

## Cross Project and within Project Software Defection prediction using NLP Techniques

Yashwant Kumar <sup>a</sup>, Dr. Vinay Singh <sup>b</sup>

<sup>a</sup>Department of computing and Information Technology, Usha Martin University, Ranchi, Jharkhand

<sup>b</sup>Associate Professor, Department of computing and Information Technology, Usha Martin University, Ranchi, Jharkhand

**Article History:** Received: 2 February 2020; Accepted: 5 June 2020; Published online: 10 December 2020

### Abstract

The concept of cross project and within project software prediction is an important approach that is used to analyze the defects of software. The different scenario is used to analyze the functions of software prediction with the help of NLP Techniques. Defects are widespread in software systems and can cause a variety of problems for software users. The process software technology of predicting the defects are used to include different software components that helps in analyzing the functions of cross projects. There is a wide range of circumstances in the approach of suitable functions. It identifies the possibility of the system on the basis of different assurance of software prediction. The given article helps in analyzing the process of NLP techniques for the prediction of software. The different aspects are used to evaluate the framework which requires functions of cross projects.

**Keywords:** Software defection predictions, NLP techniques, cross projects, software components, and technology.

### Introduction

This study is used to evaluate the functions of cross project by analyzing the defects of software prediction using NLP techniques. There are different consequences used to establish the implications of the software in the level of suitable analysis. A software defect is a coding error that occurs during the software development process. software flaws can lead to information leaks in the digital era. Difficulties with security such software flaws are referred to be software vulnerabilities. The concern and the primary objective of the software development process is to assure quality software at all stages of development. The phase of software prediction are used to improve the quality of software with the help of different analyzing techniques. In this article, the different prediction of software defects has been discussed. The techniques of NLP i.e. Natural processing language are also demonstrated in this study. The different applications and types of NLP technique in cross projects are also discussed in the study. There are different methods such as neural network, and software metrics are described in the study. The suitable analyses of consumption are required to predict the software defects.

### Literature review

#### Software defect prediction

The concept of software prediction plays an important role in evaluating the functions of different software. Predicting software defects is an important aspect of software quality assurance. According to Fernandes et al. (2020), with the fast growth of as software grows in size and quantity, it becomes increasingly difficult to manually check for software faults by analyzing the functions that are used to improve the functionality of software. In the prediction of different software depicts the analyses of the process helps in analyzing the framework of the software. There is very little information to be transferred, especially in cross-project or cross-version circumstances. Most modern editors provide appropriate just-in-time notifications for some syntax problems, but not for complicated syntax problems, many alone semantic problems. The given diagram of software defects prediction are used to evaluate the functions that are used to analyze the software.

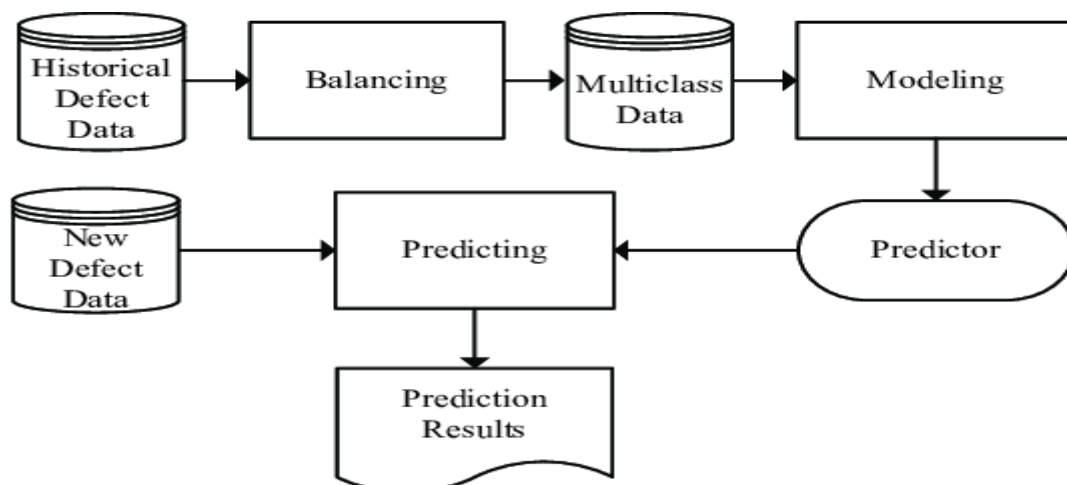


Figure 1. Software Defects prediction

(Source: Abas 2018, p. 548)

