International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT1952191

Discovery of Probable Sentiments in Hypertensive Pregnant Women using Horizontal Fragmentation and Pointwise Mutual Information

Dr. Sudhir Tirumalasetty, P. Tejaswini, R Renuka, M Naga Sirisha
Department of Computer Science and Engineering, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India

ABSTRACT

Since a decade research over sentiment analysis and opinion mining was evolving slowing and emerging widely with greater perspectives and objectives. Sentiment analysis is an important task in order to gain insights over the huge amounts of opinions that are generated on a daily basis. This analysis relies on the opinions made by the individuals. These opinions are text, may be positive or negative or a phrase which gives significance to the context. Also these opinions have the power of expressing the context besides drags the attention of new folks. Expressing such opinions ranges from documents level, to the sentence level, to phrase level, to word level and to special symbol level. All these opinion types are labelled with common name Sentiment Analysis. Sentiment Analysis is health care is evolving narrowly with wider research strings. This paper mainly focuses in identifying Sentiments in health care. These sentiments can be medical test values which may be numeric and nominal; sometimes in text too. Such sentiments are identified with pre-fragmentation of data set and Pointwise Mutual Information measure. To accomplish this data of hypertensive pregnant women is considered.

Keywords : Sentiment Analysis, Horizontal Fragmentation, Pointwise Mutual Information

I. INTRODUCTION

Amid the most recent years, the World Wide (Web) has become an immense wellspring of client created content. Numerous social media, for example, Twitter, are viewed as open journals, where a great many individuals communicate, give feelings, and get criticism from different clients. These consistently developing abstract information are, without a doubt, an amazingly rich wellspring of data for any sort of basic leadership process. To computerize the investigation of such information, the region of Sentiment Investigation has developed. It goes for distinguishing opinionative information in the Web and grouping them as indicated by their extremity, i.e., regardless of whether they convey a positive or negative implication.

Most existing opinion investigation calculations were intended for twofold grouping, implying that they allocate sentiments or surveys to bipolar classes, for example, Positive or on the other hand Negative [1]. Notion examination of item surveys utilizing content mining has turned out to be main-stream as of late.

In [1] creators concentrates how we can apply fundamental assumption examination on the surveys put together by the clients for different items. It portrays the different dimensions at which Sentiment investigation can be done. The customary techniques characterize something into positive or negative. This paper presented an idea called multi-point rating scales otherwise called rating derivation which we have joined in our undertaking. The model talked about in this paper accepts an accumulation of

surveys as info, and procedures them utilizing three centre advances, Data Preparation, Review Analysis and Sentiment Classification. The outcomes delivered by such a model are the orders of the surveys, the evaluative sentences, or conclusions communicated in the audits.

In [2] estimation investigation is connected to get data from computerized news articles about its positive or negative opinion with respect to specific legislator. This paper proposes a basic model to break down advanced paper notion extremity utilizing Naive Bayes classifier strategy. The model uses a lot of introductory information in the first place which will be refreshed when new data shows up.

The framework examined in [3] totals the audits of an item from different web sources and utilizing AI calculation would foresee the slant of clients towards the item. The main necessities are that the client should know the particular name of the item he wishes to purchase and just the audits which are in English can be handled.

In [4] authors effectively explored on how item surveys can be characterized dependent on the semantics they convey. This causes us to decrease the overhead of navigating through different locales. This framework right now centers around just content order. Later on, the framework can be prepared to process emojis in audits. Diverse dialects can be joined further into the framework. The productivity of the calculation can be expanded by utilizing troupe strategies

As sentiment analysis is expanding widely in recent evolution in many contexts, its expansion in medical sector is narrower. This paper focuses on identifying sentiments in medical sector. To do this several mining concepts and probabilistic concepts [5, 6, 7, 8, 9, 10] are studied to be a part of this publication.

II. PROPOSED METHOD

Sentiment Analysis is an important task in-order to gain insights over the huge amounts of opinion that are generated on a daily basis. The task of sentiment analysis is to label people's opinions as different categories such as positive and negative from a given piece of text. Another task is to decide whether the given text is subjective or objective expressing. All these tasks are under the umbrella of Sentiment Analysis. The project "Discovery of Probable Sentiments in Hypertensive Pregnant Woman using Horizontal Fragmentation and Point-wise Mutual Information" focuses identifying sentiments in health care. These Sentiments can be medical test values which may be numeric and nominal; sometimes in text too. Such sentiments are identified with pre-fragmentation of data set and pointwise mutual information measure. To accomplish this, data of hypertensive pregnant woman is considered. Source of data set if Guntur Medical College, Guntur, Andhra Pradesh, India.

