International Journal of Sciences and Innovation Engineering
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: www.ijsci.com
Vol.03 No.01
E-ISSN : 3049-0251
DOI: https://doi.org/10.70849/ijsci0301260356695

AI-Based Project Originality Checker for Resumes

DEV DHARRSHAN S[1], DIVYA DHARSHINI S[2]
[1,2] Kongu Engineering college,Computer Science And Design,Virudhunagar,Tamil Nadu,India.

Article Info

Article History:

Published:15 Jan 2026

Publication Issue:
Volume 3, Issue 01
January-2026

Page Number:
299-305

Corresponding Author:
DEV DHARRSHAN S

Abstract:

In recent years, resume-based project evaluation has become increasingly challenging due to the repetition and duplication of commonly implemented academic and technical projects. Recruiters and academic evaluators often struggle to verify the originality of projects listed in resumes, leading to inefficient screening processes and reduced emphasis on innovation. To address this issue, this project proposes an AI-based Project Originality Verification System that automatically analyzes and validates the uniqueness of projects mentioned in resumes.The proposed system leverages Natural Language Processing (NLP) and machine learning techniques to extract project titles, descriptions, objectives, and technologies from uploaded resumes. Using advanced semantic similarity models, the extracted project information is compared against an existing database of previously submitted projects and publicly available project repositories. Unlike traditional keyword-based matching, the system employs contextual understanding to identify concept-level similarities, ensuring accurate detection of duplicated or highly similar projects.Based on a calculated similarity score, the system classifies each project as either existing or novel. Projects that exceed a predefined similarity threshold are automatically rejected, while projects demonstrating sufficient originality are accepted. The system also provides transparency by reporting similarity scores and highlighting closely related projects when applicable.This AI-driven approach significantly reduces manual verification effort, improves screening efficiency, and promotes creativity and originality among candidates. The solution is scalable and adaptable for use in academic institutions, recruitment platforms, hackathons, and internship evaluations. Future enhancements include code-level plagiarism detection, integration with version control platforms, and multilingual resume support.

Keywords: Artificial Intelligence, Resume Screening, Project Originality Detection, Natural Language Processing (NLP), Semantic Similarity Analysis, Machine Learning, Plagiarism Detection, Automated Project Evaluation, Text Embeddings, Innovation Validation.

1. Introduction

In today's competitive academic and professional environment, projects play a crucial role in evaluating a candidate's technical skills, creativity, and problem-solving ability. However, with the rapid growth of online resources, open-source repositories, and project-sharing platforms, many commonly used project ideas are repeatedly implemented and reused. This repetition makes it difficult for recruiters, academic institutions, and evaluators to accurately assess the originality and true contribution of a candidate based solely on the projects listed in a resume.

Traditional resume screening methods rely heavily on manual review, keyword matching, or superficial project titles, which are time-consuming, subjective, and prone to human error. As a result, genuinely innovative projects may go unnoticed, while duplicated or minimally modified projects receive undue recognition. There is a growing need for an automated and intelligent system that can verify the uniqueness of projects efficiently and consistently.

This project introduces an **AI-based Project Originality Verification System** designed to analyze projects mentioned in resumes and determine whether they already exist or represent a novel idea. By employing **Natural Language Processing (NLP)** and **machine learning techniques**, the system extracts project-related information and compares it with an existing project knowledge base and publicly available repositories. Instead of relying on exact keyword matching, the system uses semantic similarity analysis to understand the underlying meaning and context of each project.

Based on similarity scores, the system classifies projects as either existing or new, enabling automatic acceptance or rejection decisions. This approach enhances fairness, reduces manual effort, and encourages innovation among candidates. The proposed system can be effectively applied in academic evaluations, campus recruitment, hackathons, and internship screening processes, making project assessment more reliable, scalable and objectives.

2. Literature Review

The rapid growth of digital repositories, open-source platforms, and online learning resources has led to an increasing repetition of similar academic and technical projects across resumes and project submissions. Several studies have highlighted the limitations of traditional resume screening methods, which primarily rely on manual evaluation or keyword-based filtering. These approaches often fail to assess the originality and actual contribution of a candidate, making them inefficient for large-scale evaluations.

Early research in **resume screening automation** focused on rule-based systems and keyword matching techniques. While these systems improved processing speed, they lacked contextual understanding and were unable to detect semantically similar project descriptions that used different wording. As a result, duplicated projects with minor textual changes often went undetected.

