AUTOMATED CODE REVIEW USING NATURAL LANGUAGE PROCESSING

Rationale/Introduction

Software development has grown exponentially in complexity, leading to an increasing need for efficient and accurate code review processes. Traditional code review methods rely heavily on human expertise, which, while effective, can be time-consuming and prone to errors. The emergence of Natural Language Processing (NLP) offers a promising approach to automating code reviews by analyzing code comments, documentation, and structure to provide actionable insights (Ray et al., 2016). Automated code review using NLP can enhance software quality by detecting potential issues, improving readability, and ensuring compliance with coding standards (Allamanis et al., 2018).

Despite its potential, NLP-driven code review systems face challenges such as understanding the semantics of different programming languages, context-aware analysis, and integration with existing development workflows. This study aims to explore the application of NLP in automated code review, evaluating its effectiveness, challenges, and potential enhancements to improve software development efficiency.

Significance of the Study

This study is significant as it contributes to advancing software engineering practices by introducing AI-driven solutions to automate and improve code review processes. By leveraging NLP, developers can reduce the time spent on manual code reviews, minimize human errors, and maintain code consistency across projects. Additionally, the findings of this study will provide insights for organizations seeking to integrate AI-based tools into their development pipelines, ultimately enhancing productivity and software reliability.

Scope and Limitations of the Study

The study focuses on the implementation and evaluation of NLP techniques in automated code review. It will analyze how NLP models can interpret code comments, detect inconsistencies, and suggest improvements in various programming languages. However, the research is limited to textual-based analysis and does not include execution-based testing or performance optimization techniques. Moreover, while integration with version control

systems will be discussed, practical deployment in large-scale software projects is beyond the study's scope.

Objectives of the Study

- To analyze the role of NLP in understanding and reviewing source code.
- To evaluate the accuracy of NLP models in detecting errors, inconsistencies, and stylistic issues.
- To identify challenges in implementing NLP-based automated code review and propose potential solutions.
- To compare NLP-driven code review with traditional manual code review methods in terms of efficiency and effectiveness.
- To propose a framework for integrating NLP-based code review into modern software development workflows.

Expected Outputs

A comprehensive analysis of NLP applications in automated code review, including its advantages, limitations, and areas for improvement. The research will present a prototype or conceptual model for an NLP-driven code review system, demonstrating its ability to detect errors and provide useful feedback. Additionally, comparative insights between manual and automated code reviews will be discussed, offering recommendations for improving AI-based code review tools. The study's findings will be beneficial for software engineers, AI researchers, and organizations looking to enhance code quality and streamline development processes.

REFERENCES

Allamanis, M., Barr, E. T., Bird, C., & Sutton, C. (2018). "A survey of machine learning for big code and naturalness." *ACM Computing Surveys, 51*(4), 81.

Ray, B., Posnett, D., Filkov, V., & Devanbu, P. (2016). "A large-scale study of programming languages and code quality in GitHub." *ACM SIGSOFT International Symposium on Foundations of Software Engineering*.