A Project Based Seminar Report

on

"Review Based Sentiment Analysis using Machine Learning"

Submitted to the

Savitribai Phule Pune University

In partial fulfillment for the award of the Degree of

Bachelor of Engineering

in

Information Technology

by

Rohan Dighe

Roll No: 3024

Class: TE-10

Under the guidance of

Prof. J.K. Kamble

Department of Information Technology

Pune Institute of Computer Technology

Sr. No 27, Pune-Satara Road, Dhankawadi, Pune - 411 043

2018-2019

CERTIFICATE

This is to certify that the project based seminar report entitled "**Review Based Sentiment Analysis Using Machine Learning**" being submitted by **Rohan Dighe (T150058537 / 3024** & **TE-10)** is a record of bonafide work carried out by him/her under the supervision and guidance of **Prof. J. K. Kamble** in partial fulfillment of the requirement for **TE (Information Technology Engineering) – 2015 course** of Savitribai Phule Pune University, Pune in the academic year 2018-2019.

Date: 15/03/2018

Place: Pune

Prof. J. K. Kamble
Guide

Dr. B. A. Sonkamble
Head of the Department

Dr. P. T. Kulkarni
Principal

This Project Based Seminar report has been examined by us as per the Savitribai Phule Pune University, Pune requirements at Pune Institute of Computer Technology, Pune – 411043 on

. . . . . .

Internal Examiner

External Examiner

ACKNOWLEDGEMENT

It is indeed a matter of great pleasure and privilege to present the Project on the "**Review Based Analysis of Pre-Launched Product**" . This is to acknowledge and thank to all the individuals who played defining role in shaping this Project. Without their constant, Guidance and assistance this Project based Seminar report would not have been completed.

We express our sincere thanks to our respected guide **Mr. J. K. Kamble** for their kind cooperation and timely guidance for Presenting the Report well in time and thanks to for constant guidance and his boundless cooperation and help extended for this Project, all through our work.

We also express our gratitude and thanks to **all the Staff Member of IT Department** for providing the necessary facilities for completion of this project work in our college. And last, but not least huge thanks to all our friends and our family members for their help in the successful completion of this project.

Thank you.

(Rohan Dighe)

ABSTRACT

Sentiment analysis is one of the fastest growing research areas in computer science, making it challenging to keep track of all the activities in the area. Today internet has become the major part of our life. Most of the people use online blogging sites or social networking sites to express their opinions on certain things. They also use these sites to know what other people's opinions are. Thus mining of this data and sentiment extraction has become an important field of research.In recent years, sentiment analysis has shifted from analyzing online product reviews to social media texts from Twitter and Facebook. Many topics beyond product reviews like stock markets, elections, disasters, medicine, software engineering and cyberbullying extend the utilization of sentiment analysis. he outbreak of computer-based sentiment analysis only occurred with the availability of subjective texts on the Web.In modern world the development of web and smartphones increases the usage of online shopping. The overall feedback about product is generated with the help of sentiment analysis using text processing.Opinion mining or sentiment analysis is used to collect and categorized the reviews of product. The proposed system uses aspect level detection in which features are extracted from the datasets. The system performs pre-processing operation such as tokenization, part of speech and limitization on the data tofinds meaningful information which is used to detect the polarity level and assigns rating to product. The proposed model focuses on aspects to produces accurate result by avoiding the spam reviews.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction to project

The numbers of people that buy the products and review them online have increased tremendously. Product reviews are helping customers make decisions. These reviews may be positive, negative or neutral in nature.So we are planning to make an application which gathers the review about pre-launched products from customer and we will providing solutions about the product and statistical data to the organization. Project Flow diagram is given below:

