*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 10 Issue: 3***
***Article Received:** 25 January 2022 **Revised:** 12 February 2022 **Accepted:** 30 March 2022*

# Feedback Analysis of BEAMS using Opinion Mining

**Arti Deshpande**

Computer Engineering Departmentline, Thadomal Shahani Engineering College , Mumbai, India, e-mail: arti.deshpande@thadomal.org

**Vivek Nawle**

Public Works Department , Finance Ministry, Government of Maharashtra, Mumbai, India

**Abstract**—Online Feedback System for BEAMS (Budget Estimation, Allocation and Monitoring) helps in collecting feedbacks from users for various projects undertaken by the PWD (Public Works Department) . Proposed opinion mining on the feedbacks helps to view and analyze the response and reviews of the users. This can be used to assess the work in progress and exploring work of PWD. So superior officer can review the work done by engineers for particular project assigned to him. The proposed technique also provides positive and negative feedback and also uses geolocation of user, google maps, authentication system. Feedback form is provided in English and regional language Marathi.

**Keywords**-data mining; natural language; NLTK API; opinion mining, sentiment analysis

## I. INTRODUCTION

Opinion mining is a type of text analysis which helps to classifies the text and make decision by extracting and analyzing the text. There are positive and negative opinions which processes the degree of positive or negative associated with that event (people, organization, social issues).

Customer reviews play an important role in project development. An easily accessible means of submitting reviews which are available to a wide range of customers would be an online feedback system.

The Online Feedback System allows users to give feedback for various projects undertaken by BEAMS. It reduces the manual work of BEAMS officials to take feedback from people. The online feedback system is much more reliable as compared to the traditional feedback system in the sense that it increases the efficiency and reduces time and effort used to collect feedback. Currently, no online system exists to collect feedback; users have to go through increasingly convoluted line of control. It also increases the scope of users giving feedback significantly. Online feedback system helps to keep track of users providing the feedback and also keeps the BEAMS officials updated with the fact as to which projects have overall negative feedback and needs more work to be done for user acceptance. It allows for users to upload pictures of the work in question to highlight the issue, if any.

It performs opinion mining on the text comments given by the users, thereby decreasing the time and effort needed by the officials to understand the gist of multitude of comments. There are two modules, first is an end user module for users to leave feedback on a selected project and the second is an admin side module for verified officials to view the reviews as well as analyze them. The analysis of the reviews is through opinion mining to classify the projects into various groups. The opinion mining is done at document level based on the comments submitted by the reviewers.

The admin has access to the list of projects under said admin's jurisdiction. On click of a project, sentiment analysis [1] of reviews through opinion mining via NLTK(Natural Language Tool Kit) [2] is done. Opinion Mining is done on comments submitted by the user. Bi-weekly, CO (Chief Officer) receives an email that consists of all opinion mining results of all projects under that CO's jurisdiction as well as all feedbacks submitted for the same.

## II. LITERATURE REVIEW

A recent publication [3] classifies customer reviews based on an existing domain-specific corpus by applying a lexicon-based sentiment analysis. The authors propose creation of a domain specific lexicon by utilizing an existing corpus. The lexicon created is straightforward and consists of tokens along with their sentiment values and POS tag (Part of Speech) from the corpus of 80,000 reviews picked from TripAdvisor.com. The reviews are divided according to class labels and only those reviews are taken which discriminate between those labels to guarantee that the lexicons are a disjoint set of tokens. Each of the five distinct lexicons are assigned separate values. Based on sentiment analysis of the document the classification function is computed to generate lexicons. In order to compute the sentiment value for one document, the sum of all identified sentiment values is generated. The evaluation of the classification on test data shows that the proposed system performs better compared to a predefined baseline: if a customer review is classified as good or bad the classification is correct with a probability of about 90%.

The results are evaluated using precision, recall and F-measure. Precision defines the proportion of reviews the system classified correctly to all reviews classified. Recall describes the proportion of reviews selected correctly to all reviews selected. The F-measure uses a parameter controlling the influence of precision and recall. The precision and recall values significantly exceed the baseline values gathered.

However, the study does not take into consideration the subjectivity of the reviews. Biased reviews can severely affect the result of sentiment analysis. The lexicons used consider singular words. In such a case collocations such as 'not good' could be classified as positive due to the high positive sentiment of 'good'. Furthermore, the processing time of the data increases when low information words (such as articles) are used. Instead, considering the sentiment values of only high information words could yield better results.

