). "A note on extracting 'sentiments' in financial news in English, Arabic & Urdu". The Second Workshop on Computational Approaches to Arabic Script-based Languages" LSA 2007 Linguistic Institute , July 21, 2007, Stanford University.
[6]. Leung, C. and Chan, S. (2008). "Sentiment Analysis of Product Reviews". Encyclopedia of Data Warehousing and Mining - 2nd Edition, Information Science Reference, August 2008
[7]. A. Andreevskaia and S. Bergler, "Mining WordNet for a fuzzy sentiment: Sentiment tag extraction from WordNet glosses."

Proceedings of the European Chapter of the Association for Computational Linguistics (EACL), 2006.

[8]. N. Archak, A. Ghose, and P. Ipeirotis, "Show me the money! Deriving the pricing power of product features by mining consumer reviews,"Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2007.

[9]. J. Blitzer, M. Dredze, and F. Pereira,"Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," Proceedings of the Association for Computational Linguistics (ACL), 2007.