Enhance Project Management System using Machine Learning

[1]Rajan Salunkhe, [2]Nilesh Kasar, [3]Sujal Bainade, [4]Vishal Pawar

MET BKC Institute of Engineering
Nashik, Maharashtra, India.

***Abstract*- Academic integrity is in focus in today's educational environment, and maintaining academic integrity is of paramount importance. Plagiarism and repetition of research topics can undermine the authenticity of academic work. Empowering Students To address this challenge, we introduce the "Enhance Project Management System using Machine Learning", a cutting-edge machine learning-based system that empowers students to uphold academic honesty and originality in their research projects. This tool provides a seamless user experience. Students log in and input essential project information, including topic, abstract, methodology, objectives, and conclusion. Leveraging a sophisticated algorithm and a vast dataset of academic materials, the system then computes a matching percentage, allowing students to gauge the uniqueness of their work and make informed decisions to enhance the originality of their projects.**

***Keywords*: Educational Environment, Machine Learning, Sales Strategies, Project Repetition**

Published in IJIRMPS (E-ISSN: 2349-7300), Volume 11, Issue 5, Sept-Oct 2023

Introduction

Academic Integrity in Focus In today's educational environment, maintaining academic integrity is of paramount importance. Plagiarism and repetition of research topics can undermine the authenticity of academic work. Empowering Students To address this challenge, we introduce the "Enhance Project Management System using Machine Learning", a cutting-edge machine learning-based system that empowers students to uphold academic honesty and originality in their research projects. the tool provides a seamless user experience. Students log in and input essential project information, including topic, abstract, methodology, objectives, and conclusion. Leveraging a sophisticated algorithm and a vast dataset of academic materials, the system then computes a matching percentage, allowing students to gauge the uniqueness of their work and make informed decisions to enhance the originality of their projects.

Academic integrity is the cornerstone of educational institutions worldwide. It ensures that the knowledge produced and disseminated is reliable and trustworthy. Plagiarism, which involves presenting someone else's work or ideas as your own, poses a significant threat to academic integrity. Similarly, repetitive research topics can hinder the advancement of knowledge by limiting the exploration of new ideas and perspectives. Therefore, in today's educational environment, there is an increasing need for tools and systems that help students maintain academic honesty and foster originality in their academic pursuits.

The "Enhance Project Management System using Machine Learning" is a groundbreaking solution that harnesses the power of machine learning to assist students in upholding academic integrity and promoting the creation of original work. This tool has been designed with user-friendliness in mind, ensuring that students can easily integrate it into their research processes. By logging in and providing essential project information such as the topic, abstract, methodology, objectives, and conclusion, students can harness the capabilities of this system.

At the heart of the "Enhance Project Management System using Machine Learning" is a sophisticated algorithm that leverages a vast and diverse dataset of academic materials. This dataset encompasses various research papers, articles, theses, and dissertations from various disciplines and academic institutions. The algorithm employs natural language processing techniques to analyze the submitted project information and cross-reference it with the extensive dataset.

The key output provided by the system is a matching percentage, which serves as a quantitative measure of the uniqueness of the student's work. This percentage is crucial as it empowers students to make informed decisions about the originality of their projects. By knowing the extent to which their work matches existing research, students can take proactive steps to enhance the originality of their projects.

Literature Survey

Text matching is one of the crucial technologies in Natural Language Processing (NLP), and it has been applied in many tasks, such as textual similarity, information retrieval and question answering. The target of text matching is to model the relationship between two input texts. In this paper, we aim to give a survey on recent advanced techniques of deep-learning-based text-matching methods. Specifically, depending on whether a model will first encode a sentence into a fixed-length vector without any incorporation from the other sentence, the existing studies can be categorized into two major categories: representation-based models and interaction-based models. The latter can be divided into two groups according to the interaction methods. In addition, we summarize the strengths and weaknesses of these methods to help beginners in this area to choose the appropriate model for their application. Finally, we conclude by highlighting several directions and open problems that need to be further explored by the community in the future [1].

Text matching is the task of matching two texts and determining the relationship between them, which has extensive applications in natural language processing tasks such as reading comprehension, and Question-Answering systems. The mainstream approach is to compute text representations or to interact with the text through an attention mechanism, which is effective in text-matching tasks. However, the performance of these models is insufficient for texts that require commonsense knowledge-based reasoning. To this end, in this paper, we introduce a new model for text matching called the Knowledge Enhanced Text Matching (KETM) model, to enrich contextual representations with real-world common-sense knowledge from external knowledge sources to enhance our model understanding and reasoning. First, we use Wiktionary to retrieve the text word definitions as our external knowledge. Secondly, we feed text and knowledge to the text-matching module to extract their feature vectors. The text matching module is used as an interaction module by integrating the encoder layer, the co-attention layer, and the aggregation layer. Specifically, the interaction process is iterated several times to obtain in-depth interaction information and extract the feature vectors of text and knowledge by multi-angle pooling. Then, we fuse text and knowledge using a gating mechanism to learn the ratio of text and knowledge fusion by a neural network that prevents noise generated by knowledge. After that, experimental validation on four datasets is carried out, and the experimental results show that our proposed model performs well on all four datasets, and the performance of our method is improved compared to the base model without adding external knowledge, which validates the effectiveness of our proposed method [2].