Minqing Hu and Bing Liu proposed a set of techniques [4] for mining and summarizing product reviews based on data mining and natural language processing methods. They provided a feature base summary of reviews about a product online. The authors mined products features using both Data mining and NLP (Natural Language Processing) [5]. Then the opinion sentences in each review were identified. This step itself has three more sub steps. First is the identification of adjective words done via NLP. The semantic orientation of these words is then found through a bootstrapping technique using WordNet. Then the orientation of the entire sentence is found. After that, the result of the entire review depending on the orientation of its encompassed sentence are found. Three perspectives are used to evaluate the system :

1. The effectiveness of feature extraction.

2. The effectiveness of opinion sentence extraction.

3. The accuracy of orientation prediction of opinion sentences.

The experiments were performed on reviews collected from Amazon.com and Cnet.com about 5 products. The recall and precision of association mining, compactness pruning, P-support pruning and infrequent feature identification were compared. Association mining returns low precision results. The pruning methods improve precision while recall remains steady while infrequent feature identification improves recall with a slight drop in precision.

The study further compares the results from the proposed system to those of FASTR, a well-known and publicly available term extraction and indexing system by Christian Jacquemin. The average recall and precision are both improved by FBS (Feature Based Summarization).

While this work too does not consider objective reviews, instead focusing on classifying all reviews under positive or negative, it does implement pruning. It prunes words which have low information value thereby increasing the efficiency of the method, as can be seen from their experimentation results. Furthermore, it does not consider N-grams and their effect on the overall phrase. This may cause a significant shift in accuracy of the method.

Peter Turney [6] proposes the use of an unsupervised learning algorithm for classifying reviews as recommended or not recommended. The PMI-IR algorithm is employed to estimate the semantic orientation of a phrase. PMI-IR uses Pointwise Mutual Information (PMI) and Information Retrieval (IR) to measure the similarity of pairs of words or phrases. The algorithm assigns a numerical rating by comparing its similarity to a positive phrase (excellent is used here) and subtracting from that its similarity to a negative phrase (poor is used here). This numerical rating indicates the semantic orientation as a number.

The algorithm proposed first extracts phrases containing adjectives or adverbs. Two consecutive words are extracted from the review if their tags conform to certain patterns. The semantic orientation of the extracted phrases is then found using PMI-IR algorithm. The third step is to calculate the average semantic orientation of the phrases in the given review and classify the review as recommended if the average is positive and otherwise not recommended. The algorithm achieves an average accuracy of 74% when evaluated on 410 reviews from Opinions, sampled from four different domains (reviews of automobiles, banks, movies, and travel destinations).

The study relies on presence of adjectives to ensure subjectivity. It does not consider sole adjectives either. Instead ensuring the subjectivity of the reviews being classified could yield better results. Compared to the earlier mentioned study, Peter Turney does consider N-grams and their combined semantic orientation. The time required for queries, as mentioned in the publication is a limitation of the work. It could be improved by considering the level of information of the phrases being evaluated.

*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 10 Issue: 3***
***Article Received:** 25 January 2022 **Revised:** 12 February 2022 **Accepted:** 30 March 2022*

An automated web-based course feedback system was proposed by V.K. Singh et al [7]. It is an online based system. It is composed of various binary and graded response questions. It also consists of a text box where students can write in prose about their experience which is stored as a varchar field in the database. The system then calculates the score of feedback. The text entered by the student in free form input text box is then used for opinion mining to label the feedback as 'positive' or 'negative'. Thus the system uses opinion mining to calculate the score of each feedback. The algorithm used is an unsupervised SO-PMI-IR. The level of classification for sentiment analysis is document-level. The system also uses other algorithms such as Naïve-Bayes classification and semantic orientation approach. The system is implemented using a mysql database.

The data provided in response to the various closed reference questions is used by various queries in the database to glean information. The system gleans information in two ways:

1. Directly from the entered data to close ended questions

2. From the opinion mining conducted on the free-form text entered by the students.

The students are hence asked to provide detail sin their feedback to help get as much information as possible from each review.