Step wise procedure for attaining sentiments are as follows:

1) The main theme is Diabetes in Pregnant Women
2) Consider
    25 samples--- without Diabetes
    25 samples--- with Diabetes
3) Initially, prefragmenting the samples as
    5 samples--- with diabetes
    5 samples---- without diabetes
    10 samples--- with diabetes
    10 samples---- without diabetes
    15 samples--- with diabetes
    15 samples---- without diabetes
    20 samples--- with diabetes
    20 samples---- without diabetes
    25 samples--- with diabetes
    25 samples---- without diabetes
4) For each set of samples for all the attributes calculate the following:

a. Average- which is the sum of all samples for every attribute divided by total number of samples.
b. Max- Maximum vavlue is the value highest among all values in every attribute w.r.t samples
c. Min- Maximum value is the value highest among all values in every attribute w.r.t samples
d. Standard Deviation- It is calculated using all individual sample values and average of all the samples.

5) For the same set of above samples for all attributes have to calculate Gaussian Distribution.

To calculate it we need Average and standard deviation for every set of samples.

6) After attaining gaussian distribution table:

Calculate average which is the central value for every attribute for all set of samples.

7) Point-wise Mutual Information:

a. Total all the central values of Gaussian Distribution and make it as a gaussian probability.
b. As mentioned that there are only 2 cases i.e., with or without diabetes; we take the probability of diabetes as 0.5(as the chance of diabetes is 50% - it is considered as a fraction of 0.5)
c. In-order to calculate PMI, initially divide each average of Gaussian distribution with total gaussian probability.
d. Next multiply the obtained value with 0.5.

$$PMI(X,Y) = \log\left(\left(\frac{GD(A)}{\sum_{i=1}^{k} GD(A_i)}\right) * 0.5\right)$$

Where k is number of attributes, X is an attribute (one among in k attributes), Y takes either 'diabetes' or 'no diabetes', GD is Guassian Distribution Probability, A is an attribute.

III. RESULTS AND DISCUSSION

14 Medical parameters of pregnant women are taken into account in this paper. They are: Age, Gestational Age (Gage), Na (Sodium levels), K (Potassium levels), Urea levels, Uric acid levels, Total Proteins (TProteins), Fasting Blood Sugar (FBS), Total Cholesterol (TCH), HDLC, TTG, Systolic, Diastolic and Weight.

PMI values for each parameter is calculated with fragments of varying size (5, 10, 15 and 25) with respect to pregnant women having 'diabetes' and 'no diabetes'.

The following Figures 1, 2, 3 and 4 reveals about the variations of PMI values with respect to each fragment of mentioned sizes.

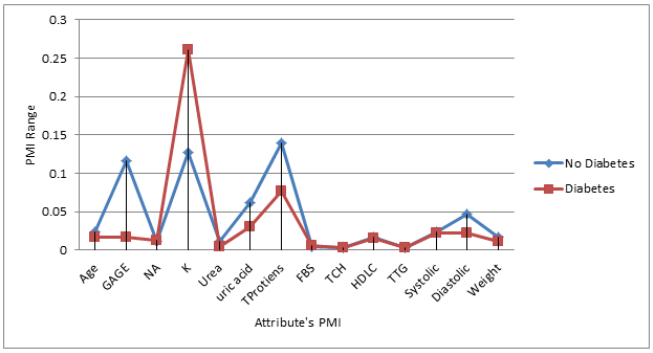

Figure 1. PMI values comparison with fragments of size 5 with 'diabetes' and 'no diabetes'

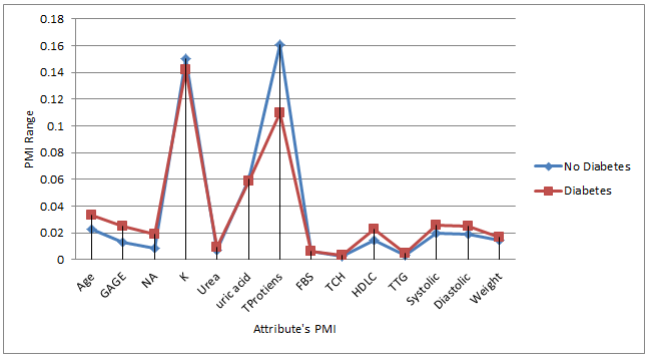

Figure 2. PMI values comparison with fragments of size 10 with 'diabetes' and 'no diabetes'

