SMARTY PIE

(PROJECT IDEA EVALUATION)

GROUP NO 8

SESSION: BSCS. SPRING 2020

PROJECT SUPERVISOR: Dr.Anam Akbar

PROJECT CO-SUPERVISOR: Dr. Amir Mehmood

SUBMITTED BY

Kashan Khan Ghori

Numair Nasir

Muhammad Zahid

Dinesh Kumar

Computer Science

Sir Syed University of Engineering & Technology

Certification

This is to certify that **Kashan Khan Ghori (BSCS-070-2020S), Numair Nasir (BSCS-072-2020S), Muhammad Zahid (BSCS-075-2020S)** and **Dinesh Kumar (BSCS-095-2020S)** have successfully completed the final project **Smarty Pie,** at the **Sir Syed University** of **Engineering & Technology**, to fulfill the partial requirement of the degree **Bachelor of Science in Computer Science.**

External Examiner

Name of Examiner

Designation

Project Supervisor

Dr.Anam Akbar

Assistant Professor, SSUET

Chairman

Department of Computer Science, Sir Syed University of Engineering & Technology

Project Title (Smarty Pie)

Sustainable Development Goals

(Please tick the relevant SDG(s) linked with FYDP)

SDG No	Description of SDG	SDG No	Description of SDG
SDG 1	No Poverty	SDG 9 ✓	Industry, Innovation, and Infrastructure
SDG 2	Zero Hunger	SDG 10	Reduced Inequalities
SDG 3	Good Health and Well Being	SDG 11	Sustainable Cities and Communities
SDG 4 ✓	Quality Education	SDG 12	Responsible Consumption and Production
SDG 5	Gender Equality	SDG 13	Climate Change
SDG 6	Clean Water and Sanitation	SDG 14	Life Below Water
SDG 7	Affordable and Clean Energy	SDG 15	Life on Land
SDG 8	Decent Work and Economic Growth	SDG 16	Peace, Justice and Strong Institutions
		SDG 17 ✓	Partnerships for the Goals

Range of Complex Problem Solving			
	Attribute	**Complex Problem**	
1	Range of conflicting requirements	Involve wide-ranging or conflicting technical, engineering and other issues.	
2	Depth of analysis required	Have no obvious solution and require abstract thinking, originality in analysis to formulate suitable models.	✓
3	Depth of knowledge required	Requires research-based knowledge much of which is at, or informed by, the forefront of the professional discipline and which allows a fundamentals-based, first principles analytical approach.	✓
4	Familiarity of issues	Involve infrequently encountered issues	
5	Extent of applicable codes	Are outside problems encompassed by standards and codes of practice for professional engineering.	
6	Extent of stakeholder involvement and level of conflicting requirements	Involve diverse groups of stakeholders with widely varying needs.	
7	Consequences	Have significant consequences in a range of contexts.	
8	Interdependence	Are high level problems including many component parts or sub-problems	
Range of Complex Problem Activities			
	Attribute	**Complex Activities**	
1	Range of resources	Involve the use of diverse resources (and for this purpose, resources include people, money, equipment, materials, information and technologies).	✓
2	Level of interaction	Require resolution of significant problems arising from interactions between wide ranging and conflicting technical, engineering or other issues.	
3	Innovation	Involve creative use of engineering principles and research-based knowledge in novel ways.	✓
4	Consequences to society and the environment	Have significant consequences in a range of contexts, characterized by difficulty of prediction and mitigation.	✓
5	Familiarity	Can extend beyond previous experiences by applying principles-based approaches.	

Abstract

In today's rapidly growing world of technology, coming up with new and unique ideas for projects is very important. But figuring out how novel a project can be a big challenge. This is even harder when trying to connect these projects with the globally accepted 17 Sustainable Development Goals (SDGs). The goals are detailed and interconnected, making it difficult for project managers, researchers, evaluators, and students to fully understand their implications. Furthermore, it is also important to check how new and novel these projects are, as fresh ideas are key to achieving these goals. This shows that there is a need for a tool that is easy to use and can do these tasks automatically and with high accuracy. The proposed model has two primary phases: novelty assessment and SDG identification. The first phase focuses on the degree of novelty of a project. The second phase primarily centers around aligning the project with the Sustainable Development Goals (SDGs). The model uses advanced technologies including Natural Language Processing (NLP), web scraping, machine learning techniques, trend analysis algorithms, and similarity scoring methods to check novelty and SGD's. Upon completing the analysis, our proposed model generates a downloadable report containing the results based on the project summaries provided by users. This aspect ensures that users have a tangible output that they can refer to, share, or use for further planning and implementation. This model eliminates the lacking between the SDGs and novelty assessment, effectively addressing the identified gap. By automating the process, it not only simplifies the tasks for users, but also enhances the accuracy of the results. Hence, this model presents a significant solution to the pressing need for efficient, accurate SDG alignment and novelty evaluation in project planning and execution.