The research doesn't take into account whether the comments are subjective or objective. In case of collocations, it reverses the SO value if the term is preceded by 'not' thereby only considering bigrams and not others. It allows for collecting information via binary responses ('yes' or 'no' in this case) thereby decreasing effort required by the users.

Ashish Bhalerao and Sachin Deshmukh [8] proposed a new methodology which is a lexicon-based approach. They also compared the accuracy on different domain of dataset using sentiment analysis. They collected reviews from Amazon.com and Cnet.com of various electronic products such as digital cameras, DVD player, mp3 player and cellular phone. Each of the format of reviews were text review and a title. They used WordNet as a dictionary to determine the opinion words and their synonyms and antonyms. Their research is based on Document level as well as sentence level sentiment analysis. The steps involved:

1. Creating a dataset

2. Data Pre-processing

3. To generate the score of reviews as positive, negative or neutral

4. Classification of score of user reviews

5. Result

To determine the sentiment of document, two different machine learning techniques such as Naive Bayes and Support Vector Machine (SVM) are used. To compute how well the system classifies each document as compared to human decision, all the documents were manually classified and the corresponding opinion was determined. The results were then compared with the result of the system. Same reviews were also applied to the other system named as "AIRC Sentiment Analyzer" available online. Finally, the results of the two systems were compared and the results have shown that the results of the document based Sentiment orientation system are better than that of AIRC Sentiment Analyzer.

## III. PROPOSED METHOD

The method for feedback submission from users is given in Figure. 1. The user first narrows the number of projects by submitting input values regarding the location of the project, the type and the year the project. List of projects with summery details are provided for the review as per the user's selection. By selecting project, user van gives feedback in marathi or english.

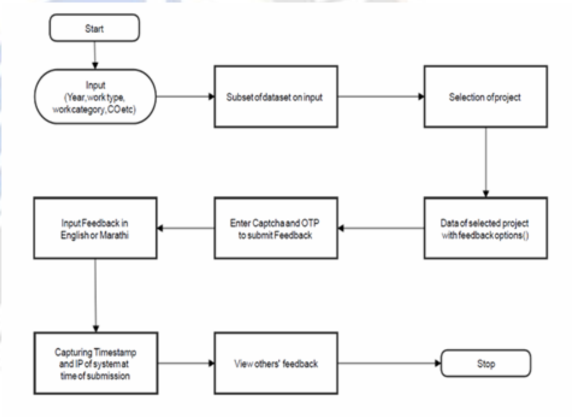


Figure 1. **Block Diagram for User Activities**

The method for admin to view analysis of the feedback is given in Figure. 2. The admin first needs to log in to view all the reviews and can also request for analysis of the feedback. Using feature-based opinion mining [9], clusters of projects are displayed to the admin based on similar reviews.

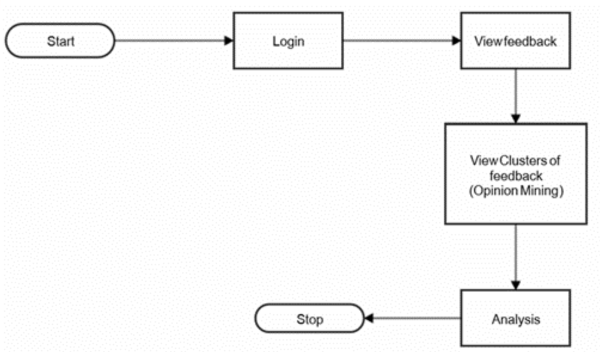


Figure 2. Block Diagram for Admin Activities

## IV. METHODOLOGY

The methodology used for the proposed system is given below as per the categorization of user.

### *A. User's View*

*Filtering of results:* First and foremost the results to be displayed to the user are filtered based on selected inputs. Mainly geographical and physical characteristics such as the location of the project and the type of construction(building, road, etc) are taken into account. The user can then select the project for which he wants to give feedback.

*Display of project details and location on map:* Only the details understandable by the general public have been displayed to the user. Details such as project name in english and marathi, type of work, work category, physical progress, total expenditure, date of work order, region, district, taluka, CO, user department have been displayed on the screen with the location using Google Maps [10]. Geocoding [11] is used in Google Maps to pinpoint the location of the project.

