



# **Construction Cost Overturn Prediction Using Artificial Neural Network on Text and Numerical Data**

Payal Singla<sup>1</sup>, Er. Mohit Kakkar<sup>2</sup>

M.Tech, Dept. of Computer Science & Engineering, Desh Bhagat University, Mandi Gobindgarh, Punjab, India<sup>1</sup>

Assitant Professor, Dept. of Computer Science & Engineering, Desh Bhagat University, Mandi Gobindgarh, Punjab, India<sup>2</sup>

**ABSTRACT:-**Now-a-days, with the increase in technology every information is available on internet but sometimes it becomes difficult to find accurate information in short. There should be a proper path from where people can get information about the main thing what they want instead of long material. In this work we have studied construction cost overturn prediction on text and numerical data in a NLP to produce a prediction of the level of cost overrun and under run using artificial neural network algorithms. Earlier model shows an average accuracy of 43.72% also creates a problem of redundancy and the work has been done by ensemble classifiers. After having this problem we have created a neural network that discusses how text can be processed, combined with numeric values then give the cost overturn with the accuracy of approx. 87% with improved performance and remove redundancy.

**KEYWORDS:-**prediction, accuracy, costoverturn, neuralnetwork, textmining, wordselection.

## **I. INTRODUCTION**

Internet has been used widely these days across the world to retrieve information for various purposes through various search engines like google, yahoo, safari, etc. where sometimes the result may be in detail. Users feel bore and think that it is time consuming. Many factors affect construction **cost overruns**. Possibly, indicators provided in bidding text documents can identify construction projects that potentially will have large cost overruns. Text summaries of what is to be constructed for a project, and project line item text are available from project bidding data collected by state transportation agencies. Additionally, numeric data are available at the time of the bid opening including the projects' magnitude, and the number of bidders. It has now become possible to combine various text and data mining methods to project data to attempt to make **predictions** for overruns. With the development of text mining algorithms that allow the extraction of information from the text, it may be possible to find directions of the projects' nature and likelihood to experience cost overturn. Text mining can be defined as the automatic discovery of previously unknown information from unstructured text data. **Text mining** involves extracting information of interest from text documents and then the use of data mining to discover new associations among the extracted information. Text mining has been found to give excellent results for some predictive applications. It is very important to know about the construction cost overturn initially so as it become bidders easy to take work according to their budget. It just give the estimation that the cost you have expected is up to the mark or it will fluctuate overrun or under run. Sometimes big projects and tenders are initiated thereby prediction of the budget is necessary to take work. So here we created a **neural network** that helps to predict the construction cost overturn accurately.

There are two basic methods of text pre-processing: (a) feature extraction and (b) feature selection, which are detailed below.

In (FE) we extract the words that are in need to find the construction cost overturn. the words may be gathered or extracted from sentence or paragraph. It deals with single words, tokenization, remove stop words and stemming-word and provides knowledge about the grammatical structure of a language known as syntax analysis. Semantic analysis deals with the meaning of the words based on Word Net-Affect.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

In (FS) we select the word that has been used in construction on the basis of noun and have expected overturn from the data that has been collected years ago, words have been selecting after stemming and N-GRAM text extraction. Prediction has been made on that basis through the neural network based on matrix based.

## II. REVIEW OF RELATED WORK

**1. Williams, Trefor P.** et. al. [1] has worked Predicting construction cost overruns using text mining, numerical data and ensemble classifiers. This paper discusses how text can be combined with numerical data to produce a prediction of cost overrun using data mining classification algorithms. The stacking model had an average accuracy of 43.72% for five model runs. It was found that a stacking model that used only numerical data produced predictions with lower precision and recall. A potential application of this research is to budget sufficient amount to complete a construction project.

**2. C.G. Wilmot** et.al.[13] has proposed the model that estimates highway construction cost in terms of cost index. The most influence factors were found for highway construction cost calculations be the cost of material, labor and equipments used in constructing the facility as state department of transportation are required to prepare highway construction programs that describe their planned construction activity in short term.

**3. J. Zhang** et.al.[6] has proposed semantic modeling, semantic natural language processing techniques .It used six phase iterative approach that contains text classification, information extraction, information transformation, compliance reasoning. Algorithms used for discussing and presenting information transformation which transform the extracted information that reduce time for checking cost manually then to check with automated compliance checking.