Keywords: *project selection, idea evaluation, decision making, computer-based sustainable goal identification, computer-based novelty assessment*

Undertaking

I certify that the project **Smarty Pie** is our own work. The work has not, in whole or in part, been presented elsewhere for assessment. Where material has been used from other sources it has been properly acknowledged/ referred.

Kashan Khan Ghori

BSCS-070-2020

Numair Nasir

BSCS-072-2020

Muhammad Zahid

BSCS-075-2020

Dinesh Kumar

BSCS-095-2020

Acknowledgement

We truly acknowledge the cooperation and help make by **Dr.Anam Akbar, Assistant Professor** of **Sir Syed University of Engineering & Technology.** She has been a constant source of guidance throughout the course of this project. We would also like to thank **Dr. Amir Mehmood**, **Associate Professor** of **Sir Syed University of Engineering & Technology** for his help and guidance throughout this project.

We are also thankful to our friends and families whose silent support led us to complete our project.

Table of Contents

Certification **i**

Abstract **iv**

Undertaking **v**

Acknowledgement **vi**

Table of Contents **vii**

List of Tables **viii**

List of Figures **ix**

List of Acronyms **x**

List of Equations **xi**

Chapter 1 1

1.1 Introduction

1.2 Statement of the Problem

1.3 Goals/Aim & Objectives

1.4 Motivation

1.5 Assumptions & Dependencies

1.6 Methods

1.7 Report Overview

Chapter 2 2

2.1 Overall System Model

2.2 System Data Flow

Chapter 3 3

3.1 Literature Review with Comparative Analysis

Chapter 4 4

4.1 Proposed Solution/Result & Discussion

Chapter 5 5

5.1 Summary and Future Work

Chapter 6 6

6.1 Conclusion and Recommendation

References **7**

Annexure **8**

N/A

List of Tables

Table-1: Comparative Analysis Table

Table-2: PERT Activity Time estimate table.

List of Figures

Figure-1: Overall System Mode Diagram

Figure-2: System Data Flow Diagram

List of Acronyms

NLP Natural Language Processing

PERT Program Evaluation Review Technique

BERT Bidirectional Encoder Representations from Transformers

SDG Sustainable Development Goals

ML Machine Learning

PIE Project Idea Evaluation

List of Equations

N/A

Chapter 1

1.1 Introduction

In today's technological renaissance, where innovation forms the cornerstone of competitiveness, the real challenge lies not just in the generation of ideas, but in their genuine novelty and global resonance. Our project redefine the parameters of project evaluation. This project is not merely a conceptual exploration but a concrete solution, meticulously engineered to assess the novelty of ideas across a wide spectrum, from fledgling startups to seasoned enterprises. The crux of the project is twofold: to deliver an objective measure of a project's innovation and to ensure its alignment with the ambitious Sustainable Development Goals (SDGs). By doing so, we address the dual challenge that current evaluative methods face—a lack of objective novelty assessment and a fragmented approach to SDG alignment. Our project transcends these hurdles by integrating advanced algorithms with a user-centric design, facilitating a nuanced analysis that discerns the uniqueness of projects and their potential global impact.