*Feedback:* The feedback itself consists of personal details such as name and number, a rating and a text comment (which is not necessary to fill) and photo of the project site. The number is used to ensure that an individual does not spam the website continuously regarding one project. The individual will be able to submit feedback again regarding the project after a stipulated amount of time. Base64 [12] encoding is used for the photos. For security purposes Captcha [13] and OTP(One Time Password) [14] are also used. A graphical display of rating given by other users to that project are shown. The user can also see the text comments and photos submitted by other users.

### *B. Admin's View*

*Display of feedback:* The admin will have access only after successful login. A list of all projects under the admin along with the CO responsible and the average rating of that project as computed from all associated user feedback. The admin can also see all feedback on projects under them.

*Display of opinion mining result:* On click of a project the opinion regarding that project is calculated and displayed. The calculation of the opinion is done by sending a request to a NLTK API. The response is then processed and the result is displayed.

A subset of the data sample [15] used in testing is as shown in TABLE I. It comprises of data available on BEAMS website.

### *C. Opinion Mining*

The API [16] used for Sentiment Analysis with Python NLTK Text Classification by J. Perkins uses the Hierarchical Classification concept. It combines a subjectivity classifier and a polarity classifier with the polarity classifier being applied if the subjectivity classifier labels the text to be subjective. If the text is objective, then a label of neutral is returned, and the polarity classifier is not used.

The polarity classifier used is Naive Bayes. It classifies the text into either positive or negative. The NLTK classifier work with featstructs [17]. A feature structure is actually just a kind of dictionary, and is accessed by indexing. Here the feature is a word while the value is True. Hence a simplified bag of words model is used. The classifiers are trained using data from BEAMS.

The three metrics to evaluate the effectiveness of a classifier are accuracy precision and recall [18]. Accuracy is the degree of closeness of measurements of a quantity to that quantity's true value. Precision measures the exactness of a classifier. Recall measures the sensitivity or completeness of a classifier. High precision decreases the number of false positives while high recall decreases the number of false negatives. Improving recall results in an increase in the sample space. This results in lower precision.

F-measure or balanced F-score combines precision and recall. It is calculated as the harmonic mean of the two. J. Perkins indicates that the use of F-measure does not improve the effectiveness of the results nor does the filtering of stopwords. The effectiveness can be improved by including significant bigrams in it and including only words with high information gains. Bigram collocations are a pair of consecutive written units.

## V. IMPLEMENTATION AND RESULT

The user can select a Work Type, Work Category, Region, Statutory Board, District and Constituency or Taluka. The values of these parameters selected narrows the applicable projects to be displayed. The list of projects will be displayed conforming to the selected parameters. The list will include the Budget Item No (Project ID), Name of the Project in English and Marathi. User can select a project from this list.

The details of the selected project are displayed to the user along with a map highlighting the location of the project as shown in Figure. 3. The feedback form consists of Name, Contact Number, Rating and Comment. The user can also attach images before submitting the review. The user is displayed a message informing about the status of the submitted response. The user can view others' feedbacks for the same project after submitting their own feedback. The user can view other user's feedbacks when the user clicks on the show other feedback button. Subset of sample dataset used for implementation is given in Table I

*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 10 Issue: 3***
***Article Received:** 25 January 2022 **Revised:** 12 February 2022 **Accepted:** 30 March 2022*

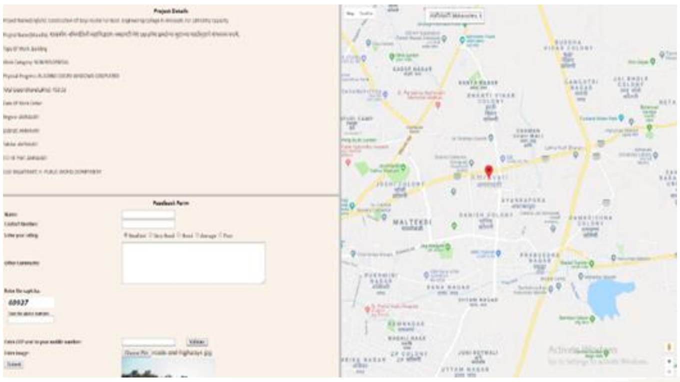

Figure 3. Feedback form

The user can view the summary of ratings for each project divided according to the rating submitted as shown in Figure. 4 only after submitting his review. This gives an overview of the feedback from other users. The admin can view the feedbacks submitted by all users on all projects. The Project ID, Name of User, User Comments and Image are displayed.