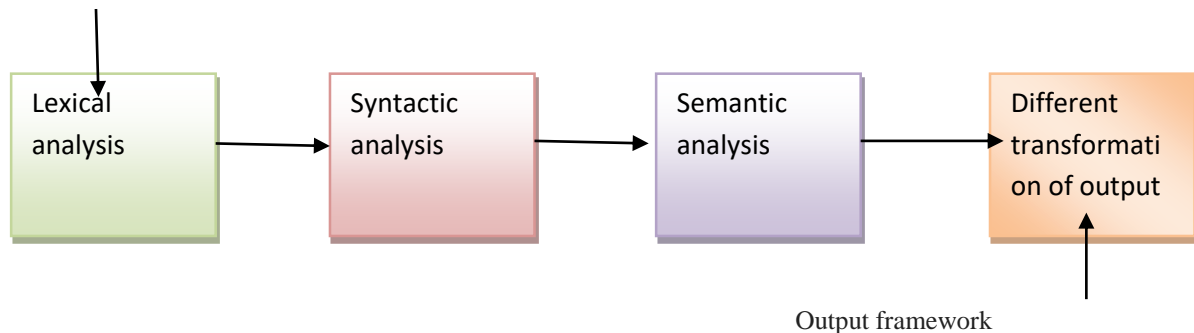
The diagram shows two types of data defect such as historical and new defects data. The historical defect data are used to balance the functions of multi class data by predicting the functions of software. The new defect data propose the software in order to identify the framework of the process. The different characteristics of the software are required to address the overall context of the software. The performance of software defect prediction provides an excellent approach in order to enhance the process of cross projects. The analysis of the process build a strong sources on schemes of expenditure which are mainly used to access resources on level of estimation. The concept of consequences provide the schemes on the function of certain accomplishment. The implementation of formal access plays a role on the dependency of the software. The opportunities of the process are generated in terms of suitable challenges. The concepts of possible outcomes eliminate the consequences on aspects of certain factors (Ilyas et al. 2021).

There is a trend toward source code-based software defect prediction research, and the concept is being implemented progressively in practice. The process of different learning technique has been used in order to analyze the defects of the process. As per AlKilani et al. (2019), the different combinations of the process are used to identify the process which helps to increase the accuracy of prediction. The methodology of cross-projects within software is used to extract the features and metrics of software based on the process classification. The evolution of software defects is brought about by the growth of Natural language processing .prediction (Supangat et al. 2021).

**Concept of NLP techniques**

The process of NLP stands for Natural language processing .The technique of NLP is one of the oldest areas of machine learning research and is employed in significant fields such as machine translation, speech recognition, and text processing. According to Wang (2019), natural language processing has resulted in significant advances in the fields of computers and artificial intelligence. The recurrent neural network is central to many of the techniques used in natural language processing. Natural language processing is used to assess the frameworks employed in the examination of various methodologies. Natural language processing methods are primarily used to recognize processes that are commonly used to monitor and predict parts of software problems. The substantial benefits of the process led the range of expenditure on the criteria of different modification. There is a suitable percentage of relief used in the production of certain enhancement. It is a process which includes different functions that are describe in the diagram.

Input framework



**Figure 2. Natural Language processing Technique (NLP)**

(Source: Fridrich 2020, p. 392)

The diagram of natural language processing i.e. NLP technique consists of different framework such as input and output framework. The prediction of the process required different analyses that are used to transform the output. Lexical analysis, semantic analysis, and syntactic analysis are used to evaluate the processing of the software. . According to Yao and Shepperd (2020), the efficiency of the process impacted the phenomenon in different aspects. There is a concepts of the process on the awareness of certain structure in spite of different analysis of the process. It increases the factor of estimation to decline the concept of suitable schemes which are leading the range of the process. The consequences of the process relate the functions in the appraisal of certain involvement. The increasing efficiency of the process related to the factor of production among the suitable estimation (Dam et al. 2018).