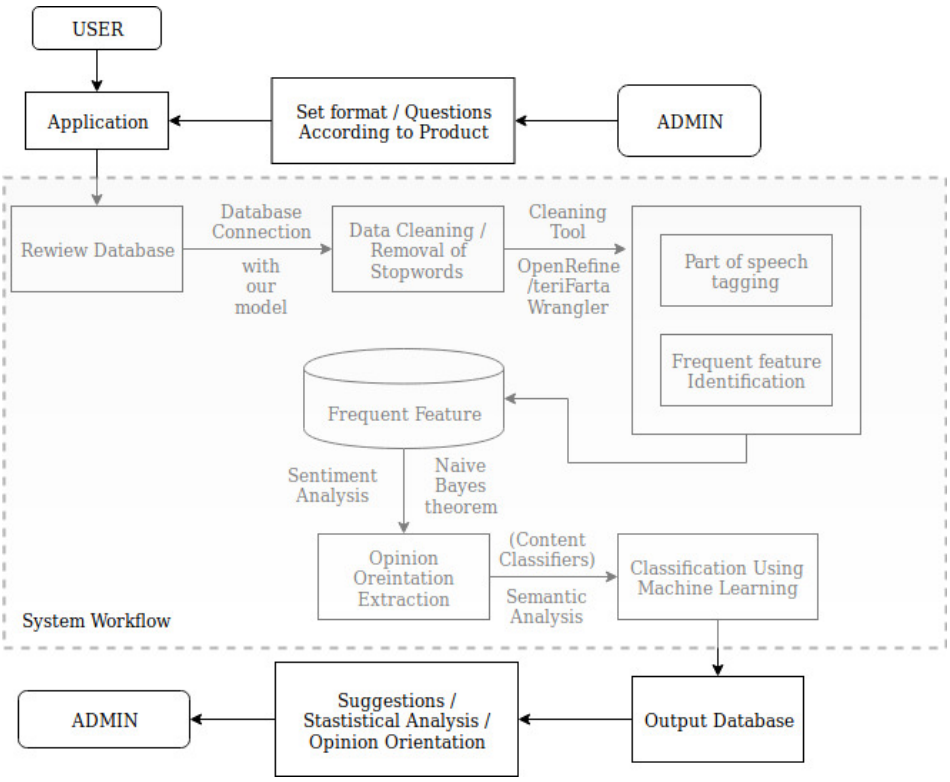

Figure 1.1: Fig.1 of Project Flow Diagram for Review Based Analysis of Pre-Launched Product

1.2 Motivation behind the project topic

Important information can be obtained from reviews which consists of user's feedback about the product. Reviews can guide the customers into buying the right product, help choosing between 2 products.

1.3 Aim and Objective(s) of the work

Project Aim - To develop an application using Python to perform data analysis based on user reviews on pre-launched product.

Project Objectives -

1) To minimize the cost of production
2) To suggest changes about product to manufacturer
3) To provide summary about product

1.4 Introduction to seminar topic

The growth of social web contributes vast amount of user generated content such as customer reviews, comments and opinions. This user generated content can be about products, people, events, etc. This information is very useful for businesses, governments and individuals. While this content meant to be helpful analyzing this bulk of user generated content is difficult and time consuming. So there is a need to develop an intelligent system which automatically mine such huge content and classify them into positive, negative and neutral category. Sentiment analysis is the automated mining of attitudes, opinions, and emotions from text, speech, and database sources through Natural Language Processing (NLP). The objective of this paper is to discover the concept of Sentiment Analysis in the field of Natural Language Processing, and presents a comparative study of its techniques in this field.

Sentiment Analysis :

Sentiment Analysis is the study of people opinions, attitudes and emotions toward an entity. The entities are the event, aspect or any other that concerned with people opinion. They express some type of mutual meaning. Some of the researcher stated that Opinion Mining is using to extracts the people's opinion about the any entity or aspect. On the other side Sentiment Analysis identifies the sentiment that expressed into the text then analyzes it. So, the target of SA is to find opinions, identify the sentiments they express, and then classify their polarity i.e. positive, negative or neutral. Sentiment Analysis is the classification process. The classification process are divided into the several part depends on there working these are Document -Level, sentence-level and the aspect level.

CHAPTER 2
LITERATURE SURVEY

Selected Paper :

Study of supervised machine learning approaches for sentiment analysis.

Keywords Used:

1. Naive Bayes
2. Support Vector Machine (SVM)
3. Maximum Entropy (ME)
4. Sentiment,Opinion

Published Year: Apr-2018 (IEEE)

Summary:

In this paper, they have been provided that Text extraction is a crucial stage of analyzing Journal papers. Journal papers generally are in PDF format which is semi structured data. Journal papers are presented into different sections like Introduction, Methodology, Experimental setup, Result and analysis etc. so that it is easy to access information from any section as per the reader's interest. The main importance on section extraction is to find a representative subset of the data, which contains the information of the entire set. Various approaches to extract sections from research papers include stastical methods, NLP, Machine Learning etc. In this paper we present review of various extraction techniques from a PDF document.