**4. Sibarani, E.M.**[11] has Researched on Natural Language Processing (NLP). In a series of natural language processing, the initial step is parsing the sentence based on the grammar in order to understand the meaning of a sentence. The three main stages are: 1) preprocessing to generate events files, 2) lexical analysis to convert into tokens, and 3) syntax analysis to build parse tree that requires file events to calculate the probability of the grammar by count the occurrence frequency on file events to determine the best sentence trees. An evaluation was performed to the parser using 30 simple sentences and the outcomes were able to generate a file and probability calculations.

**5. Hyojoo. Son, Changmin. Kim.**[7] has researched an accurate prediction of cost performance of commercial buildings using pre-project planning stage that is to develop various techniques in which one can address the risk and decide resourced to maximize the chance for successful project. It is more efficient than other models. It is also known as front –end planning. thus it is very important to measure the project performance from the earliest stage of a project lifecycle.

**6. Steven.M. Trost, G.D. Oberlender.**[5] predicts the accuracy of early cost estimates that represents a key in any business. Main aim is to accomplish objectives, quantitative data collected from the completed construction projects in the industry. Accuracy is based on estimate score that helps the team to score an estimate and then predict its accuracy based on estimate score.

**7. K. Petrousatou, E. Georgopoulos, S. Lambropoulos, J.P. Pantouvakis.**[3] predicts the early cost estimation of road tunnel construction using neural network that is subjected to underground uncertainties and risk where it is difficult to predict the final construction cost where issues are evaluated and design are made. In this final construction cost has been determined.

## III. OBJECTIVES

1. To create and design the proposed solution for text mining construction cost overturn prediction to mitigate the shortcomings of the existing models.
2. To implement the proposed model using the NETBEANS.
3. To evaluate the performance of the proposed model implementation using various types of construction text data and SCG Neural Network.

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

- 4.To improve the accuracy and speed of the proposed model.
- 5.To obtain and conclude the results obtained for the final implementation of the proposed model.

## IV. PROPOSED METHODOLOGY

The following figure represents the diagram of the propose system. Proposed model has the two stages:-

- 1.In step 1 Firstly the data has been read and files have been loaded .
- 2.After that some keywords have been extracted that has been used to make a keyword table and the redundancy or duplicity has been removed so as not t o overwrite.
- 3.Next is to train the data and neural network so as to give iteration and accurate value to minimum error.
- 4.In step 2 we test the network by joining it with the training data.
- 5.At last final output has been obtained that tells the prediction of cost overturn.