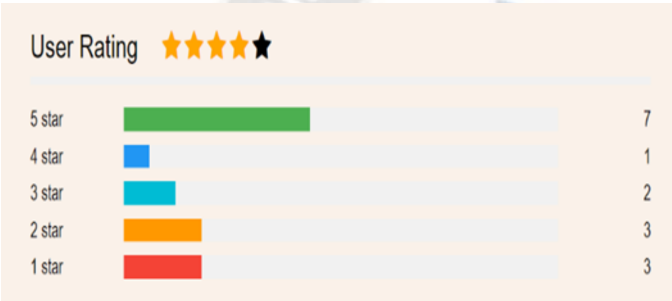


Figure 4. Summary of Ratings

The admin can view the Project ID, Project Name, CO of Project and The Average Rating of the Project as shown in Figure. 5. The average rating is calculated from all the rating submitted by users for that particular project as shown in table II

TABLE I. SUBSET OF DATA SAMPLES(A)

| Budgetitem No. | District | ProjectName | Type | Work Category | Physical Progress | Cost |
|---|---|---|---|---|---|---|
| 1003/400/10388 | AMRAVATI | Improvement To Talani Rajurwadi Road | Road | BASE COURSE PLUS SURFACING | In Progress | 41.37 |
| 1003/441/10433 | AMRAVATI | Land Acquisition For Arala Armgaon Road | Road | W PLUS BASE COURSE PLUS SURFACING | | 7.9 |
| 1103/013/300175 | GONDIA | Improvement To Sangadi Arjuni Road | Road | WIDENING SURFACING | | 274.96 |
| 1212/159/03353 | NAGPUR | Constructio Of Four Laning Road Nagpur City | Road | RIGID PAVEMENT | IN PROGRESS | 556.64 |
| 0207/227/01099 | AURANGABAD | Land Acquisition Of Nagpur | Road | BOT OR DBFOT | UNAVOIDABLE EXPENDITURE | 588.6 |
| 0803/108/02204 | JALNA | Imp. To VR 71 To Shirpur Road Km | ROAD | BASE COURSE PLUS SURFACING | UNAVOIDABLE EXPENDITURE | 86.39 |
| 1103/041/40005 | AMRAVATI | Const. Of Sub-Divisional Office | Building | NON-RESIDENTIAL | | 60.01 |

TABLE II. SUBSET OF DATA SAMPLES (B)

| Region | Constituency | Taluka | CO | UserDepartment |
|---|---|---|---|---|
| AMRAVATI | MORSHI | MORSHI | SE PWC AMRAVATI | H-PUBLIC WORKS |

| | | | | |
|---|---|---|---|---|
| | | | | DEPARTMENT |
| AMRAVATI | DARYAPUR | DARYAPUR | SE PWC AMRAVATI | H-PUBLIC WORKS DEPARTMENT |
| NAGPUR | ARJUNI MORGAON | MORGAON | SE PWC NAGPUR | H-PUBLIC WORKS DEPARTMENT |
| NAGPUR | NAGPUR NORTH | NAGPUR | SE PWC NAGPUR | H-PUBLIC WORKS DEPARTMENT |
| AURANGABAD | VAIJAPUR | VAIJAPUR | SE PWC AURANGABAD | |
| AURANGABAD | PARTUR | MANTHA | SE PWC AURANGABAD | H-PUBLIC WORKS DEPARTMENT |
| AMRAVATI | ACHALPUR | ACHALPUR | SE PWC AMRAVATI | H-PUBLIC WORKS DEPARTMENT |

TABLE III. EXAMPLE PROJECT COMMENTS

| COMMENT | OPINION MINING RESULT |
|---|---|
| THE PART OF WORK COMPLETED IS GOOD. | POSITIVE |
| THE WORK DONE IS SATISFACTORY. | POSITIVE |
| THE WORK IS BEING DONE VERY SLOW. | NEGATIVE |
| THE WORK COMPLETED IS EXCELLENT. | POSITIVE |
| NO WORK IS BEING DONE. | NEGATIVE |
| THE QUALITY OF WORK IS GOOD BUT A LONG TIME IS BEING TAKEN TO COMPLETE IT. | NEUTRAL |
| VERY RUDE WORKERS. | NEGATIVE |
| WORK COMPLETED IS GREAT! | POSITIVE |
| THE ROAD LOOKS VERY GOOD. | POSITIVE |
| GOOD WORK. | POSITIVE |

Counting the number of positives, negatives and neutrals we get:

Positive:6

Negative:3

Neutral:1

Since a greater number of comments show a positive response, we consider the overall opinion regarding the selected project to be positive.


