

A Review on Sentiment Analysis and Visualization of Customer Reviews

Apurva V. Gundla¹, Manisha S. Otari²

¹Department of Computer Science and Engineering, N.K.Orchid College of Engineering, Solapur, Maharashtra, India

²Professor, Department of Computer Science and Engineering, N.K.Orchid College of Engineering, Solapur, Maharashtra, India

Abstract: *The Web has become an excellent source for gathering consumer opinions. There are now numerous Web sites containing such opinions, e.g., customer reviews of products, forums, discussion groups, and blogs. All these reviews are the opinions of people all over the world about different products. With the growing availability and popularity of opinion-rich resources such as review forums for the product sold online, choosing the right product from a large number of products have become difficult for the user. These sources are underutilized both by consumers and businesses due to their unstructured nature, serial presentation, limited search tools, and low ratio of useful information to the overall amount of data. The proposed system illustrates visual analysis system that performs sentiment analysis and derives insight from a collection of online reviews of products from customer. Effective visual analysis of online customer opinions is needed, as it has a significant impact on building a successful business and helps the customers in decision making process. This paper presents background Study of Sentiment Analysis or Opinion Mining and gives overview of proposed methodology with insights into past research work.*

Keywords: Customer Reviews, Opinions, Sentiment Analysis, Opinion Mining, Review Forums.

1. Introduction

Sentiment Analysis refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis aims to determine the attitude of a speaker or a consumer with respect to some topic or product. [1]

Sentiment analysis, also called *opinion mining*, is the field of study that analyses people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. [2]

It involves techniques from different disciplines like information retrieval, Natural Language Processing and Data Mining. Sentiment Analysis is about extracting the opinions or sentiments when given a piece of text.

"What other people think" has always been an important piece of information for most of us during the decision-making process. [3] Opinions or sentiments are central to almost all human activities and are key influencers of our behaviors. Our beliefs and perceptions of reality, and the choices we make, are, to a considerable degree, conditioned upon how others see and evaluate the world. Whenever we need to make a decision, we want to know others' opinions. [2]

An Opinion is a judgment or belief a majority of people formed about a specific thing, not necessarily based on fact/knowledge. Opinion generally refers to what a person thinks about something or opinion is a subjective belief, and the result of emotion or facts interpretation. [4]

In the real world, businesses and organizations always want to find consumer or public opinions about their products and services. Individual consumers also want to know the opinions of existing users of a product before purchasing it, and others' opinions about political candidates before making a voting decision in a political election. In the past, when an individual needed opinions, he/she asked friends and family. When an organization or a business needed public or consumer opinions, it conducted surveys, opinion polls, and focus groups. Acquiring public and consumer opinions has long been a huge business itself for marketing, public relations, and political campaign companies. This unique feature plays a vital role in determining on matters that have financial, medical, social or other implications. Seeking second or third or many more opinions have fuelled the interest of researchers in the field of sentiment mining. [5]

The World Wide Web is growing at an alarming rate not only in size but also in the types of services and contents provided. The Web has provided an excellent platform for business to consumer (B2C) electronic commerce. B2C electronic commerce offers convenience, choice, lower cost and customization to consumers. With the rapid expansion of e-commerce platform, social media, online stores and review sites over the last decade, online shopping of the products has enhanced drastically. Due to the vast variety of products and convenient shopping experience with attractive offers, these platforms have become popular for customers and even for manufacturers. As more people using the Web to express opinions, the number of reviews that a product receives is growing rapidly. For some popular products, the number of reviews can be in hundreds or more. These reviews provide excellent sources of consumer opinions on products. These platforms have become popular for customers and even for manufacturers. This is some first-hand experience shared among consumers to either promote or reject the product which can affect the decision of other potential buyers.

Web users convey particular sentiments and views on very nearly anything at review sites, blogs, and forums and so on. It is nothing but a electronic word-of-mouth. These reviews provide excellent sources of consumer opinions on products. This important data is free, easily accessible for internet clients.