### Step 1 Train Model

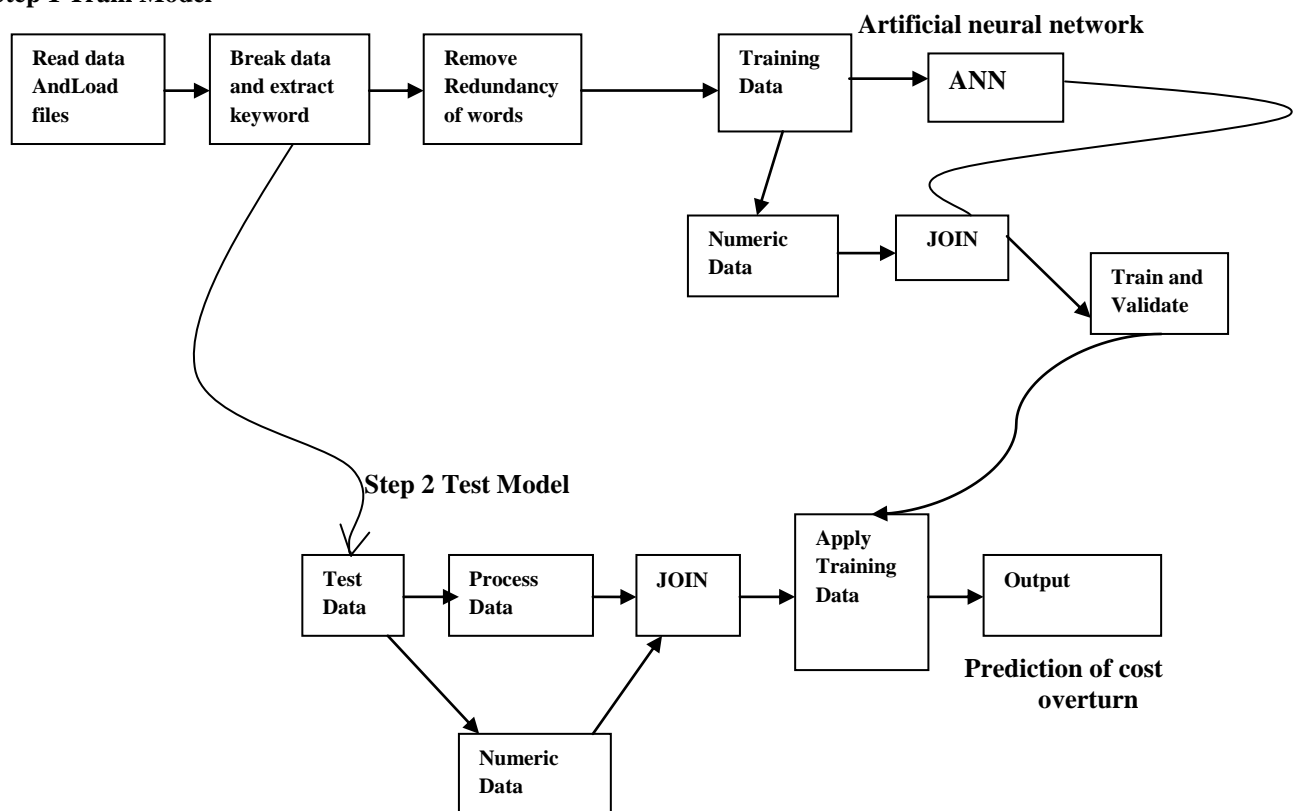


Fig. 1 : Proposed Methodology

## V. IMPLEMENTATION AND RESULTS

This experiment is conducted by programming in Java using Netbeans IDE. We created the code in Java and ran to collect the results. The coding is done using multiple third party tools like Neuroph for Neural Network Support and Stand ford Core NLP's POS Tagger for Part of Speech based Tagging and speech Segmentation. The classification is performed using Multilayer Perceptron aka. Artificial Neural Networks for Pattern recognition.

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

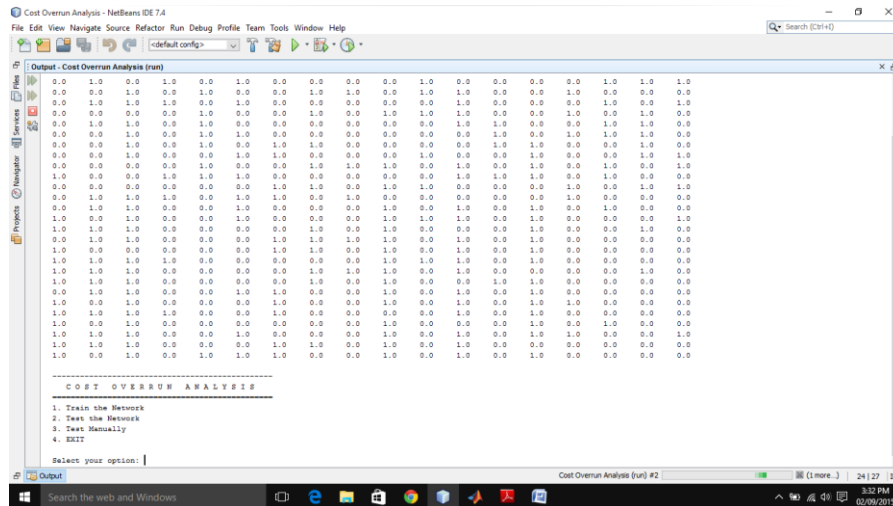


Fig. 2

Figure 2 shows program execution and the POS library initialization. POS library is used to segment the speech to extract keywords frequently used. Basically Nouns and Adjectives are extracted using this method.