Text matching is one of the fundamental research tasks in the field of natural language processing. It can be applied to many NLP tasks, such as information retrieval, question, and answer systems and text repetition. In this paper, we propose a text-matching model with multi-granularity term alignment (MGTA). The model extracts word information at different granularities through convolutional neural networks and enhances the model effect by aligning the original location features at different word granularities, enabling the model to obtain multiple granularities of information during text matching. We conduct experiments on the Q&A dataset, the text-implication dataset, and the paraphrase recognition dataset, respectively, and compare them with current mainstream models in terms of accuracy, MAP and MRR evaluation metrics, and has fewer parameters, which greatly improves the inference speed [3].

A new talent training mode of collaboration between industry and education was proposed, which aims to reduce the gaps in talent definition between enterprise demand and college education. We formulated a scenario as a text-matching task of similarity between course details and job position information and developed an ERNIE+DPCNN-based model for the task. The model has been benchmarked on the LCQMC dataset, the results show that it achieves the art-of-state performance of Acc 0.8809, Recall 0.8745, and F1 score 0.8739, We conducted many experiments about the text similarity calculation between domains of industry and educations. The experiment results show our approach has much potential to optimize the curriculum system while applying in this field [4].

Aim and Objectives

- The aim is to empower students to actively engage in the research process by offering them the means to assess the uniqueness of their projects and make informed adjustments, thereby fostering a sense of ownership and responsibility in their academic work. To elevate the quality of academic output by encouraging students to produce high-quality, original research projects that contribute meaningfully to their chosen fields of study.
- The main objectives are to cultivate a culture of honesty and originality within educational institutions by providing students with a tool that aids in preventing unintentional plagiarism and the repetition of research topics.

Motivation

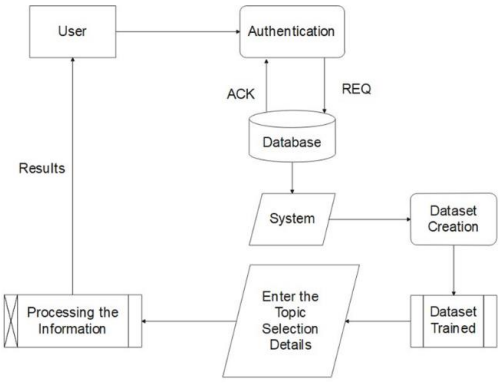
The motivation behind developing the system arises from a genuine concern for upholding the principles of academic integrity and excellence. In today's educational landscape, plagiarism and unintentional repetition of research topics are pervasive issues that compromise the authenticity and quality of student work. This system aims to motivate students by providing them with a tool that enables them to take control of their academic journey. By offering a means to assess the originality of their projects and make informed adjustments, it empowers students to contribute meaningfully to their chosen fields of study.

Proposed System

The system is an innovative machine learning-based system designed to assist students in ensuring the originality and uniqueness of their project submissions. Upon logging into the system, students can input crucial project details, including the project's topic, abstract, methodology, objectives, and conclusion. The system then employs a sophisticated algorithm to compare this information against an extensive dataset of previous projects, research papers, and academic materials. The result is a comprehensive matching percentage that the system provides to the user, enabling students to assess the potential similarity of their work to existing literature and make necessary adjustments to enhance the originality of their projects.

System Architecture

Figure 1: System Architecture Diagram



Application

- **Increased Sales and Revenue:** By tailoring sales offers to individual customers based on their preferences and behaviours, the project can lead to higher conversion rates and increased customer

engagement. Personalized discounts can incentivize customers to make more purchases, thereby boosting overall sales and revenue.

- **Enhanced Customer Loyalty:** The multi-tiered discount system and tailored discount percentages can foster a deeper sense of loyalty among customers. When customers feel that a company understands their needs and preferences, they are more likely to remain loyal, make repeat purchases, and become brand advocates.
- **Improved Customer Satisfaction:** The project's focus on providing discounts aligned with each customer's unique tendencies can lead to higher levels of customer satisfaction. When customers receive offers that resonate with their preferences, they are more likely to have positive interactions with the brand, leading to improved satisfaction scores and positive word-of-mouth.