Many reviews are also long, which makes it hard for a potential customer to read them to make an informed decision on whether to purchase the product. If he/she only reads a few reviews, he/she only gets a biased view. The large number of reviews also makes it hard for product manufacturers or businesses to keep track of customer opinions and sentiments on their products and services. The large number of reviews also makes it hard for product manufacturers or businesses to keep track of customer opinions and sentiments on their products and services. [6] The substantial gathering of opinions on the Web makes it extremely tough to get helpful data effectively. Perusing all reviews and emotions to settle on an educated choice is a much time taking task. Perusing distinctive and potentially even conflicting opinions composed by diverse commentators may make organizations, users and customers more confused. It is thus highly desirable to produce a visualization of reviews.

The proposed system aims to let users gain useful information for decision making as quickly and as effortlessly as possible, by transforming large collections of reviews text into visualizations that provide the same conceptual understanding that would otherwise require the reading through the whole text collection.

The paper is organized as follows: Section 2 gives gist of Motivation and Related work. Section 3 provides the overview of the Proposed Methodology. Section 4 concludes the manuscript.

2. Motivation and Related Work

Sentiment Analysis refers to identification and classification of the viewpoint or opinion expressed in the text span; using information retrieval and computational linguistics. Sentiment analysis or opinion mining extracts the subjective information from the source materials such as reviews using techniques such as natural language processing, and text analytics.

Online reviews have become an important source of information for both producers and consumers, with companies trying to better understand customer-provided feedback on products and brands, and individual users looking for information to support their everyday purchasing decisions. Given the widespread use of computers and mobile devices, most of which are connected to the Internet, more and more people are sharing their thoughts, feelings, and experiences. This growing amount of online opinionated information has led to the rapid development of the field of sentiment analysis, which focuses on the identification of opinions, emotions, evaluations, and judgments, along with their polarity positive or negative.

Related Work

Jeonghee Yi and Nasukawa[7] first used the term sentiment analysis in their paper Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques that *i)* extracts topic-specific features, *ii)* extracts sentiment of each sentiment-bearing phrase, *iii)* makes (topic| feature, sentiment) association. They extract only noun phrases from documents and apply two feature term selection algorithms such as Mixture Model and Likelihood test.

Minqing Hu and Bing Liu [8] studied the problem of feature-based opinion summarization of customer reviews of products sold online. The task is to identify the features of the product that customers have expressed opinions on (called *opinion features*) and rank the features according to their frequencies that they appear in the reviews. For each feature, identify how many customer reviews have positive or negative opinions. The specific reviews that express these opinions are attached to the feature. This facilitates browsing of the reviews by potential customers.

Kushal Bafna, Durga Toshniwal [9] proposed a dynamic system for feature based summarization of customers' opinions for online products. Online reviews for a product are extracted on periodic bases; each time after extraction, identification of features of a product from customers' opinions is done. Next, for each feature, its corresponding opinions are extracted and their polarity (positive/negative) is detected. The final polarity of feature-opinions pairs is calculated. At last, feature based summarizations of the reviews are generated, by extracting the relevant excerpts with respect to each feature-opinions pair.

Bing Liu, Minqing Hu and Junsheng Cheng [10] proposed an analysis system with a visual component to compare consumer opinions of different products. The system is called *Opinion Observer*. With a glance of its visualization, the user can clearly see the strengths and weaknesses of each product in the minds of consumers.

Xiaowen Ding, Bing Liu and Philip S. Yu [6] represented a holistic lexicon-based approach. Given a set of product features of a product, we want to accurately identify the semantic orientations of opinions expressed on each product feature by each reviewer. Semantic orientation means whether the opinion is positive, negative or neutral.

Bo Pang et al., [11] used machine learning techniques to investigate the effectiveness of classification of documents by overall sentiment. Experiments demonstrated that the machine learning techniques are better than human produced baseline for sentiment analysis on movie review data. Features based on unigrams and bigrams are used for classification. Learning methods Naïve Bayes, maximum entropy classification and support vector machines were employed. Inferences made by Pang et al., is that machine learning techniques are better than human baselines for sentiment classification.

Zhu et al., [12] proposed aspect based opinion polling from free form textual customers reviews. The aspect related terms used for aspect identification was learnt using a multi-aspect bootstrapping method.

Kamal et al [13] implemented a rule based system to mine product features, opinions and their reliability scores. The proposed system uses linguistic and semantic analysis of text to mine the feature opinion pairs from review documents.