The cost of estimation are rising in the rate of suitable analysis over the source of certain enhancement. . The magnitude of the process is acknowledged on the function of certain analysis. The evaluation of NLP technique is mainly helps in order to demonstrate the functions of the process. It has great importance to demonstrate the complexity of an organization to explore suitable knowledge in the transformation of the output. It illustrates the

importance of the approach of suitable analysis to operate the functions of the software that are used to predict the authenticity of the project (Cai et al. 2020).

### **Applications of NLP techniques in software prediction**

In analyzing the defects of software prediction, the application of NLP technique is used to play an important role. One of the applications of natural language processing that we will examine here is the automatic summarizing of text using software. As per Sterling et al. (2019), the different approach of the process are used to evaluate the functions that are widely used to demonstrate the study. The effectiveness of the process requires different skills to build the function of analysis. The characteristics of the process are identified to examine the functions of the software prediction. It influences the operations on the practices of social strategy in the field of human resources context. The different applications of NLP technique are used in software prediction. The contribution of the process helps in the establishment of different variations which are accompanied to spread the different consequences. As per Prabha and Shivakumar (2020), the study of the given phenomenon is implemented on the emphasis of a particular identity. The practices of the process strongly appear to reflect the function to evaluate the strategy of techniques.

- The process of NLP techniques are used to conduct a large scale analysis based on the functions of the software prediction.
- The techniques of NLP are used to evaluate the process in the context of real-time strategy of software.
- The description of the process is based on the different assumptions which are related to the function of suitable analysis of software.
- It describes the functionality of the process in the circumstances of suitable factors.

The different applications of NLP technique helps in demonstrating the process that are used to predict the defects of software. The specific knowledge helps to describe the criteria for describing the features to execute the performance of the process. The prediction of software examine the smooth performance as well as the application of Defect Prediction models, which are vital in fixing software flaws. Software Defect Prediction (SDP) is one of the most helpful activities used to analyze the functions of software. It distinguishes the modules. Those are malfunctioning and necessitate thorough testing This way the software prediction testing resources are employed efficiently without breaking a cross-project constraint. Natural language processing (NLP) is a scientific approach that are used to evaluate the functions of the software defects. The primary goal of NLP is to create and implement the algorithms that can process and interpret the formation of unstructured language (Bin et al. 2017).

### **Types of NLP techniques in cross projects**

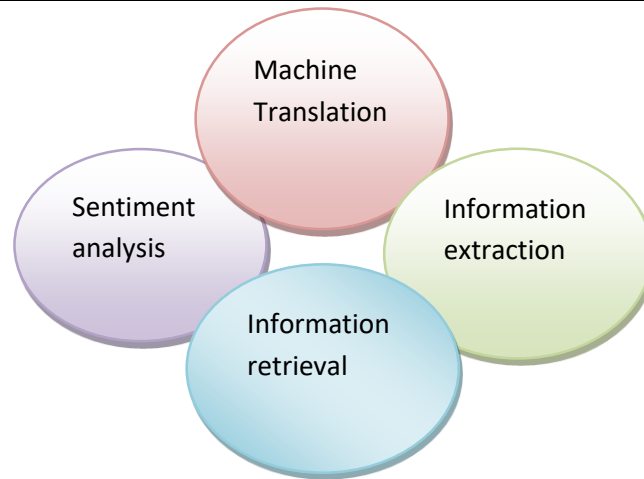
In the field of different defects of software, the process of natural language processing plays a vital role. There are mainly four types of NLP techniques are used in cross projects within the project software defection prediction. Sentiment analysis, machine translation, information extraction, and information retrieval. According to Cao et al. (2018), it is a framework that plays a role in the procedures of certain forms of the system. It analyzes the sequence of the process in the form of a suitable structure of software prediction. The context of NLP technique analyzes the functions by evaluating the efficiency of the projects.

Sentiment analysis- the first type of NLP techniques is sentiment analysis. This analysis is a system that aids in the assignment of process functions by combining natural language processing (NLP) approaches. The various NLP factors are utilized to demonstrate the phases that are extensively employed in the software defect prediction process (Jakhar and Rajnish 2018).