Selected Paper :

Product Analysis Based on user Reviews

Methodology Used:

1. Semantic Analysis
2. Sentiment Analysis
3. Data Extraction

Published Year: 2017 (IEEE)

Summary:

In the proposed system we analyze the reviews based on user's sentiments. The system also generates reports for the manufacturers as well as customers. A dashboard of all the products will be created wherein the user can view the product and its ratings based on reviews, also the user can get to know about the features of the product. Thus, it would be easier for users looking out for a specific set of features, to make a correct choice of a product. Also, due to feature based product analysis, manufacturers can get to know the exact reason of the failure of their product. Manufacturers can identify the features which were not found to be up to the mark based on the reviews, and thus work upon them to deliver a better product. Since the opinion of a product changes over time, timeline analysis is a key concept introduced. The user can view the performance of the product over time and can compare it with similar products available in the market. Since our analysis is totally review based, a product's popularity would be a reflection of the product's quality. With declining popularity, users can look for better products which serve the same purpose. The system also gives product recommendation to customer if the currently viewed product has negative opinion overall. A lso by keeping a track of the kind of products a user reviews, we aim to build a recommendation engine that could suggest similar products to the user based upon his liking for products.

Sr. No.	Author	Year	Title	Contents of Paper
1.	Sangharshjit S. Kamble1, Prof. A. R. Itkikar2 Dr. Sanjeev Sharma	Apr-2018	Study of supervised machine learning approaches for sentiment analysis	NB, SVM, ME, Sentiment, opinion graph, etc.
2.	A.Helen Victoria, M.Vijayalakshmi Assistant Professor, SRM University,	2016	An Outcome Based comparative Study of Different Text Classification Algorithms	Information Retrieval, K-nearest Neighbor, Decision Trees, Support vector Machine etc.
3.	R. Ragupathy , Lakshmana Phaneendra Maguluri	2018	Comparative analysis of machine learning algorithms on social media test	Sentimental Analysis, Social Reviews, Text Pre-processing, Sentiment Score, Machine Learning Techniques, Comparative Study etc.
4.	Mr. S. M. Vohara, Prof. J. B. Terayia	Oct-2013	A Comparative Study of Sentiment Analysis Techniques	Natural Language Processing, Sentiment Analysis, Sentiment Lexicon, Sentiment Score etc.
5.	Mohammad Mohaiminul Islam,Naznin Sultana	Oct-2018	Comparative Study on Machine Learning Algorithms for Sentiment Classification	Natural Language Processing, Sentiment Analysis, Opinion mining, Machine Learning.

Table 2.1: Comparison between diffent selected papers

CHAPTER 3

METHODOLOGY AND ALGORITHMS

3.1 Techniques ::

3.1.1 Machine Learning Based Techniques

- Naive Bayes (NB)
- maximum entropy (ME)
- support vector machines (SVM)

3.1.2 Lexicon Based Techniques

- Sentiment lexicon contains lists of words and expressions used to express people's subjective feelings and opinions
- For example, start with positive and negative word lexicons, analyze the document for which sentiment need to find. Then if the document has more positive word lexicons, it is positive, otherwise it is negative