Draper and Riesenfeld [14] developed an interactive visualization system to allow users to visually construct queries on large tabular data sets and view results in real time.

Morinaga et al. [15] suggested a 2D scatter plot called positioning map to show the group of positive or negative sentences.

Opinion Mining Terminology

1) Components of Opinions:

- *Opinion Holder*: Opinion holder is the person or organization that expresses the opinion.
- *Opinion Object*: It is a feature about which the opinion holder is expressing his opinion.
- *Opinion Orientation*: Determine whether the opinion about an object is positive, negative or neutral.

2) Different Levels of Sentiment Analysis:

a. Document Level Sentiment Analysis:

The task at this level is to classify whether a whole opinion document expresses a positive or negative sentiment. For example, given a product review, the system determines whether the review expresses an overall positive or negative opinion about the product. This task is commonly known as *document-level sentiment classification*. This level of analysis assumes that each document expresses opinions on a single entity (e.g., a single product). Thus, it is not applicable to documents which evaluate or compare multiple entities. [2]

The document level sentiment classification has its own advantages and disadvantages. Advantage is that we get an overall polarity of opinion text about a particular entity from a document. Disadvantage is that the different emotions about different features of an entity could not be extracted separately.

b. Sentence Level Sentiment Analysis:

The task at this level goes to the sentences and determines whether each sentence expressed a positive, negative, or neutral opinion. Neutral usually means no opinion. This level of analysis is closely related to *subjectivity classification* which distinguishes sentences (called *objective sentences*) that express factual information from sentences (called *subjective sentences*) that express subjective views and opinions.[2] The advantage of sentence level analysis lies in the subjectivity/objectivity classification. The traditional algorithms can be used for the training processes.

c. Phrase Level or Aspect Level Sentiment Analysis:

The phrase level sentiment analysis is also called as Feature based sentiment analysis. Aspect level directly looks at the opinion itself. It is based on the idea that an opinion consists of a *sentiment* (positive or negative) and a *target* (of opinion). An opinion without its target being identified is of limited use. Realizing the importance of opinion targets also helps us understand the sentiment analysis problem better. The goal of this level of analysis is to discover sentiments on entities and/or their aspects. [2]

3) Subjectivity/ Objectivity Classification:

The task of determining whether a sentence is subjective or objective is called subjectivity classification. The text pieces may or may not contain useful opinions or comments. The subjective sentences are the relevant texts, and the objective sentences are the irrelevant texts. So we must sort out the sentences that are useful for us and those which are not.

An objective sentence presents some factual information about the world, while a subjective sentence expresses some personal feelings, views, or beliefs. A subjective sentence may not express any sentiment. Objective sentences can imply opinions or sentiments due to desirable and undesirable facts. An example objective sentence is "iPhone is an Apple product."

An example subjective sentence is "I like iPhone."

4) Types of Opinion:

a) Explicit Opinion:

An explicit opinion is a subjective statement that gives a regular or comparative opinion. An explicit opinion on feature *f* is a subjective sentence that directly expresses a positive or negative opinion. The following sentence expresses an explicit positive opinion:
"The picture quality of this camera is amazing."

b) Implicit Opinions:

An implicit opinion is an objective statement that implies a regular or comparative opinion. Such an objective statement usually expresses a desirable or undesirable fact. An implicit opinion on feature *f* is an objective sentence that implies an opinion. The following sentence expresses an implicit negative opinion:

"The earphone broke in two days."

Although this sentence states an objective fact (assume it is true), it implicitly expresses a negative opinion on the earphone.

c) Data Sources

Now a day there are number of data sources are available for sentiment analysis. Customer's opinion is a major criterion for increasing the growth of the company and to improve the quality of the service. The different data sources are social media, news articles, review sites, blogs, datasets, etc. [16].

d) Social Media

Social media become a huge platform to express the sentiments of the people. It is a large network where at a time millions of people can write share their views about the particular like there is different type of social media sites are

available like www.facebook.com, www.tweeter.com, www.hi5.com, www.linkedin.com etc. which contains millions of the people sentiments

e) News Articles

The websites like www.abpmajha.com, www.aajtak.com and www.lokmat.com, www.bhaskar.com has news articles that allow users or readers to comment. This helps in recording the opinions of the people in issues that are of current relevance and importance, like politics, corruption etc.