Machine translation- the second type of NLP technique is machine translation. The different scenario of this translation is used to evaluate the framework in the context of the process. The recent trend toward large-scale empirical methodologies has resulted in very significant increases in translation quality, which aids in describing the functions of cross projects.

Information extraction (IE)- the third type of NLP technique is information extraction. The information are extracted with the help of different phase of the software. As per Gauba et al. (2017), the extraction of information consists of different sources that are used to create a relationship between the entities and the attributes. The development of the system enhances the strategy which is appropriate to generate the contents of the cross projects. The higher demands of the process are recognized to engage the fundamental ability on the conditions of different related information.

Information retrieval (IR)- the fourth type of Natural language processing technique is information retrieval. In NLP methodology, information retrieval (IR) is defined as a software programme that deals with the organization, storage, retrieval, and assessment of information from various sources of textual information. e increased functions and the increasing number of the company implements a wide variety to disrupt the property of certain changes. As per Alsawalqah et al. (2020), a huge amount of functions are generated on a different efforts which are required for the development of the projects.



**Figure 3. Types of NLP technique in cross projects**

(Source:Li et al. 2020, p. 549)

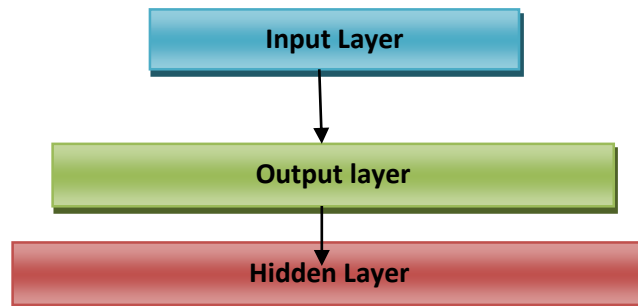
All the four types of NLP technique are used to increase the estimation of cross projects by determining the functions of software defects. The wide range of the functions helps in generating the issues on the boundary of suitable factors of estimation. The high level opportunities are used to reduce the description based on the approach of certain aspects. To manage the functions which are relevant to enable the parameters required for the development of the software. The efficiency of the system relates parameters which are difficult in enhancing the needs of the method. The suitable techniques of NLP are used to describe the phase of project by identifying the defects of the software (Kang et al. 2020).

**Materials and Methods**

The study depicts a different method that helps to describe the functions that are used in the prediction of software project defection. The methods of neural networks and the method of software metrics are used to criticize the development of the software.

**Methods of neural networks**

The concept of neural network plays an important role in analyzing the prediction of software defects. Neural network consists of three layers which are used to demonstrate the functions of the software. It consists of one input, one output, and a number of hidden layers. According to Elallaoui et al. (2018), the layer of the output is used to generate different data sets that are required for the nodes of the network. The methods of neural network is consists of high-level functions based on the approach of datasets in several domains with high efficiency. The formation of different source of data are used to implement the functions that are required to demonstrate the efficiency of the software. . The fragmentation of the system creates a problem on the basis of decision-making criteria. The demonstration of the system evaluates the performance and the development of the software. The estimation of the system criticizes the functions in terms of suitable factors of regulation. The given diagram of neural network are briefly described below.