3.2 Machine Learning Based Techniques :

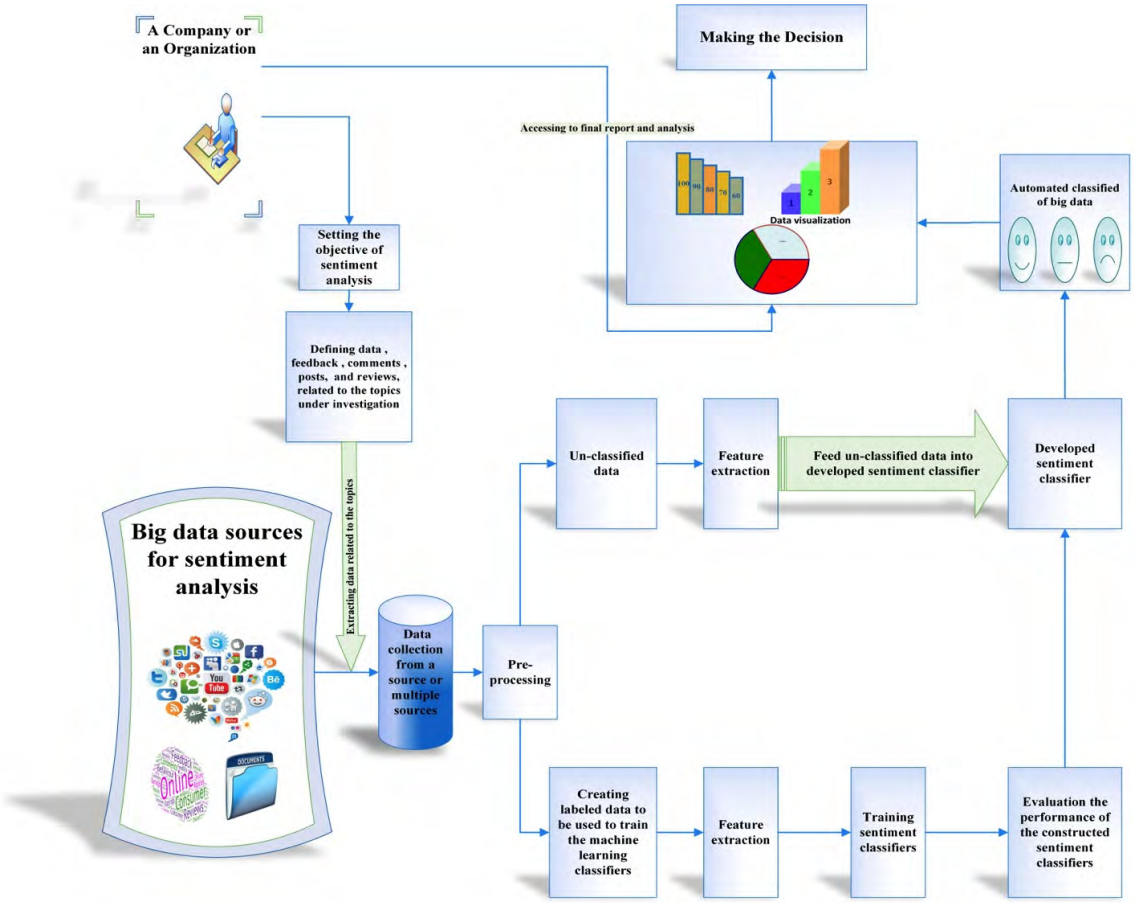

Figure 3.1: Steps for constructing machine learning based method for Sentiment Analysis

Steps for constructing machine learning based method for Sentiment Analysis

3.2.1 Comparative Study of ML Techniques

Comparison among Algorithms	**Decision Tree**	**Naïve Bayes**	**KNN**	**Support Vector Machine**
Input Type	Continuous& Categorical Values	Gaussian Naïve Bayes can deal with Continuous Values	Continuous	Continuous
Missing or Noisy Data	Can deal with Noisy or Incomplete Data	Good Resistance to missing or noisy data	Approximates the missing value by the value nearest to it	Performs poorly When data set has more noise
Error Rate	Error rate is high because of overfitting and under fitting when the sample size is small	When data size is huge it may double count features since it assumes events to be independent.	Error rate depends on the choice of K value in KNN	Error rates are high when svm doesn't filter out the samples that is not sure about.
Best	Best for Larger Samples	Works best for smaller samples /Training Data	Best when Sample size is small	uitable for small and medium data sets. Performs poorly when attributes are less.
Computational Demand & Memory Requirements	Best compared to others, data is classified without much calculations.	Better results compared to others.	Larger the data set bigger the Memory Requirements. It has costly testing phase.	Memory Efficient