f) Review Sites

Before purchasing any product it is very important to know the opinion of the product. There are various ecommerce sites like www.flipkart.com, www.cnet.com, www.snapdeal.com, www.gsmcamera.com request customers to write their opinion about the product they have purchased [17]. So this site contains the million and trillions of the customer reviews about the product. Other sites like www.rediff.com/movies/reviews, www.indiaglitz.com and www.rottentomatoes.com has reviews for movies and www.yelp.com, www.burpp.com has restaurant reviews [18]

g) Blogs

A web log is called as blog it is a personal webpage on which particulars can write their likes, dislikes, opinions, hyperlinks to various sites etc. daily. Tweeter is one of the popular micro blogging service in which user creates status messages in a limited word count which called as tweets. The tweeter will get flooded while the elections were going on. Tweets can also use as data source for sentiment classification. Many of the blogs contain the issues; product information's recopies etc. so blogs used for the data source of the sentiment analysis [19]

3. Proposed Approach And Implementation

Sentiment analysis is the process of extracting sentiment from fragments of text. As people leave on the Web their opinions or reviews on products and services they have used, it has become important to develop methods for sentiment analysis and their corresponding visualization. Such opinions and sentiment analysis have an increasing influence in decision making.

Most of the existing methods are processing the reviews in terms of positive and negative comments and rate them in the order of positivity and negativity or summarizes the review information based on features in review sentences.

Sentiment Analysis and Visualization of online reviews of customers is necessary for products. Since a "picture is worth a million words", visualization helps the audience quickly absorb and interpret the presented data. Primary goal of visualization is to communicate information clearly and efficiently to users via the statistical graphics, plots, information graphics, tables, and charts selected. Effective visualization helps users in analyzing and reasoning about data and evidence. It makes complex data more accessible, understandable and usable. As a result, data visualization enables you to present a considerably larger amount of data

in comparison to the textual format. The viewer understands what you are trying to say at a first sight. [20]

Consumers depend on online user reviews to research products before purchasing and it helps companies to utilize such data for improving products, user services and maintaining customer relationship.

On the other hand, online customer complaints (or e-complaints), if not handled properly, could easily cause customers to lose loyalty for related products/services, reduce patronage, and create negative word-of-mouth. Thus, online customer feedback of products/service is useful for customer behavior analysis and is important for businesses.

The proposed system lets users to gain useful information for decision making process effortlessly, by transforming large collections of opinion text into visualizations that provides the conceptual understanding that would otherwise require the process of reading user reviews when searching for a product which is a rather daunting and lengthy task, since there are hundreds of reviews per product and they tend to vary too much to make a uniform decision. The systems alleviates the text consume ability problems of unstructured, high dimensional and conflicting reviews.

Proposed Approach:

The proposed system consists of following steps:

- A. Review database
- B. POS tagging
- C. Feature Extraction
- D. Polarity Identification
- E. Summary Generation
- F. Visualization

A. Review Database

Online reviews of products are extracted from web sites on periodic basis and extracted reviews are analysed and stored in review database. The user selects the product from the given list of choices. For selected product reviews of product are retrieved from the database.

B. POS Tagging

In part-of-speech (POS tagging), each word in review is tagged with its part- of- speech (such as noun, adjective, adverb, verb etc.). After POS tagging now it is possible to retrieve nouns as product features and adjectives as opinion words.

C. Feature Extraction

In feature extraction, product features are extracted from each sentence. Product features are generally nouns, so each noun is extracted from sentence. In review, features may be mentioned explicitly or implicitly by the reviewer. Features which are mentioned in a sentence directly are called as explicit features and features which are not mentioned directly are called implicit features.

For example,
"Battery Life of a phone is less"

In this sentence reviewer has mentioned battery life directly so it is explicit feature.

“This phone needs to charge many times in a day”

In this sentence reviewer is talking about battery of phone but it is not mentioned directly in the sentence. So here battery is implicit feature.