Furthermore, this project is a testament to the synergistic potential of technology and sustainability. It serves as a beacon for responsible innovation, guiding project creators on a path that not only fulfills market needs but also propels us towards a more sustainable future. With this initiative, we are setting in motion a paradigm where the assessment of project viability is as much a science as it is an art, promising a future where every creative endeavor is evaluated against the touchstone of global standards and ethical benchmarks

1.2 Statement of the problem

The process of evaluating project summaries and aligning them with Sustainable Development Goals (SDGs) presents a multifaceted challenge, often constrained by time and expertise. Traditional manual evaluations, reliant on extensive literature reviews, are not only labor-intensive but also fall prey to subjective biases, leading to inconsistent outcomes. This subjectivity potentially stifles the recognition and nurturing of truly novel projects, especially in fields where evaluators may lack domain-specific knowledge. The absence of a specialized tool for such assessments hinders the identification and development of innovative project ideas. Therefore, a pressing need arises for an intuitive, comprehensive application that can automate and refine the evaluation process, bolstering accuracy, and fostering innovation in project design.

1.3 Goals/Aims & Objectives

The proposed smart app aims to revolutionize project evaluation by achieving the following objectives:

1. **Automated Project Evaluation:** Efficiently assesses students' projects with minimal human intervention.
2. **SDG Alignment Identification:** Employ Machine Learning techniques to meticulously map projects with the United Nations' 17 Sustainable Development Goals, ensuring global relevance.
3. **Novelty Assessment and Enhancement:** Analyze projects against existing benchmarks in the domain to gauge novelty, offering constructive feedback for potential enhancements to elevate originality.
4. **User-Centric Design for Broader Impact:** Craft a user-friendly interface that caters to a wide audience, including project managers, researchers, evaluators, and students, fostering a culture of sustainable development and innovation.
5. **Facilitation for Students and Evaluators**: Aids in creating and assessing projects that meet global excellence standards.

1.4 Motivation

Our motivation stems from the urgent need to enhance the objectivity and efficiency in project evaluations. Recognizing the limitations of current manual methods, which are often subjective and inconsistent, we were driven to create a solution that harnesses the power of technology. By integrating Natural Language Processing and Machine Learning, our tool not only evaluates project novelty but also aligns it with the Sustainable Development Goals. This alignment is crucial in promoting global standards in academic and professional projects. Our vision is to empower students and educators with a tool that guides, improves, and revolutionizes the way project ideas are brought to life and evaluated against global benchmarks.

1.6 Assumption and Dependencies

The project operates under several key assumptions and dependencies:

1. **Data Availability**: It presumes access to a rich dataset of project summaries for analysis.
2. **Technological Reliability**: The project's effectiveness is closely tied to the performance and accuracy of current NLP and ML technologies.
3. **Novelty Assessment:** The project assumes a consistent and evolving understanding of what constitutes 'novelty' within various project domains. This includes the availability of comprehensive and up-to-date databases against which the novelty of new projects can be accurately assessed and benchmarked.
4. **SDG Framework Stability:** It assumes that the Sustainable Development Goals will remain a stable and relevant benchmark for evaluating project alignment.

5. **User Engagement:** The success of the project is contingent on active participation from its target audience, including students, evaluators, and project managers, who are expected to use the tool regularly for project assessment.

1.7 Methods

SDG Methodology & Approach

A) Multilabel Text-Classification

To accurately align project summaries with the 17 Sustainable Development Goals, we employ a multi-label text classification algorithm. This sophisticated approach enables the assignment of multiple SDGs to a single text, reflecting the multifaceted nature of most projects. Central to our method is the integration of BERT (Bidirectional Encoder Representations from Transformers), a cutting-edge NLP algorithm renowned for its efficacy in processing complex language tasks. This integration not only enhances the precision of our SDG alignment but also ensures that each project is evaluated in a comprehensive and nuanced manner, truly reflecting its alignment with multiple global goals.

B) Bert Algorithm

In our project, we employ the Bidirectional Encoder Representations from Transformers (BERT) architecture, a neural network framework by Google, which has revolutionized Natural Language Processing tasks like text classification. We first partition our dataset, "abstract.csv", into training and testing subsets, allocating 70% for training and 30% for testing. Our data is then formatted to be BERT-compliant, involving the creation of 'InputExample' classes where 'text_a' contains the classification text and 'label' denotes the corresponding SDG labels. Pre-processing steps are undertaken to meet BERT's criteria, including text lowercasing and tokenization. We augment BERT with a new layer, specifically trained for multi-label classification. This phase is critical for learning the textual characteristics of the 17 SDG classes. Post-training, the model, equipped with its learned parameters, is tasked with classifying texts with unassigned classes, predicting their SDG alignment as probabilities using a Logit function. To enhance model robustness and prevent overfitting, dropout techniques are incorporated. The model's accuracy, standing at 95%, is measured by comparing its predicted labels against the actual ones, showcasing its efficacy in SDG classification.