Table 3.1: Comparative Study of Different Machine Learning Techniques

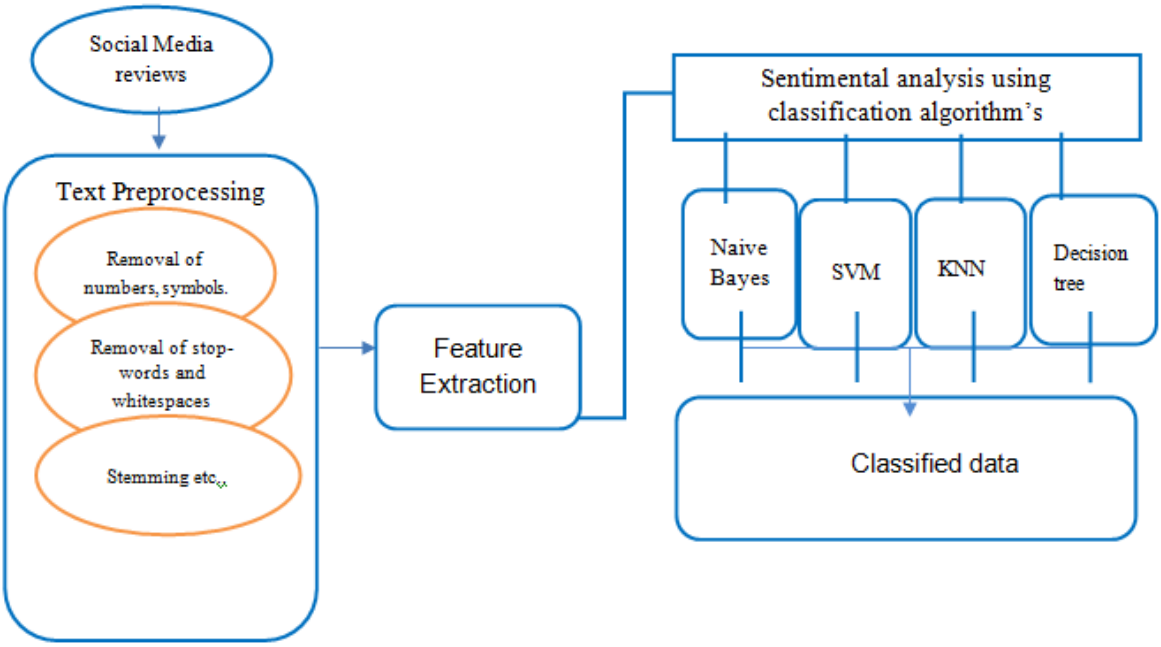

Figure 3.2: Outline of the Comparative Model of Sentiment Analysis

Outline of the Comparative Model of Sentiment Analysis

CHAPTER 4

ADVANTAGES AND DISADVANTAGES

Sr.No.	Machine Learning Classifier	Advantage	Disadvantage
1.	KNN	It is simple and also used for multiclass categorization Of document.	It requires more time to ategorize when huge number data are inclined. Takes lot of memory for running a process.
2.	Decision Tree	This is very fast in learning data set. Easy for understanding purpose.	It has problem that it is difficult handle data with noisy data.Over fitting of data.
3.	Naïve Bayesian	Simple and work well with textual as well as numerical data. Easy to implement. Computationally cheap.	Performs very poorly when feature set Is highly correlated. It gives relatively low classification performance for large data set.
4.	Support Vector Machine	High accuracy even with large data set. Works well with many number of dimensions. No over fitting.	Problems in representing document into numerical vector.

Table 4.1: Comparison between Machine Learning Methods

CHAPTER 5

APPLICATIONS

- Sentiment analysis has many applications and benefits to your business and organization. It can be used to give your business valuable insights into how people feel about your product brand or service.
- When applied to social media channels, it can be used to identify spikes in sentiment, thereby allowing you to identify potential product advocates or social media influencers.
- It can be used to identify when potential negative threads are emerging online regarding your business, thereby allowing you to be proactive in dealing with it more quickly.
- Sentiment analysis could also be applied to your corporate network

CHAPTER 6
ENHANCEMENTS

In this report, we conducted a comprehensive and systematic survey on Sentiment Analysis using machine learning techniques. In this review article I am try to cover these three supervised machine learning algorithm. In this study we learn what exactly the sentiment analysis is and the various approaches that are use for sentiment analysis. These two are unsupervised and supervised machine learning approach. Further supervised machine learning is use two types of data set i.e. Train data set and test data set.

CHAPTER 7

CONCLUSION

Sentiment analysis (or) opinion mining plays a significant role in business decision making. Many of the organization and enterprises will take their business decision only based on their customer review. There are several techniques for performing sentiment analysis. This paper specifies the sentiment analysis under 3 categories such as Machine learning, dictionary based, Ontology based and gives the clear knowledge about various approaches. This survey gives the knowledge about the sentiment analysis issues such as Polarity shift problem, data sparsity, binary classification briefly and how they are handled in different domains.in this paper we have provided an insight about when to use which classification algorithm assuming that the information of the dataset is fully understood. This might help analysts to choose a better classification algorithm.

CHAPTER 8

BIBLIOGRAPHY

[1] Blinov P. D., Klekovkina M. V., Kotelnikov E. V., Pestov O. A. "Research of lexical approach and machine learning methods for sentiment analysis ".

[2] B. Pang, L. Lee, and S. Vaithyanathan, Thumbs up? Sentiment Classification using Machine Learning Techniques, Proc. Of the Conf. on Empirical Methods in Natural Language Processing (EMNLP)

[3] Doaa Mohey El-Din Mohamed Hussein" A survey on sentiment analysis challenges" , In Ain Shams Engineering Journal (2014) 5, April 2016

[4] David Osimo and Francesco Mureddu "Research Challenge on Opinion Mining and Sentiment Analysis.

[5] Kumar Ravi, Vadlamani Ravi "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications" In Knowledge-Based Systems 89 (2015)

[6] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, Manfred Stede "Lexicon-BasedMethods for Sentiment Analysis"