Algorithm

Building an "Enhance Project Management System using Machine Learning" using machine learning and natural language processing (NLP) involves using NLP techniques to detect and analyze repetitions in each text or set of texts. Here is a step-by-step approach to creating such a system:

Step 1: Data Collection and Pre-processing

Start by gathering a diverse dataset of texts where repetitions or redundancies are likely to occur. This dataset should include a wide range of project submissions or academic materials. Pre-process the collected data. Pre-processing involves tasks like removing punctuation, converting text to lowercase, and tokenizing the text (breaking it into individual words or tokens). This makes the text data consistent and ready for analysis.

Step 2: Feature Extraction

Extract features from the pre-processed text data. In this case, feature extraction could involve using word embeddings or word vectors. Word embeddings capture the semantic meaning of words in a numerical form. Train a Word2Vec or similar model on your pre-processed data. This model will learn to represent words as vectors in a high-dimensional space. These word vectors can be used as features for the NLP model.

Step 3: Labeling Data

Annotate the dataset to indicate instances of repetition or redundancy. This step requires manual or automated labeling of the data. You may assign labels, like "0" for non-repetition, and "1" for repetition, to the project submissions based on the degree of similarity.

Step 4: Model Training

Train a machine learning model using the labeled dataset. The chosen model, in this case, could be an NLP model, such as a Random Forest Classifier, or Support Vector Machine (SVM). During model training, the algorithm should learn to identify patterns associated with repetition in the project submissions. It will use the word embeddings as features to make predictions.

Step 5: Model Evaluation

After training, evaluate the model's performance using a test dataset that was not used during training. Common evaluation metrics include accuracy, precision, recall, and F1-score. Fine-tune the model as needed to improve its accuracy in detecting repetitions and redundancies accurately.

Step 6: Integration with the Project Management System

Integrate the trained model into your "Enhance Project Management System." This system should allow students to input their project details, such as the topic, abstract, methodology, objectives, and conclusion. The system should then utilize the NLP model to compare these details against the extensive dataset of previous projects and research papers. The result should be a matching percentage that indicates the potential similarity of the submitted work to existing literature. Provide feedback to students based on the

matching percentage so that they can make necessary adjustments to enhance the originality of their projects. Ensure a user-friendly interface for students to interact with the system effectively.

System Requirements

Software Used

- Windows 10 or above
- Python 3.10
- SQLite
- Django
- VS Code

Hardware Used

- Intel Core i3 Processor or above
- 150 GB Hard Disk or above
- 4 GB RAM or above

Conclusion

In conclusion, the "Enhance Project Management System using Machine Learning" stands as a pivotal asset in preserving academic integrity and championing originality in research. With its user-friendly interface, seamless integration, and educational orientation, it serves as a valuable resource for students and educators. By instilling self-awareness, guiding research refinement, and facilitating interdisciplinary exploration, this system marks a significant leap forward in the pursuit of academic excellence and integrity. Its transparent interface encourages student responsibility, while its integration capabilities simplify its adoption into academic workflows. This educational focus transforms it into a tool for nurturing original research, bridging the gap between theory and practice. Moreover, it promotes continuous learning and growth, creating a culture of improvement. By allowing students to draw from various disciplines, it stimulates creativity and innovation. In essence, it empowers students, enriches teaching, and contributes to the ever-evolving landscape of academia.

REFERENCES:

1. Kalervo J. Arvelin and Jaana Kek al ainen, "Ir evaluation methods for retrieving highly relevant documents", ACM SIGIR Forum, vol. 51, pp. 243-250, 2017.
2. Yi Yang, Wen-tau Yih and Christopher Meek, "Wikiqa: A challenge dataset for open-domain question answering", Proceedings of the 2015 conference on empirical methods in natural language processing, pp. 2013-2018, 2015.
3. Tushar Khot, Ashish Sabharwal and Peter Clark, "Scitail: A textual entailment dataset from science question answering", Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, 2018.
4. Tomas Mikolov, Kai Chen, Greg Corrado and Jeffrey Dean, Efficient estimation of word representations in vector space, 2013.
5. Jeffrey Pennington, Richard Socher and Christopher D Manning, "Glove: Global vectors for word representation", Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532-1543, 2014.
6. Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, et al., "Deep contextualised word representations", Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 1, pp. 2227-2237, 2018.
7. Yichen Gong, Heng Luo and Jian Zhang, "Natural language inference over interaction space", International Conference on Learning Representations, 2018.
8. Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero and Larry Heck, "Learning deep structured semantic models for web search using clickthrough data", Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, pp. 2333-2338, 2013.