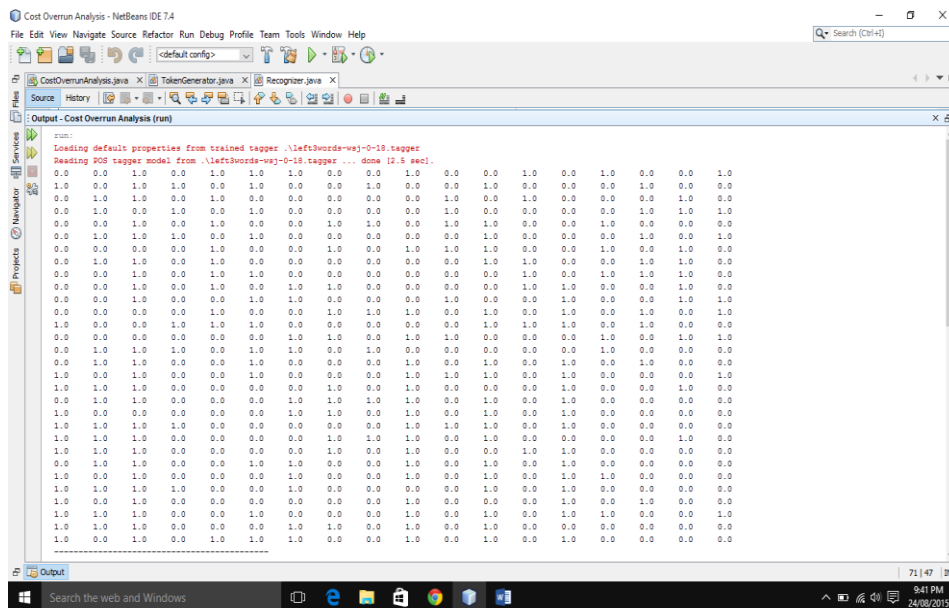


Fig. 3

Figure 3 shows the generation of a feature matrix. This matrix is generated by marking all the words that are presented in a sample as 1 and not present as 0. The columns are all the unique words found after initial thresholding. Rows are sample documents for training.

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

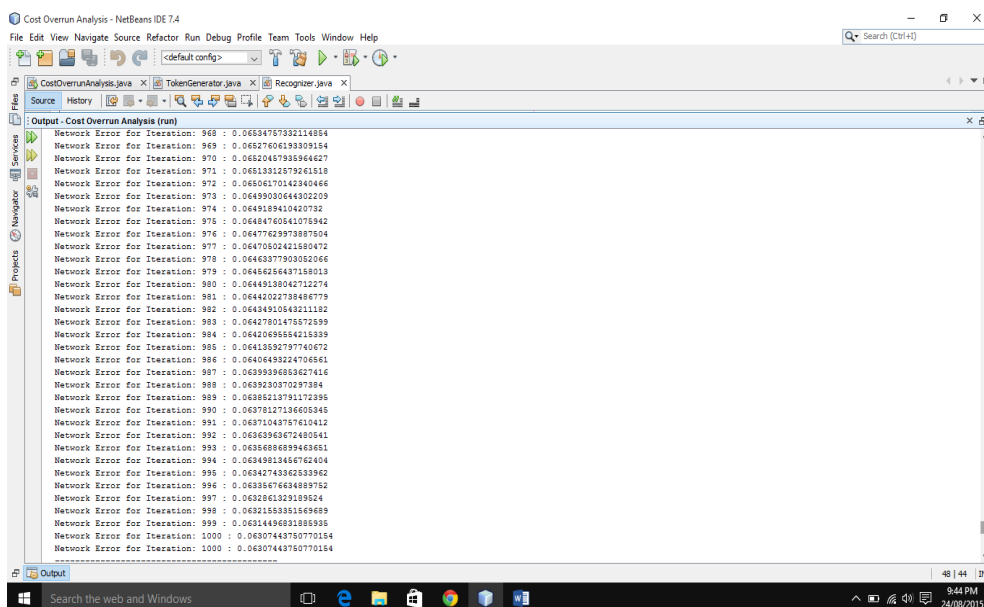


Fig. 4

Figure 4 shows the process of training the Artificial Neural Network. The neural network library called Neuroph is used to implement the Artificial Neural Networks in the programming. Here’ the system is trained to maximum Iteration value set for it (1000) and the total Error of the network is shown.