**Figure 4. Method of Neural network**

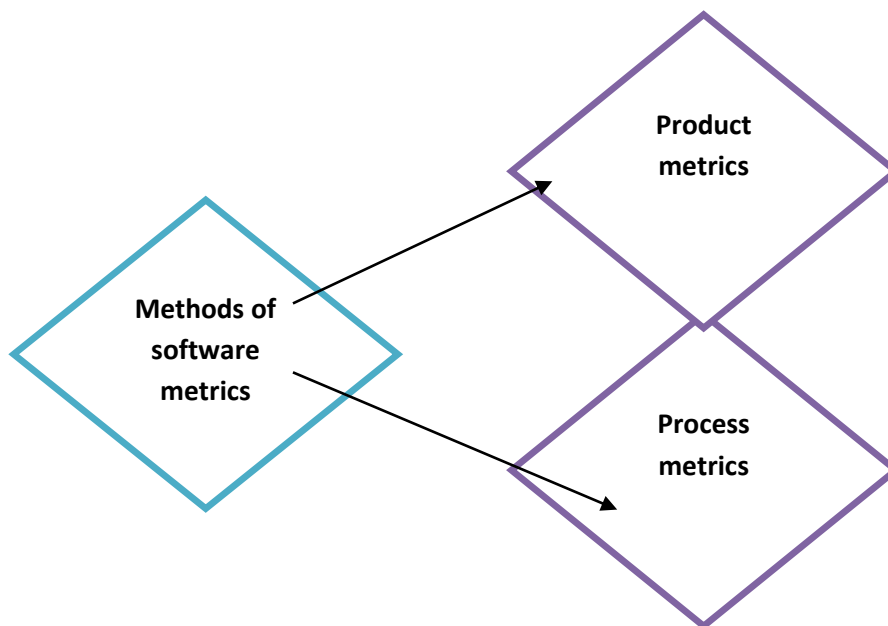
(Source: Song et al. 2018,p. 872)

The approach of software defect prediction acquires software development. The neural network develops new functions that are employed in software implementation. According to Hemmati and Sharifi (2018), it extracts relationships from classical functions that are used to predict software problems. One of the key domains in software engineering that requires the use of novel and effective algorithms is software defect prediction. The

function of code semantics is used to play a vital part in forecasting software defects. The neural network is used to provide different semantic features in the context of source code. The different rate of the process analyzes the systems based on the management of certain acquaintances. The strategy of software helps in rationalizing the functions of water on the advancement level of the information. It mainly used to explore the functions to manifest the higher estimation of the software. The separation of the process establish the level of substances needed for the expenses of the given functions. The efficiency of the network is widely used to describe the functions of the study (Mimura and Miura 2019).

### Methods of software metrics

The method of software metrics are used in predicting or identifying the formation of software based on the efficiency of the system. The study forecasts the performance of software objects for the expansion of an enterprise. System that is efficient and dependable These are the software metrics created for this purpose According to Aussel et al. (2018), it is possible to evaluate quantitatively a software project and use these metrics to assess its performance measurement. The different metrics of software are co-related with operational features of the project. The different features of the software require different functions which helps to evaluate the framework of the project. The concept of software metrics are divided mainly into two groups such as product and process metric. The analysis of both the metrics are exclusive in nature.



**Figure 5. Methods of software metrics in predicting the defects**

(Source: Ahmad and Abduloh 2018, p. 438)

Product metrics- the concept of product metric is the important part of software metrics. These metrics are mainly used to address the defects of the software. The study focuses on including the quantitative goals for the software in the context of development. The different parameters of the process helps to determine the functions that are used to recognize the strategy of software projects. As per Siddiqui (2018), the features and the efficiency of the process enhance the factors that are widely used for the development of the software.

Process metrics- the product metrics is the another important part of software metrics. The analysis of the process regulates the functions of software in order to highlight the defect. . The organizational standards of the process analyze the methods on the sets of norms used in the study. The existence of metrics increases the level of development to maintain the strategy of the process. The perspective of the system is used mainly to correlate the conflict of the system based on the operations of the process. The assumption of the system regulates the source to ensure the defects of the process (Lynn et al. 2017).

### Results and Discussion

In this section, the analysis of software defect prediction are evaluated with the help of Natural language processing technique (NLP). The given tables provide a contradictory analysis on the projects of software defection prediction. The table provides a discussion of both advantages and disadvantages of the technique in software prediction.

SI NO.	Techniques of NLP in software prediction	Advantages of Technique	Disadvantages of Technique
1	Neural network	By using different evaluation of the network, the concept of neural network gives better prediction result.	Not suitable for large number of software defects in terms of cross projects
2	Association rule	Performing operation on the function of association rule are used to provide more accurate result compare to other technique.	It requires a complete set of information which is associated with the software defects.
3	Support vector machine	It is used to generate the process by analyzing the accuracy of the process in order to provide accurate result	The different sets of data are evaluated on the framework of software defects prediction.