With advancements in **Natural Language Processing (NLP)**, more sophisticated techniques such as **Term Frequency–Inverse Document Frequency (TF-IDF)** and cosine similarity were introduced to measure textual similarity between documents. Although these methods improved detection accuracy compared to simple keyword matching, they still struggled with understanding semantic meaning and context, especially in technical project descriptions.

Recent studies have demonstrated the effectiveness of **word embeddings and transformer-based models**, such as Word2Vec, GloVe, BERT, and Sentence-BERT, in capturing semantic relationships between texts. These models enable comparison of project descriptions at a conceptual level rather than relying on exact word matches. Research in plagiarism detection and document similarity has shown that embedding-based similarity analysis significantly improves the identification of paraphrased or structurally similar content.

Several existing plagiarism detection systems focus on academic papers or source code comparison. However, limited research has been conducted on applying AI-driven originality detection specifically

to **resume-based project evaluation**. Most commercial Applicant Tracking Systems (ATS) prioritize skill matching and experience ranking rather than verifying project novelty.

Recent developments in **AI-based decision support systems** have shown promise in automating acceptance and rejection decisions using similarity thresholds and classification models. These systems improve fairness, scalability, and consistency in evaluation processes. However, challenges such as dataset availability, threshold calibration, and bias reduction remain active research areas.

Based on the reviewed literature, there is a clear research gap in developing a dedicated AI system that evaluates the originality of projects mentioned in resumes using semantic similarity analysis. The proposed project addresses this gap by integrating advanced NLP techniques with machine learning-based decision logic to create an efficient and scalable project originality verification framework.

3. Case and Methodology

Case Description

The proposed system is designed for scenarios where resumes containing project details are submitted for evaluation, such as campus placements, internship screening, hackathons, or academic project reviews. In these cases, evaluators need to determine whether the projects mentioned by candidates are original or already widely implemented. Manual verification is inefficient, subjective, and not scalable when handling a large number of resumes.

In the considered case, a candidate uploads a resume containing one or more project descriptions. The system automatically analyzes each project and compares it with an existing repository of previously submitted projects and publicly available project data. Based on the similarity analysis, the system decides whether the project is novel or already existing. If the project is found to be highly similar to existing projects, it is rejected; otherwise, it is accepted. This automated process ensures consistent, unbiased, and efficient project evaluation.

Methodology

The methodology of the proposed AI-based project originality verification system consists of the following stages:

1. Resume Input and Preprocessing

Resumes are accepted in common formats such as PDF or DOC. Text extraction techniques are applied to convert the resume into machine-readable text. Preprocessing steps such as tokenization, lowercasing, stop-word removal, and lemmatization are performed to clean and normalize the extracted content.

2. Project Information Extraction

Using NLP-based techniques, the system identifies and extracts project-related sections from the resume. Key elements such as project title, description, objectives, tools, and technologies used are captured for further analysis.

3. Feature Representation

The extracted project text is converted into numerical representations using semantic embedding models. Transformer-based models such as Sentence-BERT are used to generate embeddings that capture the contextual meaning of project descriptions.

4. Similarity Analysis

The generated embeddings are compared with embeddings stored in the project database using similarity metrics such as cosine similarity. This step measures how closely the new project matches existing projects at a semantic level.

5. Decision Making

A predefined similarity threshold is applied to classify the project:

>If the similarity score exceeds the threshold, the project is classified as **existing** and rejected.

>If the similarity score is below the threshold, the project is classified as **new** and accepted.

6. Output Generation

The system displays the final decision along with the similarity score. Optionally, it may highlight the most similar existing projects for transparency and further review.

4. Results & Analysis

Results

The proposed AI-based Project Originality Verification System was tested using a dataset consisting of resumes containing both commonly repeated projects and unique project ideas. The system successfully extracted project information from resumes and performed semantic similarity analysis against the existing project database.

Experimental results show that the system was able to accurately classify projects as either **existing** or **new** based on similarity scores. Projects that were widely implemented or frequently repeated achieved high similarity scores and were correctly rejected by the system. In contrast, projects that introduced new problem statements, methodologies, or innovative approaches produced lower similarity scores and were accepted.