Novelty Methodology & Approach

To assess the novelty of project ideas, our methodology integrates a sophisticated text generation model and web scraping techniques. Initially, we process the input project summary to generate specific queries using the google/flan-t5-large text generation model. These queries are designed to pinpoint similar projects on the internet. Once potential matches are identified, we employ the BeautifulSoup API for web scraping, extracting project summaries from the located web pages. This gathered data serves as a comparison point against the entered project abstract. We then apply a series of algorithms to this comparative analysis, meticulously calculating the novelty of the submitted project idea. This process is not only automated but also ensures a comprehensive evaluation of the project's uniqueness by considering a wide array of existing projects in similar domains.

1.8 Report Overview

Chapter 1 of this report provides a foundational understanding of our innovative project. It begins with an introduction to the necessity for advanced project evaluation tools in our technologically evolving landscape. The problem statement highlights the existing challenges in evaluating project novelty and SDG alignment. Our goals and objectives focus on utilizing NLP and ML for effective project analysis. The motivation behind this initiative is outlined, emphasizing the need for accurate and objective project assessments. Key assumptions and dependencies crucial for the project's execution are discussed, followed by a detailed explanation of the methodologies employed. This chapter sets the stage for a deep dive into the project's intricacies and its potential impact.

Chapter 2

2.1 Overall System Model

The figure presented meticulously outlines the interconnected components of our system model. The starting point is the user submission of a project summary through a device, which then undergoes scanning. The core of the system is our sophisticated model, which employs multi-label text classification to analyze and align the project with multiple SDGs. This approach leverages the BERT algorithm for its powerful NLP capabilities. Concurrently, novelty assessment methodology is applied, using query generation and web scraping techniques to evaluate the project's uniqueness against existing databases. The outcome of this dual-faceted analysis is a comprehensive report that provides a holistic evaluation of the project's innovation and its potential contribution to global goals. Users can download this report, which embodies the system's utility in guiding projects towards achieving a higher standard of novelty and sustainability.

