Web Based Syllabus to Skill Mapping System Using Machine Learning and Natural Language Processing

Krishna Veni Ampolu[1], Umamaheswararao Mogili[2], P. Kavya[3], G. Roja[4], V. Hema[5], J. Suresh[6], M. Siva[7], Biddika Aneel Kumar[8]

[1,2]*Assistant Professor, Department of Computer Science and Engineering, Avanthi's St Theressa Institute of Engineering and Technology, Garividi, Andhra Pradesh, India*

[8]*Assistant Professor, Department of Humanities & Basic Sciences, Avanthi's St Theressa Institute of Engineering and Technology, Garividi, Andhra Pradesh, India*

[3,4,5,6,7]*B. Tech, Department of Computer Science and Engineering, Avanthi's St Theressa Institute of Engineering and Technology, Garividi, Andhra Pradesh, India*

***Abstract*—The rapid growth of technology has increased the demand for graduates with practical and industry-relevant skills in addition to academic knowledge. However, many students face challenges due to the gap between university curricula and current industry requirements. To address this issue, this research proposes a Web-Based Syllabus to Skill Mapping and Skill Gap Analysis System that uses machine learning and natural language processing techniques to analyze academic syllabus content and identify missing industry skills. The system allows users to upload syllabus documents, which are processed using Natural Language Processing (NLP) methods to extract important topics and keywords. These extracted topics are compared with an industry skill database using TF-IDF vectorization and cosine similarity techniques to measure their relevance. Based on this comparison, the system identifies covered skills, missing skills, and calculates the overall skill coverage percentage. Furthermore, the platform generates skill gap reports and provides recommendations that help students improve their technical competencies. The system is implemented using Django for backend processing, a relational database for data storage, and web technologies for user interaction. By combining curriculum analysis with skill evaluation, the proposed system helps students understand their readiness for industry requirements and supports better career preparation.**

***Index Terms*—Machine Learning, Natural Language Processing (NLP), TF-IDF Vectorization, Cosine Similarity, Curriculum Analysis.**

I. INTRODUCTION

In the modern digital era, technological advancements are rapidly transforming industries across the globe. Organizations are constantly adopting new technologies such as artificial intelligence, data analytics, cloud computing, and automation to improve productivity and innovation. As a result, the demand for skilled professionals who possess relevant technical competencies has increased significantly. Educational institutions play a critical role in preparing students for these evolving industry requirements by designing curricula that incorporate modern technologies and practical skills. However, despite the efforts made by universities and colleges, there is often a mismatch between the skills taught in academic programs and the skills required by industries. Many graduates complete their academic studies with strong theoretical knowledge but lack practical expertise in tools and technologies that are widely used in professional environments. This mismatch creates challenges for both students and employers. Students struggle to secure suitable job opportunities, while companies must spend additional time and resources training newly recruited employees. The concept of skill mapping has emerged as an effective approach to address this issue. Skill mapping involves identifying the skills present in academic curricula and comparing them with the skills required by industry. By analyzing the alignment between syllabus content and industry requirements, educational institutions can evaluate whether their

curriculum adequately prepares students for real-world careers.

The need for skill mapping has become increasingly important due to the dynamic nature of technological development. Modern industries continuously evolve as new programming languages, software frameworks, and digital tools are introduced. Educational institutions must adapt their teaching methods and course structures to ensure that students remain competitive in the job market. In many universities, syllabus updates occur only after several years due to administrative procedures and academic review cycles. During this time, industry requirements may change significantly. As a result, students may graduate with knowledge that is outdated or incomplete in relation to current technological trends Skill mapping helps address this problem by providing a systematic method to analyze the content of academic syllabi and compare it with industry skill requirements. Through this process, institutions can identify missing competencies and update their curriculum accordingly. The primary objective of this project is to develop a web-based system capable of analyzing academic syllabi and identifying skill gaps between educational programs and industry requirements. In the current educational environment, it is essential for students to acquire practical skills that align with industry expectations. Therefore, this project aims to provide an automated solution that can assist students, educators, and institutions in understanding the relationship between syllabus content and the skills demanded by employers. One of the major objectives of the system is to design and develop a web-based platform that allows users to upload syllabus documents for analysis. The system processes the uploaded documents and extracts relevant technical skills and keywords from the syllabus content. These extracted skills are then compared with a predefined database containing industry-required skills. Through this comparison, the system is able to identify missing skills and highlight areas where the syllabus does not fully match current industry requirements.