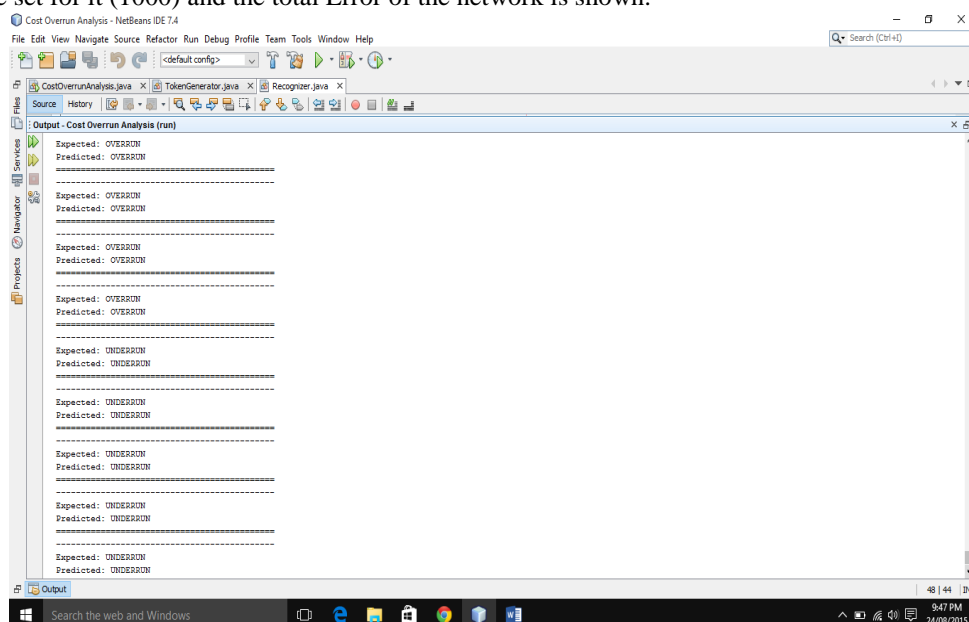


Fig. 5

Figure 5 shows the testing phase of the Neural Network System which we trained initially. The results we get from our test are 90% accurate as we can see that the Expected Class and Predicted Class are matched in all the 10 cases (we used 10 test cases).

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

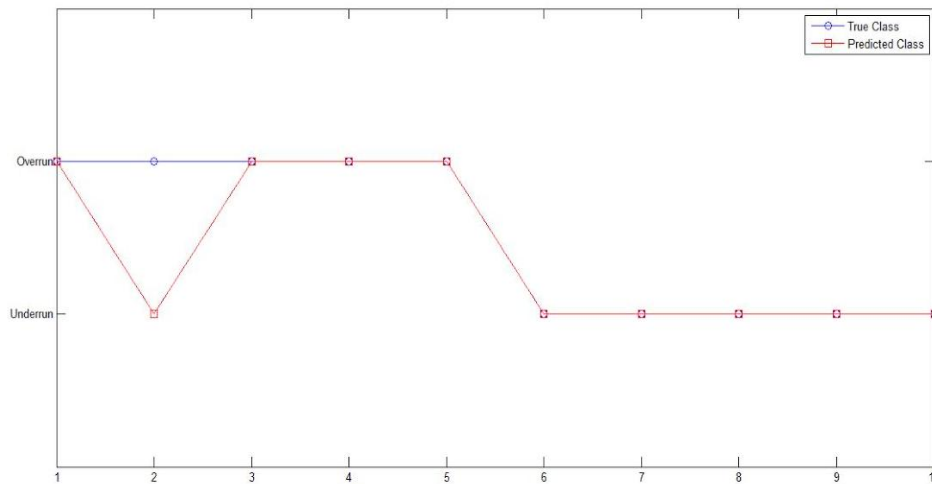
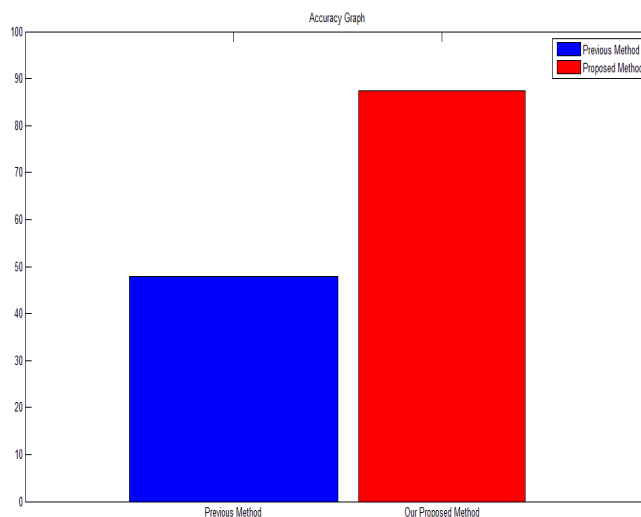