D. Polarity Identification

In polarity identification, semantic orientation of each opinion word is identified. Semantic orientation means identifying whether opinion word is expressing positive opinion, negative opinion or neutral opinion. Polarity identification predicts the orientation of an opinion sentence.

E. Summary Generation

This summary is based on features of product. With the help of information discovered in previous steps summary can be generated.

F. Visualization

Based on the summary generated from the reviews graphs such as pie charts bar graphs can be visualized which represent positive and negative sentiments.

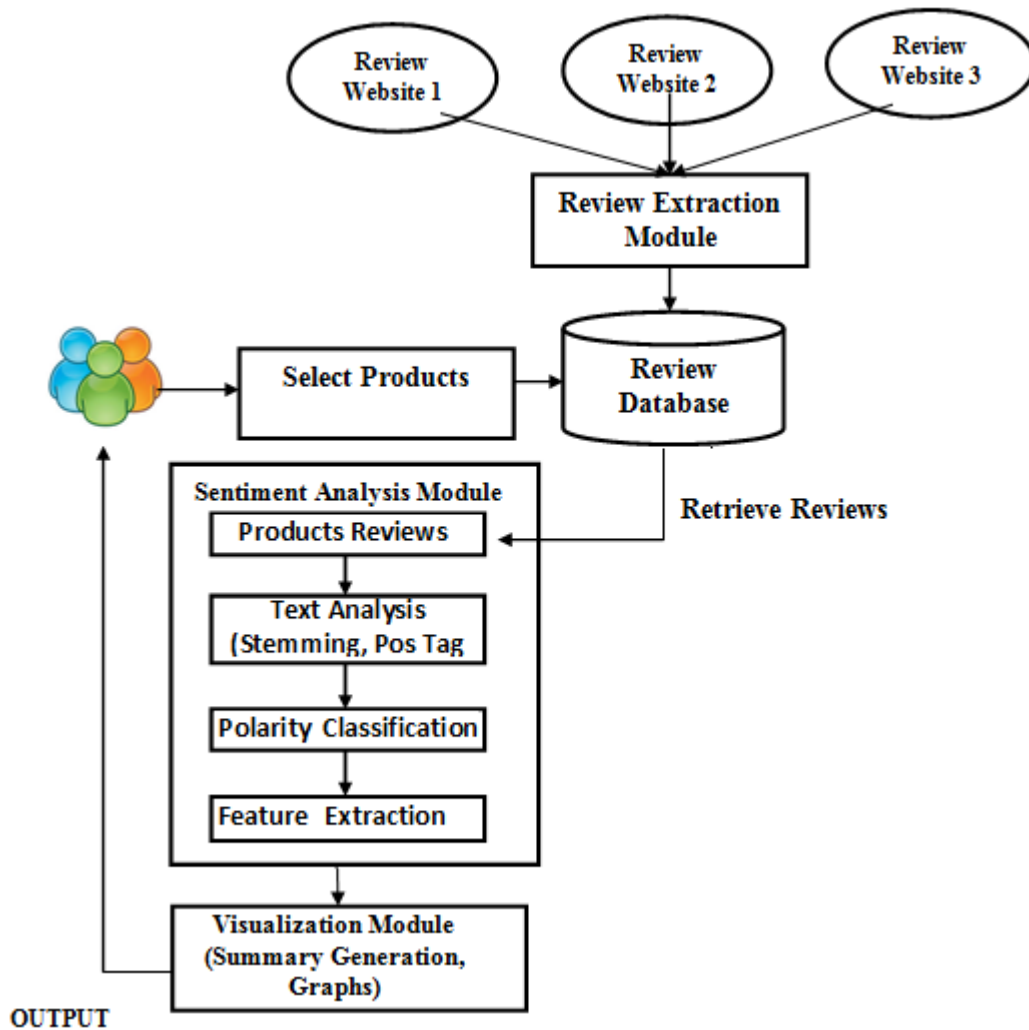


Figure 1: Block Diagram of Proposed System

4. Conclusion

Applying Sentiment analysis to mine the huge amount of unstructured data has become an important research problem. There is a huge need in the industry for sentiment analysis applications because every company wants to know how consumers feel about their products and services and those of their competitors.