II. LITERATURE SURVEY

The rapid growth of modern technologies has increased the demand for skilled professionals in areas such as artificial intelligence, data analytics, cloud computing, and cybersecurity. However, many studies have identified a mismatch between academic curricula and industry skill requirements, which is commonly known as the skill gap. This gap creates challenges for graduates seeking employment and for organizations looking for industry-ready professionals. Some of the sample artificial intelligence, machine learning and deep learning models for prediction for skill mapping models are described in details [1-9]. Mikolov et al. [10] introduced efficient methods for learning word representations in vector space. Their work helps machines understand relationships between words in large text datasets and is widely used in Natural Language Processing applications. Salton and Buckley [11] proposed term-weighting approaches for automatic text retrieval systems, which are important for identifying relevant terms in documents. Ramos [12] explained the use of the TF-IDF technique to determine word importance in document collections. TF-IDF is widely used in text mining and document similarity analysis. Bird et al. [13] introduced practical Natural Language Processing techniques using Python, which support text preprocessing tasks such as tokenization, stemming, and stop-word removal. Manning et al. [14] discussed important information retrieval techniques used for indexing and retrieving documents efficiently. Han et al. [15] presented data mining concepts and techniques that help extract useful patterns and knowledge from large datasets.

Joachims [16] applied Support Vector Machines for text categorization and demonstrated that machine learning algorithms can effectively classify textual information. Liu [17] studied sentiment analysis techniques and opinion mining methods used to analyze textual data. Rajaraman and Ullman [18] explored techniques for mining large datasets and discussed algorithms that support scalable data analysis systems. Feldman and Sanger [19] explained text mining techniques that help extract meaningful knowledge from large collections of documents. Sommerville [20] described software engineering principles and structured development methods for designing reliable software systems. Pressman and Maxim [21] also discussed modern software engineering practices used in the development of complex software applications. Gruber [22] proposed ontology specifications that enable knowledge sharing

and semantic understanding in information systems. Porter [23] introduced a stemming algorithm that reduces words to their base form, which improves the performance of text analysis systems. Sebastiani [24] presented a comprehensive survey on machine learning methods used for automated text categorization. Breiman [25] introduced the Random Forest algorithm, which is widely used for classification and prediction tasks due to its accuracy and efficiency. Mitchell [26] provided foundational concepts of machine learning and explained how algorithms learn patterns from data. Jurafsky and Martin [27] discussed speech and language processing techniques used in Natural Language Processing systems. Murphy [28] introduced probabilistic machine learning models that support predictive analytics and intelligent decision-making. Resnick and Varian [29, 30] introduced recommender systems that provide personalized recommendations based on user behaviour and preferences. Based on these research studies, it is evident that combining Natural Language Processing, machine learning, text mining, and recommendation systems can effectively analyze academic syllabi and identify skill gaps between education and industry requirements. These techniques provide a strong foundation for developing intelligentsystems that support curriculum improvement and enhance student employability.

III. METHODOLOGY

The methodology of the proposed system describes the procedures, techniques, and technologies used to develop the Web-Based Syllabus to Skill Mapping and Skill Gap Analysis System. The objective of this methodology is to create an automated platform capable of analyzing syllabus documents, extracting relevant technical skills, comparing them with industry skill requirements, and generating recommendations for skill development. The proposed system follows a structured approach that integrates text processing techniques, skill mapping algorithms, and recommendation models within a web-based architecture. The methodology includes several modules such as syllabus upload and preprocessing, skill extraction, skill database comparison, skill gap analysis, and recommendation generation. Each module is designed to perform specific tasks within the overall system workflow. The integration of these modules enables the system to analyze academic syllabi efficiently and generate meaningful insights regarding curriculum relevance and skill deficiencies. The methodology also emphasizes scalability and accessibility by implementing the system as a web application. This allows users such as students, educators, and training institutions to access the platform through standard web browsers and analyze syllabus documents without requiring complex installations. The methodology of the system is divided into the following stages:

3.1. Data Pre-processing

The collected data is cleaned and prepared for analysis. Missing values are handled, numerical attributes are normalized, and categorical attributes such as genre and artist names are converted into machine-readable formats.