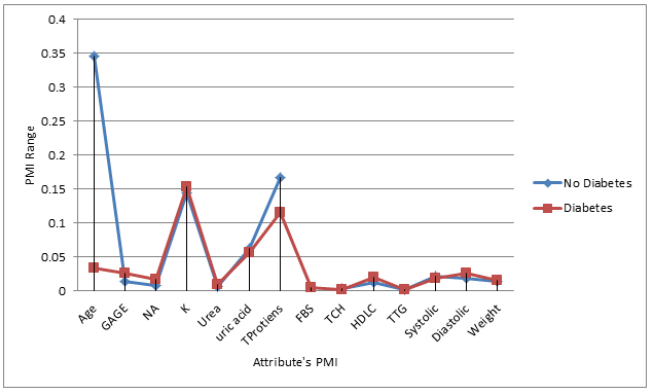

Figure 3. PMI values comparison with fragments of size 15 with 'diabetes' and 'no diabetes'

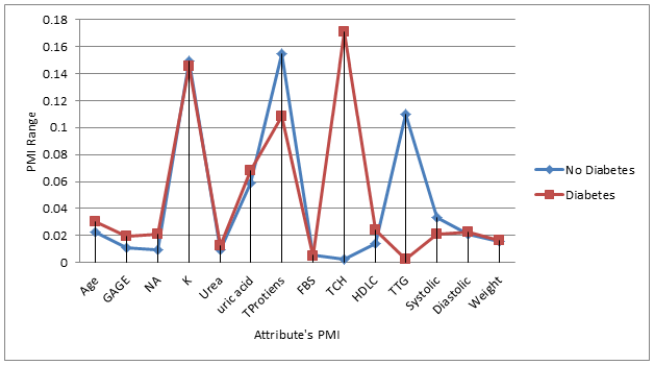

Figure 4. PMI values comparison with fragments of size 25 with 'diabetes' and 'no diabetes'

From the above figures it is evident and infers that increase in Potassium levels (K) and decrease in Total Proteins (TProteins) always drives a pregnant women to diabetes. In some cases this also leads to increase in Triglycerides (TCH) i.e. total cholesterol, which is a risk factor to Pregnant Women, may lead to cardiac problems.

IV. CONCLUSION

In our framework, we proposed an information mining procedure to pin attributes which influence pregnant women. The input put together by understudies is the best approach to think for formulating a way of identifying features in Health Care. As research expansion in health care sector is narrower, this paper provided a novel approach of finding features, which influences pregnant women. This is an initiation for promoting sentiment analysis in health sector.

V. REFERENCES

[1]. Cane W.K. Leung & Stephen C.F. Chan. Sentimental Analysis of Product Reviews. The Hong Kong Polytechnic University, Hong Kong SAR, 2008.

[2]. Yustinus Eko Soelistio and Martinus Raditia Sigit Surendra. Simple Text Mining For Sentiment Analysis of Political Figure Using Naive Bayes Classier Method. CoRR, 2015.

[3]. Jyoti Jain, Asst. Prof. Archana Shinde & Prachi Panchal, Sentiment Analysis using Machine Learning. Department of Information Technology, Sinhgad Academy of Engineering, Maharashtra, India, 2016.

[4]. Krishnaveni K S, Rohit R Pai and Vignesh Iyer, "Faculty Rating System Based on Student Feedbacks Using Sentimental Analysis", 978-1-5090-6367-3/17/, IEEE, 2017.

[5]. E.J. Fortuny, T.D. Smedt, D. Martens & W. Daelemans , Media Coverage In Times of Political Crisis: A Text Mining Approach, Expert Systems with Applications, Sciendirect, 2012.

[6]. G.F. Luger. Artificial Intelligence Structures And Strategies For Complex Problem Solving, Pearson Education Inc., 182-185., 2009.

[7]. Yamanishi, K., & Li, H. Mining open answers in questionnaire data. IEEE Intelligent Systems, 17(5),pp. 58-63, 2002.

[8]. Jiawei Han, Micheline Kamber & Jian Pei. Data Mining Concepts and Techniques. Waltham: Morgan Kaufmann Publishers, 2012.

[9]. Liu, B., Hsu, W., & Ma, Y. Integrating classification and association rule mining. Proceedings of Knowledge Discovery and Data Mining, pp. 80-86, 1998.

[10]. C. P. Prathibhamol, Ashok, A. Solving multi label problems with clustering and nearest neighbor by consideration of labels, Advances in Intelligent Systems and Computing, vol. 425, pp. 511-520, 2016.