The use of transformer-based semantic embedding models significantly improved detection accuracy compared to traditional keyword-based methods. Even when project descriptions were paraphrased or written using different terminology, the system effectively identified semantic similarities. The automated decision engine provided consistent results across multiple resume evaluations, reducing dependency on manual judgment.

Analysis

The analysis indicates that semantic similarity-based evaluation is more reliable than exact text matching for project originality detection. The system demonstrated robustness in handling variations in language, formatting, and project description styles commonly found in resumes.

Similarity threshold selection played a critical role in system performance. Lower thresholds increased rejection rates and risked false rejections of partially novel projects, while higher thresholds allowed near-duplicate projects to be accepted. An optimal threshold balanced precision and recall, ensuring fair evaluation.

Performance evaluation showed improved efficiency, with the system capable of processing multiple resumes in significantly less time compared to manual review. This makes the solution scalable for large-scale applications such as campus recruitment and hackathon evaluations.

However, the system's accuracy is influenced by the completeness of the project database. Projects that are novel but conceptually similar to existing ones may occasionally be flagged as duplicates. This limitation highlights the need for continuous database updates and adaptive learning mechanisms.

Overall, the results validate the effectiveness of the proposed system in automating project originality assessment, improving fairness, reducing evaluator workload, and encouraging innovation among candidates.

5. Conclusion

This project successfully presents an **AI-based Project Originality Verification System** that addresses the growing challenge of evaluating the uniqueness of projects mentioned in resumes. By integrating Natural Language Processing and machine learning techniques, the system automates the extraction, analysis, and comparison of project descriptions, enabling efficient and objective decision-making.

The use of semantic similarity models allows the system to move beyond traditional keyword-based matching and accurately identify duplicated or highly similar projects even when they are paraphrased or presented in different formats. Experimental results demonstrate that the proposed approach effectively classifies projects as either existing or novel, reducing manual effort while maintaining consistency and fairness in evaluation.

The system proves to be scalable and adaptable for various real-world applications, including academic assessments, campus recruitment, hackathons, and internship screening. By promoting originality and innovation, the solution encourages candidates to focus on meaningful and creative project development rather than repeated implementations.

Although the system performs effectively, its accuracy depends on the quality and coverage of the project database. Future improvements can enhance reliability through continuous learning, integration with code-level analysis, and expansion to multilingual support. Overall, the proposed system provides a practical, intelligent, and impactful solution for automated project originality assessment.

6. Funding Details

This project was supported by institutional resources provided by the academic department. No external grants or commercial funding were involved. The support was limited to access to computing facilities, software tools, and academic guidance.

7. Disclosure Statement

The authors declare that this project titled **"AI-Based Project Originality Verification System for Resume Evaluation"** is an original work developed for academic purposes. All data, tools, algorithms, and software components used in this project are either publicly available, open-source, or properly referenced. No confidential, proprietary, or personal data has been intentionally used or disclosed during the development of this system.

References

1· Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of NAACL-HLT*, 2019.

2· Reimers, N., & Gurevych, I., "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," *EMNLP*, 2019.

3· Salton, G., Wong, A., & Yang, C. S., "A Vector Space Model for Automatic Indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613–620, 1975.

4· Manning, C. D., Raghavan, P., & Schütze, H., *Introduction to Information Retrieval*, Cambridge University Press, 2008.

5· Potthast, M., Stein, B., Barrón-Cedeño, A., & Rosso, P., "Cross-Language Plagiarism Detection," *Language Resources and Evaluation*, vol. 45, pp. 45–62, 2011.

6· Mikolov, T., Chen, K., Corrado, G., & Dean, J., "Efficient Estimation of Word Representations in Vector Space," *ICLR Workshop*, 2013.

7· GitHub Repository API Documentation, "Accessing Public Repositories for Project Analysis," https://docs.github.com/en/rest, 2025.

8· Vaswani, A., Shazeer, N., Parmar, N., et al., "Attention is All You Need," *NeurIPS*, 2017.

9· Alzahrani, S. M., Salim, N., & Abraham, A., "Understanding Plagiarism Linguistic Patterns, Textual Features, and Detection Methods," *IEEE Transactions on Systems, Man, and Cybernetics*, 2012.

10· Zhang, Y., & Wallace, B., "A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification," *IJCAI*, 2015.

DEV DHARRSHAN S

AUTHOR

DIVYA DHARSHINI S

AUTHOR