3.2. Feature Extraction

Important song features such as genre, tempo, and energy are converted into feature vectors. These vectors represent the characteristics of each song and help in measuring similarity between songs.

3.3. Content Based Filtering

The system recommends songs that are similar to the ones previously liked or played by the user. Cosine similarity is used to compare song feature vectors and identify similar tracks.

3.4. Collaborative Filtering

User listening patterns are analyzed to find users with similar preferences. Songs liked by similar users are recommended to others with comparable music tastes.

3.5. Hybrid Detection Model

Content-based filtering and collaborative filtering are combined to improve recommendation accuracy.

This hybrid approach balances song similarity with user behaviour patterns. The system ranks songs based on similarity scores and predicted user interests. The top recommended songs are then displayed to the user through the system interface.

IV. RESULTS AND DISCUSSION

The experimental results obtained from the implementation of the Web-Based Syllabus to Skill

Mapping and Skill Gap Analysis System demonstrate the effectiveness of the proposed methodology in analyzing academic syllabus documents and identifying industry-relevant skill gaps. The system was tested using multiple syllabus documents related to computer science and information technology courses. These documents were uploaded through the web interface and processed by the system to extract relevant textual information and identify technical skills Fig 1, 2 & 3.

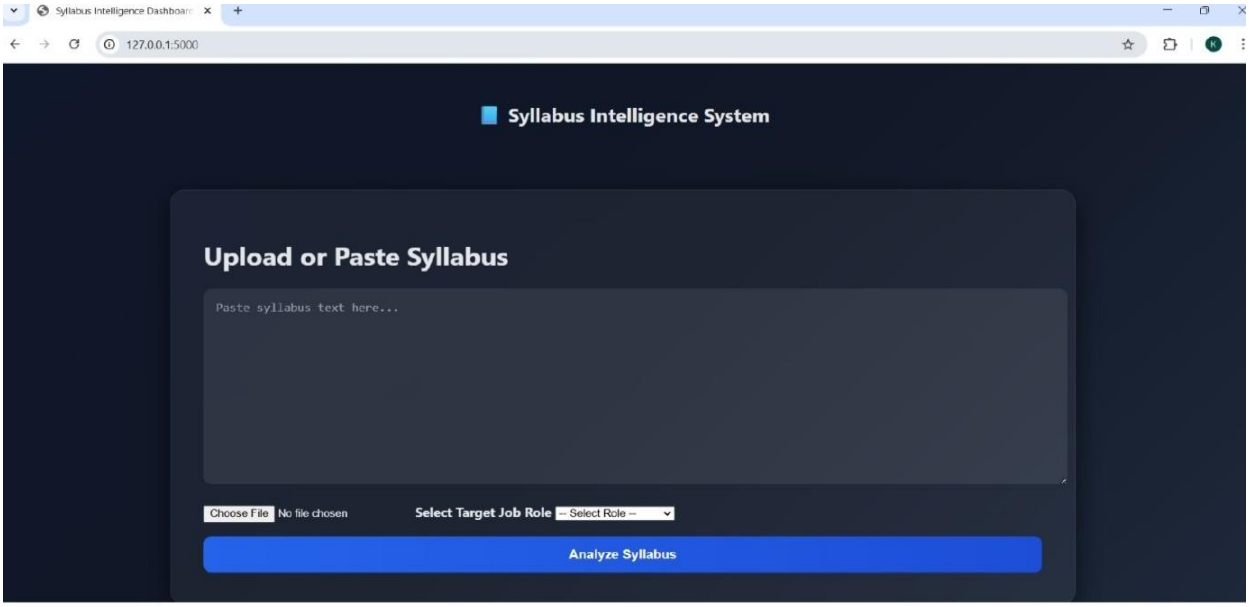

Fig.1. Syllabus intelligence system

The extracted skills were compared with a predefined industry skill dataset, which contains commonly required technical competencies in fields such as data science, software development, artificial intelligence, cloud computing, and database management. By comparing the skills present in the syllabus with those in the industry dataset, the system successfully identified missing skills that are important for professional job roles sown in Fig 4, 5 & 6.