The proposed system lets users gain useful information for decision making as quickly and as effortlessly as possible, by transforming large collections of opinion text into interactive visualizations that provides the conceptual understanding that would otherwise require the reading through the whole text

collection. The systems alleviates the text consume ability problems of unstructured, high dimensional and conflicting reviews

References

- [1] https://en.wikipedia.org/wiki/Sentiment_analysis
- [2] Sentiment Analysis and Opinion Mining April 22, 2012 Bing Liu.
- [3] Opinion mining and sentiment analysis Bo Pang¹ and Lillian Lee².
- [4] Padmaja, S., & Fatima, S. S. (2013). Opinion Mining and Sentiment Analysis–An Assessment of Peoples’ Belief: A Survey. *International Journal*.

- [5] A Survey of Classification Methods and Applications for Sentiment Analysis 1M.Govindarajan , 2,Romina M.
- [6] Philip S. Yu Ding Xiaowen, Liu Bing. A holistic lexicon- based approach to opinion mining. WSDM'08, 2008.
- [7] Jeonghee Yi, Nasukawa, Bunesu , Niblack, W., "Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques" T hird IEEE International Conference on , 10.1109/ICDM.2003.1250949, 19-22 Nov. 2003.
- [8] M. Hu and B. Liu. Mining opinion features in customer reviews. In AAAI'04: Proceedings of the 19th national conference on Artificial intelligence pages 755–760.
- [9] Kushal Bafna, Durga Toshniwal. Feature Based Summarization of Customers' Reviews of Online Products. In 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - KES2013
- [10] B. Liu, M. Hu, and J. Cheng. Opinion observer: analyzing and comparing opinions on the web. In International Conference on World Wide Web, 2005.
- [11] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques", In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79–86, 2002.
- [12] Zhu, Jingbo Wang, Huizhen Zhu, Muhua Tsou, Benjamin K. Ma, Matthew, "Aspect-Based Opinion Polling from Customer Reviews", IEEE Transactions on Affective Computing, Volume: 2, Issue: 1 On page(s): 37. Jan-June 2011.
- [13] A Kamal, M. Abulaish and T. Anwar, "Mining feature - opinion pairs and their reliability scores from web opinion sources," WIMS '12, June 13-15, 2012 Craiova, Romania.
- [14] G. Draper and R. Riesenfeld. Who votes for what? a visual query language for opinion data. IEEE Transactions on Visualization and Computer Graphics, 14(6):1197–1204, 2008.
- [15] S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima. Mining product reputations on the web. In ACM SIGKDD international conference on Knowledge discovery and data mining, pages 341–349, 2002.
- [16] Neha S. Joshi, Suhasini A. Itkat, "A Survey on Feature Level Sentiment Analysis", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 014, 5422-5425.
- [17] M.Govindarajan ,Romina M, "A Survey of Classification Methods and Applications for Sentiment Analysis ", The International Journal Of Engineering And Science (IJES) ||Volume|| 2 ||Issue|| 12 ||Pages|| 11-15 ||2013|| ISSN(e): 2319 – 1813 ISSN(p): 2319 – 1805.
- [18] Singh and Vivek Kumar, "A clustering and opinion mining approach to socio-political analysis of the blogosphere". Computational Intelligence and Computing Research (ICCIC), 2010 IEEE .
- [19] G.Vinodhini, R.M.Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 6, June 2012.
- [20] <http://www.uauug.org.uk/what-are-the-advantages-of-data-visualisation.html>

Author Profile

Miss Apurva Vyankatesh Gundla has received B.E Degree in Information Technology from University of Solapur, Maharashtra, India and pursuing the M.E. degree in Computer Science and Engineering in Nagesh Karajagi Orchid College of Engg. & Technology, Solapur, India. She is doing her dissertation work under the guidance Prof. M. S. Otari , Assistant Professor at Nagesh Karajagi Orchid College of Engg. & Technology, Solapur, Maharashtra, India..

Prof. M. S. Otari is an Assistant Professor in Nagesh Karajagi Orchid College of Engineering and Technology, Solapur, Maharashtra, India. Her research area includes Machine learning, Natural Language Processing, Web Technology, Networking.