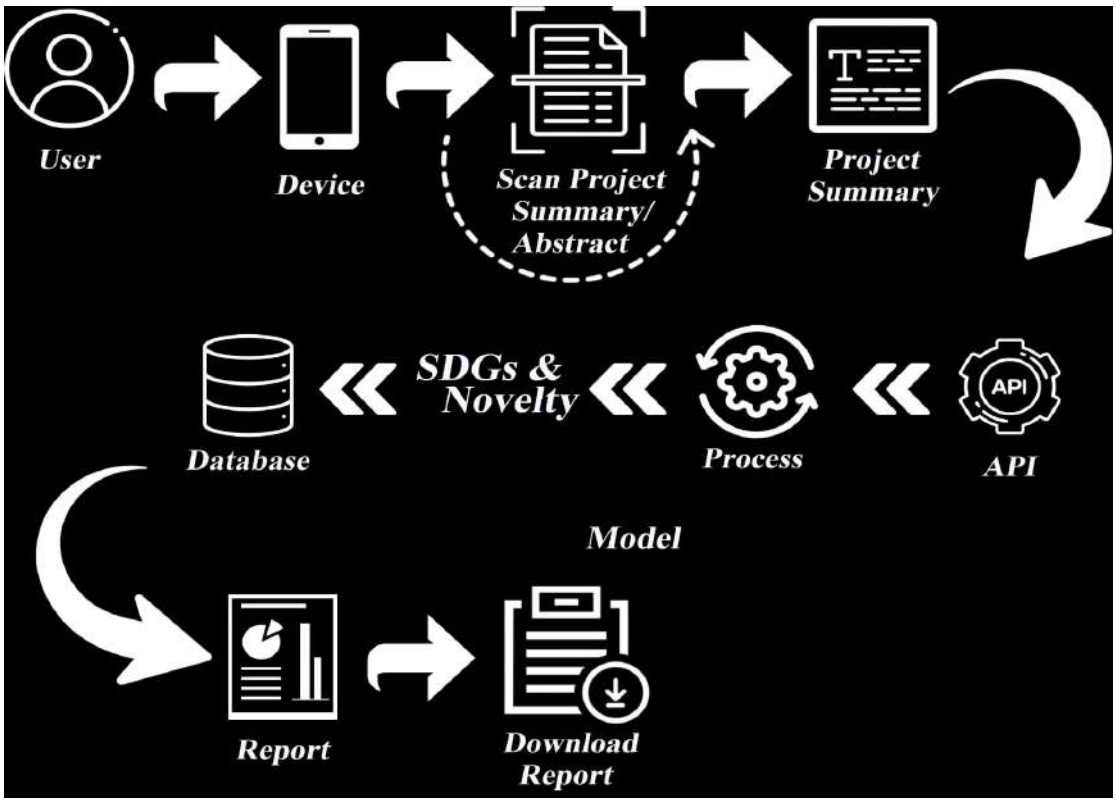

Figure 1: Overall System Mode Diagram

2.2 System Data Flow

The process initiates with the user submitting a project summary or abstract through a designated interface. This abstract is then scanned, preparing the text for analysis. The central node of the flow is the proposed model, which represents the core analytical engine of the system. It applies advanced algorithms to interpret the summary, aligning the project with Sustainable Development Goals and assessing its novelty. Following the analysis, the model generates a comprehensive report that encapsulates the evaluative findings. This report is subsequently made available for the user to download, providing a tangible output that reflects the project's evaluation.

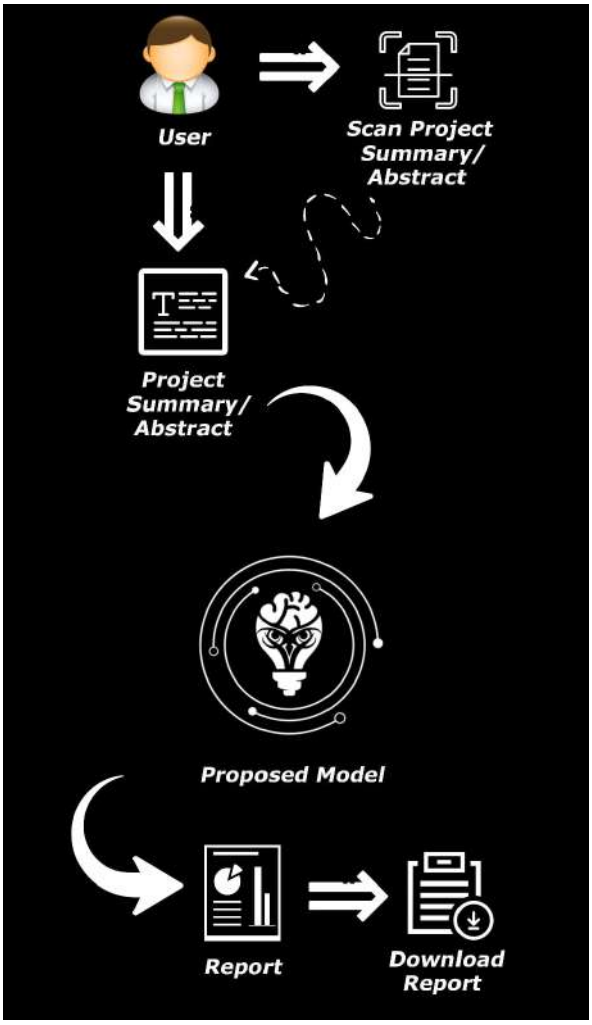

Figure 1: System Data Flow Diagram

Chapter 3

3.1 Literature Review with Comparative Analysis

we systematically examine existing projects and research works in relation to our project's capabilities. Our analysis highlights four key features where our project, "Smarty PIE," distinguishes itself from the current landscape.

Feature	Our Project	Existing Projects	Existing research work
Analyzing the student projects through project summary	✓	✗	✗
Identify which of the SDGs the project aligns with	✓	✗	✓
Checking the project's novelty	✓	✗	✗
Generating a comprehensive report	✓	✗	✗

Table 1: Comparative Analysis

Smarty PIE" excels in analyzing project summaries, an area where existing projects and most research work do not provide functionality. Moreover, our tool adeptly identifies relevant SDGs for each project, a feature uncommon in existing projects and selectively found in research work. The assessment of project novelty is another domain where "Smarty PIE" is pioneering, as both existing projects and research generally lack in providing such assessments. Lastly, our project's ability to generate comprehensive reports is unparalleled in the landscape we reviewed, offering users critical insights that are otherwise unavailable. This comparison underscores the innovative edge and adds value "Smarty PIE" brings to the table. The integration of these unique features into "Smarty PIE" not only fills a significant gap identified in the literature but also enhances the project evaluation process, making it more comprehensive and aligned with international sustainability goals.