Fig .6

Fig 6 shows class prediction graph showing the accuracy of the system. The Blue Line with Circle shows True Class and the Red with Square is Predicted Class

This problem is solved by our neural network that it improves the accuracy approx 50% from the previous system which is very important part in predicting construction cost overrun to know about the cost of big projects so as to deal with it according to our budget. This is the accuracy comparison graph. We are comparing the accuracy of our system with the current system and we can see that our proposed system (red) is better and has an aggregate accuracy of 87.5%.



## VI. CONCLUSION

In this thesis, we have studied the natural language processing and its application in real life scenario. A problem of Construction Cost Overturn Prediction based on previous knowledge is studied and then implemented using Artificial Neural Network as Classifier. The results were compiled and we found out that due to a high correlation in the data extracted from text-file, the Artificial Neural Network when trained is able to classify the construction cost overrun by



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

90% accuracy. Although, the results can be approximated to 87% accuracy when tested on a very large test set. The result shown above concludes that the system we have implemented will produce a high performance with great accuracy that also helps to improve the speed and remove redundancy that makes it easy to interpret and process fast.

## REFERENCES

1. Williams, Trefor P., and Jie Gong. "Predicting construction cost overruns using text mining, numerical data and ensemble classifiers." *Automation in Construction* 43 (2014): 23-29.
2. Khan, Atif, NaomieSalim, and Yogan Jaya Kumar. "A framework for multi-document abstractive summarization based on semantic role labelling." *Applied Soft Computing* 30 (2015): 737-747.
3. K. Petrousatou, E. Georgopoulos, S. Lambropoulos, J.P. Pantouvakis, Early cost estimating of road tunnel construction using neural networks, *J. Constr. Eng.* 138 (6) (2011) 679–687.
4. C. Gkritska, S.S. Labi, Estimating cost discrepancies in highway contracts: multistep econometric approach, *J. Constr. Eng.* 134 (12) (2008) 953–962
5. S.M. Trost, G.D. Oberlender, Predicting accuracy of early cost estimates using factor analysis and multivariate regression, *J. Constr. Eng.* 129 (2) (2003) 198–204.
6. J. Zhang, N.M. El-Gohary, Information transformation and automated reasoning for automated compliance checking in construction, *Proceedings of the ASCE International Workshop on Computing in Civil Engineering*, Los Angeles, CA, 2013, pp. 701–708.
7. H. Son, C. Kim, C. Kim, Hybrid principal component analysis and support vector machine model for predicting the cost performance of commercial building projects using pre-project planning variables, *Autom. Constr.* 27 (2012) 60–66.
8. K.M. Nassar, W.M. Nassar, M.Y. Hegab, Evaluating cost overruns of asphalt paving project using statistical process control methods, *J. Constr. Eng.* 131 (11) (2005) 1173–1178.
9. T.P. Williams, Bidding ratios to predict highway project costs, *Eng. Constr. Archit. Manag.* 12 (1) (2005) 38–51.
10. S.M. Weiss, N. Indurkha, T. Zhang, *Fundamentals of Predictive Text Mining*, SpringerVerlag, New York., 2010
11. Sibarani, E.M. ; Nadial, M. ; Panggabean, E. ; Meryana, S.. "A Study of Parsing Process on Natural Language Processing in Bahasa Indonesia". 2013. *Computational Science and Engineering (CSE)*, 2013 IEEE 16th International Conference.
12. M.A. Hearst, *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics, Untangling Text Data Mining*, 1999, pp. 3–10.
13. C.G. Wilmot, G. Cheng, Estimating future highway construction costs, *J. Constr. Eng.* 129 (3) (2003)272–279.