Figure 5. Admin's View of List of Projects

## VI. CONCLUSION AND FUTURE WORK

The proposed system has been successfully implemented. It decreases the amount of time and effort required by the users to submit a feedback by reducing the paperwork and allowing for more improved feedback rather than the earlier tedious method. It also decreases the effort to be expended by the officials to acquire subjective feedback to understand the public opinion on projects under them. It also allows for an estimation of the efficiency of the various teams in charge of their projects. Research on incorporating Opinion Mining for Marathi comments as well will be done. The project will be incorporated into the existing BEAMS website for public use in the future.

## ACKNOWLEDGMENT

Through this acknowledgment, we express our sincere gratitude to all those people who have helped us with this assignment and made it a worthwhile experience.

Firstly, we extend our thanks to the various people who have shared their opinions and experiences through which we received the required information crucial for our report.

Finally, we express our thanks to BEAMS who gave us this opportunity to learn the subject in a practical approach.

## REFERENCES

[1] Singh, Vandana, and Sanjay Kumar Dubey. "Opinion mining and analysis: A literature review." In *Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference*-, pp. 232-239. IEEE, 2014.

[2] "Natural Language Toolkit — NLTK 3.3 Documentation". 2018. *Nltk.Org*. http://www.nltk.org.

[3] Gräbner, Dietmar, Markus Zanker, Gunther Fliedl, and Matthias Fuchs. *Classification of customer reviews based on sentiment analysis*. na, 2012.

[4] Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews." In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168-177. ACM, 2004.

[5] Chowdhury, Gobinda G. "Natural language processing." *Annual review of information science and technology* 37, no. 1 (2003): 51-89.

[6] Turney, Peter D. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." In *Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 417-424. Association for Computational Linguistics, 2002.

[7] Singh, V. K., P. Kumari, A. Singh, and J. Thapa. "An automated course feedback system using opinion mining." In *Information and Communication Technologies (WICT), 2011 World Congress on*, pp. 72-76. IEEE, 2011.

[8] Ashish A. Bhalerao, and Sachin N. Deshmukh. "Opinion Mining from Multi-domain User Reviews Using Sentiment Analysis." In *International Conference On Recent Advances In Computer Science, Engineering And Technology*, pp 55-60. IOSR-JCE.

[9] Mishra, Nidhi, and C. K. Jha. "Classification of opinion mining techniques." *International Journal of Computer Applications* 56, no. 13 (2012).

[10] "Overview | Maps Javascript API | Google Developers". 2018. *Google Developers*. https://developers.google.com/maps/ documentation/javascript/tutorial.

[11] "Geocoding Service | Maps Javascript API | Google Developers". 2018. *Google Developers*. https://developers.google.com/maps /documentation/javascript/geocoding.

[12] Josefsson, S. "RFC 4648: The Base16, Base32, and Base64 Data Encodings, 2006." (2009).

[13] "What Is CAPTCHA (Completely Automated Public Turing Test To Tell Computers And Humans Apart)? - Definition From Whatis.Com". 2018. *Searchsecurity*. https://searchsecurity.techtarget. com/definition/CAPTCHA.

[14] "Geocoding Service | Maps Javascript API | Google Developers". 2018. *Google Developers*. https://developers.google.com/maps/ documentation/javascript/geocoding.

[15] "BEAMS :: Budget Estimation, Allocation & Monitoring System". 2018. *Beams.Mahakosh.Gov.In*. https://beams.mahakosh.gov.in/ Beams5/BudgetMVC/index.jsp.

[16] Perkins, Jacob. 2018. "Hierarchical Classification". *Streamhacker*. https://streamhacker.com/2011/01/05/hierarchical-classification/.

[17] 2018. *Nltk.Org*. http://www.nltk.org/book_1ed/ch09.html.

[18] Forman, George. "An extensive empirical study of feature selection metrics for text classification." *Journal of machine learning research* 3, no. Mar (2003): 1289-1305