Chapter 4

1. 4.1 Proposed Solution/Results & Discussion

Our proposed solution for SDG alignment utilizes the BERT model, enhancing text classification to identify congruence with the Sustainable Development Goals. The sophisticated architecture of BERT facilitates deep contextual understanding, enabling our model to provide accurate mappings even in complex multi-label scenarios. The results of our implementation are promising, as evidenced by the following performance metrics obtained in the final training epoch:

Training Loss	Validation Loss	F1 Score	ROC AUC	Accuracy
2%	2%	98%	98%	95%

These results demonstrate the model's robustness and its potential as a reliable tool for aligning projects with SDGs.

In parallel, we are developing a model for assessing the novelty of project ideas. While we await empirical results, the proposed methodology promises to scrutinize the originality of projects by comparing them against existing works. This component is expected to complement the SDG alignment tool by providing a dual analysis that will ensure projects are not only innovative but also contribute towards global sustainability goals. The combination of these two analytical tools stands to offer a comprehensive solution for project evaluation, setting a new benchmark in both educational and professional settings.

Activity	Description	Optimistic (a)	Most Likely (m)	Pessimistic (b)	Expected (Te)
A	Requirement Analysis	1 week	2 weeks	4 weeks	2 weeks

B	Data Collection & Preprocessing	2 weeks	4 weeks	6 weeks	4 weeks
C	SDG Model Development	2 weeks	5 weeks	8 weeks	5 weeks
D	Novelty Model Development	2 weeks	6 weeks	12 weeks	6 weeks
E	Graphics Work	1 week	2 weeks	3 weeks	2 weeks
F	Testing & QA	1 week	3 weeks	5 weeks	3 weeks
G	Deployment	0.5 weeks	1 week	2 weeks	1 week
H	Post-Deployment Review	0.5 weeks	1 week	1.5 weeks	1 week

Table 2: PERT Activity Time estimate table.

Chapter 5

6.1 Summary

This project, titled "Smarty PIE," is an exploration into automating the assessment of project ideas, primarily focusing on their novelty and alignment with the Sustainable Development Goals (SDGs). The purpose is to provide a systematic approach to evaluating projects, facilitating decision-making for educators, students, and professionals by leveraging advanced computational methods. To address this challenge, the project employs state-of-the-art Natural Language Processing (NLP) and Machine Learning (ML) algorithms. BERT, a transformer-based model, is utilized for its proficiency in text classification tasks, enabling the identification of SDG-relevant features within project summaries. For novelty assessment, the project proposes a unique model that analyzes project abstracts, cross-referencing them with existing databases to ascertain originality. The results from the SDG alignment phase are promising, with the BERT model demonstrating high accuracy and reliability. Although the novelty assessment model is still under development, the expected outcome is to provide a comprehensive measure of a project's uniqueness.

Conclusively, the project advocates for the integration of such automated tools in educational and professional settings to enhance the quality and relevance of project ideas. Recommendations include further refinement of the novelty assessment model and expansion of the database for a more robust comparison.

6.2 Future Work

For future work, our project aims to finalize and rigorously test the novelty assessment model, ensuring its effectiveness across diverse projects. We plan to explore broader applications, potentially adapting the tool for various languages and cultural contexts. Further research will focus on the tool's impact in promoting global sustainability efforts and fostering innovation. We also intend to investigate the integration of additional AI technologies to enhance the tool's capabilities, such as incorporating sentiment analysis for a more nuanced understanding of projects' potential societal impacts.