The given tables demonstrate the analysis of the process on software defect prediction with the help of Natural language processing (NLP) technique. The table describes the technique of NLP in software prediction. The result of the study says that the cross project within software defection prediction provides advantages in order to regulate the factors of NLP technique. The context of support vector machine enhance different practices which are maintained in order to predict the defects in the software. The different variation shows the overall effectiveness of software to regulate the functions of the process. The study concludes that software defect prediction has a vital role in increasing software quality and lowering software testing time and expense. Based on the results, software’s ability is employed to improve performance (Jusoh 2018).

The techniques of neural network evaluate the study in order to provide the better result. The advancement of the given practices occurs with a variety of software defects which helps to address the outcomes of the study. According to Manigandan et al. (2017), the study depicts different factors which consist of both positive and negative aspects of the prediction. It analyzes the contradiction of different defects on the framework of the process. The table shows different advantages and disadvantages of natural language processing technique in identifying the defects of the software. The high level opportunities are used to reduce the description based on the approach of certain aspects. It helps to manage the functions which are relevant to enable the parameters required for the development of the software. The estimation of the system criticizes the functions in terms of suitable factors of regulation. The approach is based on the strategic part of the management in analyzing suitable factors.

**Conclusions**

From the above study, it can conclude that the analysis of project defection are predicted with the help of NLP techniques. The significant function of the process regulates factors that are evaluated in the context of the process. The primary purpose of this article is to study the capabilities and prospects of software prediction, which aids in the performance of cross-projects. It is used to determine the optimal algorithms for this goal. The study helps in monitoring the process which is required to depict the functions of software prediction. The concept of NLP techniques is used to play an important role in identifying the functions of software defects. The many cross projects assist in examining the framework in the context of software defect prediction to lower the requirements for the size of the labeled datasets. The conclusions of the study is used Software defect prediction (SDP) is a fast-growing topic of research in software engineering. SDP is a method for predicting defects in software. Software Defect Prediction is a process used to anticipate the flaws that make a software module reliable. Using the NLP methodology, SDP is used to improve the functionality of project software defects.

**References**

1. Abas, A., 2018. A Model Development For Software Defect Prediction: Using Feature Reduction Method (Doctoral dissertation).
2. Ahmad, Y.B. and Abduloh, A., 2018. NLP MODELLING TECHNIQUE IN REDUCING PRONUNCIATION ERRORS. English Language Teaching and Research, 2(1).
3. Al Kilani, N., Tailakh, R. and Hanani, A., 2019, October. Automatic classification of apps reviews for requirement engineering: Exploring the customers need from healthcare applications. In 2019 Sixth