Chapter 6

Conclusion & Recommendation

In this project, we tackled the challenge of objectively evaluating project ideas, focusing on their novelty and alignment with Sustainable Development Goals. By employing advanced NLP and ML techniques, particularly the BERT model, we developed a tool that automates this evaluative process. The key accomplishment of this project is the successful integration of the BERT model for SDG alignment, demonstrating high accuracy and reliability. While the novelty assessment model is still in development, its potential to revolutionize project evaluation is substantial. Conclusively, this project stands not just as an academic exercise but as a real-world tool with far-reaching implications. It paves the way for more objective, efficient, and globally conscious project evaluations, fostering innovation and sustainability in various fields. The ongoing development of the novelty assessment model and its future applications pose exciting questions for further research and practical implementation.

References

Ayele, W. Y. (2022). A toolbox for idea generation and evaluation: machine learning, data-driven, and contest-driven approaches to support idea generation. *arXiv preprint arXiv:2205.09840*.DOI: https://doi.org/10.48550/arXiv.2205.09840

Banken, V., Ilmer, Q., Seeber, I., & Haeussler, S. (2019). A method for Smart Idea Allocation in crowd-basedideaselection. *DecisionSupportSystems*, *124*,113072.DOI: https://doi.org/10.1016/j.dss.2019.113072

Chikkamath, R., Endres, M., Bayyapu, L., & Hewel, C. (2020, December). An empirical study on patent novelty detection: A novel approach using machine learning and natural language processing. In *2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)* (pp. 1-7). IEEE.DOI: 10.1109/SNAMS52053.2020.9336557

Doboli, S., Kenworthy, J., Paulus, P., Minai, A., & Doboli, A. (2020, July). A cognitive inspired method for assessing novelty of short-text ideas. In *2020 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.DOI: 10.1109/IJCNN48605.2020.9206788

Guisiano, J., & Chiky, R. (2021, January). Automatic classification of multilabel texts related to Sustainable Development Goals (SDGs). In *TECHENV EGC2021*.DOI: https://hal.science/hal-03154261

Efremova, N., West, D., & Zausaev, D. (2019). AI-based evaluation of the SDGs: The case of crop detection with earth observation data. *arXiv preprint arXiv:1907.02813*.DOI: https://doi.org/10.48550/arXiv.1907.02813

Just, J., Ströhle, T., Füller, J., & Hutter, K. (2023). AI-based novelty detection in crowdsourced idea spaces. *Innovation*, 1-28.DOI: https://doi.org/10.1080/14479338.2023.2215740

Ma, N., Politowicz, A., Mazumder, S., Chen, J., Liu, B., Robertson, E., & Grigsby, S. (2021, November). Semantic novelty detection in natural language descriptions.

In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 866-882).DOI: https://aclanthology.org/2021.emnlp-main.66

Mirghaderi, S. H. (2020). Using an artificial neural network for estimating sustainable development goals index. *Management of Environmental Quality: An International Journal*, *31*(4), 1023-1037.DOI: https://doi.org/10.1108/MEQ-12-2019-0266

Peinado, F., Francisco, V., Hervás, R., & Gervás, P. (2010). Assessing the novelty of computer-generated narratives using empirical metrics. *Minds and Machines*, *20*, 565-588.DOI: 10.1007/s11023-010-9209-8

Shin, M., Kim, J., van Opheusden, B., & Griffiths, T. L. (2023). Superhuman artificial intelligence can improve human decision-making by increasing novelty. *Proceedings of the National Academy of Sciences*, *120*(12), e2214840120.DOI: https://doi.org/10.1073/pnas.2214840120

Sukhov, A. (2018). The role of perceived comprehension in idea evaluation. *Creativity and Innovation Management*, *27*(2), 183-195.DOI: 0.1111/caim.12262

Wang, K., Dong, B., & Ma, J. (2019, May). Towards computational assessment of idea novelty. In *Proceedings of the 52nd Hawaii international conference on system sciences*.DOI: http://hdl.handle.net/10125/59531

Wang, C., & Bingsheng Guan, M. (2014). Author's Accepted Manuscript. *J. Memb. Sci. https://doi. org/10.1016/j. memsci*, *55*.DOI: http://dx.doi.org/10.1016/j.ijhcs.2015.01.005

Zhou, G. (2021). Research on text summarization classification based on crowdfunding projects. In *MATEC Web of Conferences* (Vol. 336, p. 06020). EDP Sciences.DOI: https://doi.org/10.1051/matecconf/202133606020

Annexure

Annexure (if any) should be placed at the end of the project report.