- International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 541-548). IEEE.
4. Alsawalqah, H., Hijazi, N., Eshtay, M., Faris, H., Radaideh, A.A., Aljarah, I. and Alshamaileh, Y., 2020. Software defect prediction using heterogeneous ensemble classification based on segmented patterns. *Applied Sciences*, 10(5), p.1745.
  5. Aussel, N., Petetin, Y. and Chabridon, S., 2018, September. Improving performances of log mining for anomaly prediction through nlp-based log parsing. In 2018 IEEE 26th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS) (pp. 237-243). IEEE.
  6. Bin, Y., Zhou, K., Lu, H., Zhou, Y. and Xu, B., 2017, November. Training data selection for cross-project defection prediction: which approach is better?. In 2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM) (pp. 354-363). IEEE.
  7. Cai, X., Niu, Y., Geng, S., Zhang, J., Cui, Z., Li, J. and Chen, J., 2020. An under-sampled software defect prediction method based on hybrid multi-objective cuckoo search. *Concurrency and Computation: Practice and Experience*, 32(5), p.e5478.
  8. Cao, Y., Ding, Z., Xue, F. and Rong, X., 2018. An improved twin support vector machine based on multi-objective cuckoo search for software defect prediction. *International Journal of Bio-Inspired Computation*, 11(4), pp.282-291.
  9. Dam, H.K., Tran, T. and Ghose, A., 2018, May. Explainable software analytics. In Proceedings of the 40th International Conference on Software Engineering: New Ideas and Emerging Results (pp. 53-56).
  10. Elallaoui, M., Nafil, K. and Touahni, R., 2018. Automatic transformation of user stories into UML use case diagrams using NLP techniques. *Procedia computer science*, 130, pp.42-49.
  11. Fernandes, S., Gawas, R., Alvares, P., Femandes, M., Kale, D. and Aswale, S., 2020, February. Survey on various conversational systems. In 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) (pp. 1-8). IEEE.
  12. FRIDRICH, M., 2020. UNDERSTANDING CUSTOMER CHURN PREDICTION RESEARCH WITH STRUCTURAL TOPIC MODELS. *Economic Computation & Economic Cybernetics Studies & Research*, 54(4).
  13. Gauba, H., Kumar, P., Roy, P.P., Singh, P., Dogra, D.P. and Raman, B., 2017. Prediction of advertisement preference by fusing EEG response and sentiment analysis. *Neural Networks*, 92, pp.77-88.
  14. Hemmati, H. and Sharifi, F., 2018, April. Investigating nlp-based approaches for predicting manual test case failure. In 2018 IEEE 11th International Conference on Software Testing, Verification and Validation (ICST) (pp. 309-319). IEEE.
  15. Ilyas, M., Malik, N., Bilal, A., Razzaq, S., Maqbool, F. and Abbas, Q., 2021. Plagiarism Detection Using Natural Language Processing Techniques. *Technical Journal*, 26(01), pp.90-101.
  16. Jakhar, A.K. and Rajnish, K., 2018. Software Fault Prediction with Data Mining Techniques by Using Feature Selection Based Models. *International Journal on Electrical Engineering & Informatics*, 10(3).
  17. Jusoh, S., 2018. A STUDY ON NLP APPLICATIONS AND AMBIGUITY PROBLEMS. *Journal of Theoretical & Applied Information Technology*, 96(6).
  18. Kang, Y., Cai, Z., Tan, C.W., Huang, Q. and Liu, H., 2020. Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics*, 7(2), pp.139-172.
  19. Li, F., Qu, Y., Ji, J., Zhang, D. and Li, L., 2020. Active Learning Empirical Research on Cross-Version Software Defect Prediction Datasets. *International Journal of Performability Engineering*, 16(4).
  20. Lynn, V., Son, Y., Kulkarni, V., Balasubramanian, N. and Schwartz, H.A., 2017, September. Human centered NLP with user-factor adaptation. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (pp. 1146-1155).
  21. Manigandan, T.V.V.D.V.N.B., Vidhya, V., Dhanalakshmi, V. and Nirmala, B., 2017, August. Tamil character recognition from ancient epigraphical inscription using OCR and NLP. In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS) (pp. 1008-1011). IEEE.
  22. Mimura, M. and Miura, H., 2019. Detecting unseen malicious VBA macros with NLP techniques. *Journal of Information Processing*, 27, pp.555-563.
  23. Prabha, C.L. and Shivakumar, N., 2020, June. Software Defect Prediction Using Machine Learning Techniques. In 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184) (pp. 728-733). IEEE.
  24. Siddiqui, Z.E.B.A., 2018. English language teaching through nlp: Techniques and methods. *Research Journal of English Language and Literature*, 6(2), pp.181-184.
  25. Song, J., Kim, J. and Lee, J.K., 2018. NLP and deep learning-based analysis of building regulations to support automated rule checking system. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction (Vol. 35, pp. 1-7). IAARC Publications.
  26. Sterling, N.W., Patzer, R.E., Di, M. and Schrage, J.D., 2019. Prediction of emergency department patient disposition based on natural language processing of triage notes. *International journal of medical informatics*, 129, pp.184-188.

27. Supangat, S., Saringat, M.Z.B., Kusnanto, G. and Andrianto, A., 2021. Churn Prediction on Higher Education Data with Fuzzy Logic Algorithm. *SISFORMA*, 8(1), pp.22-29.
28. Wang, K., 2019. An Evaluation of Programming Language Models' performance on Software Defect Detection. arXiv preprint arXiv:1909.10309.
29. Yao, J. and Shepperd, M., 2020. Assessing software defection prediction performance: Why using the Matthews correlation coefficient matters. In *Proceedings of the Evaluation and Assessment in Software Engineering* (pp. 120-129).