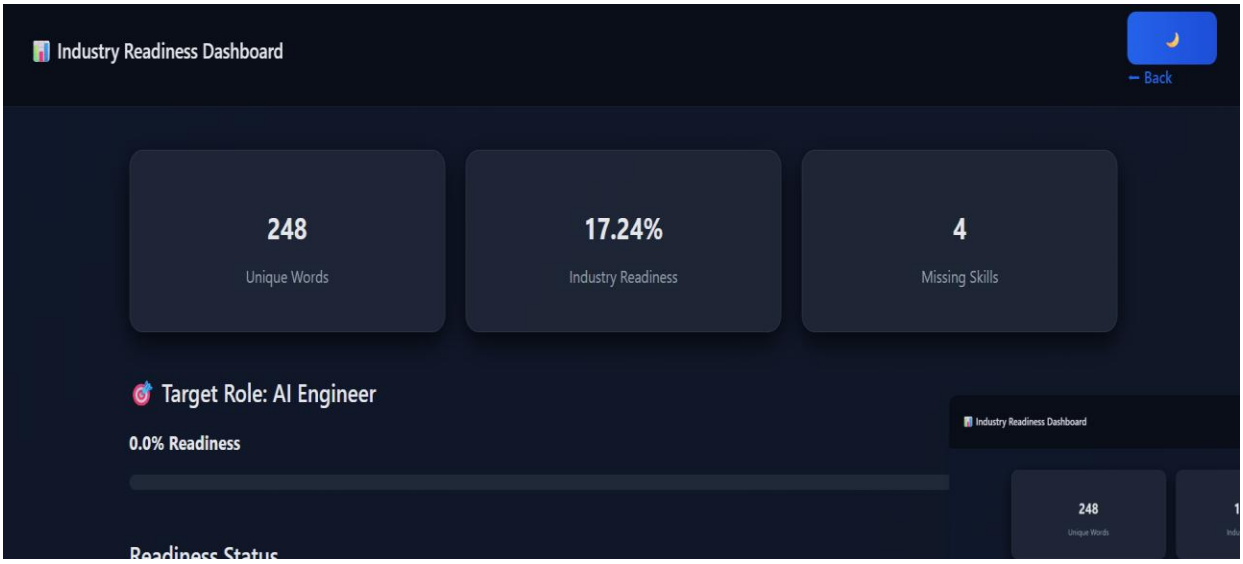

Fig.2. Industry readness dashboard

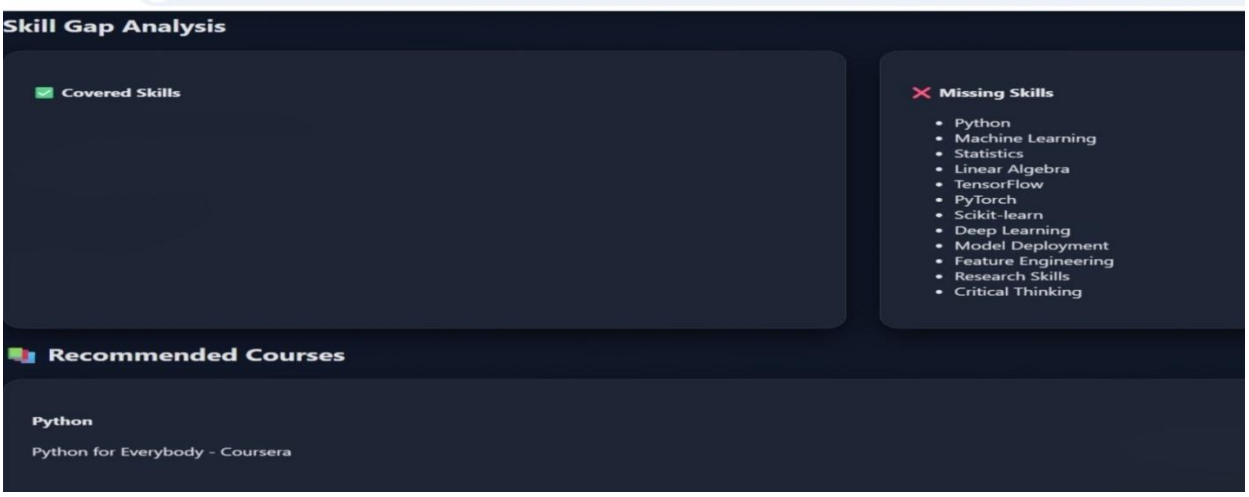

Fig .3. Skill gap analysis

Fig .4. Skill gap Explanation

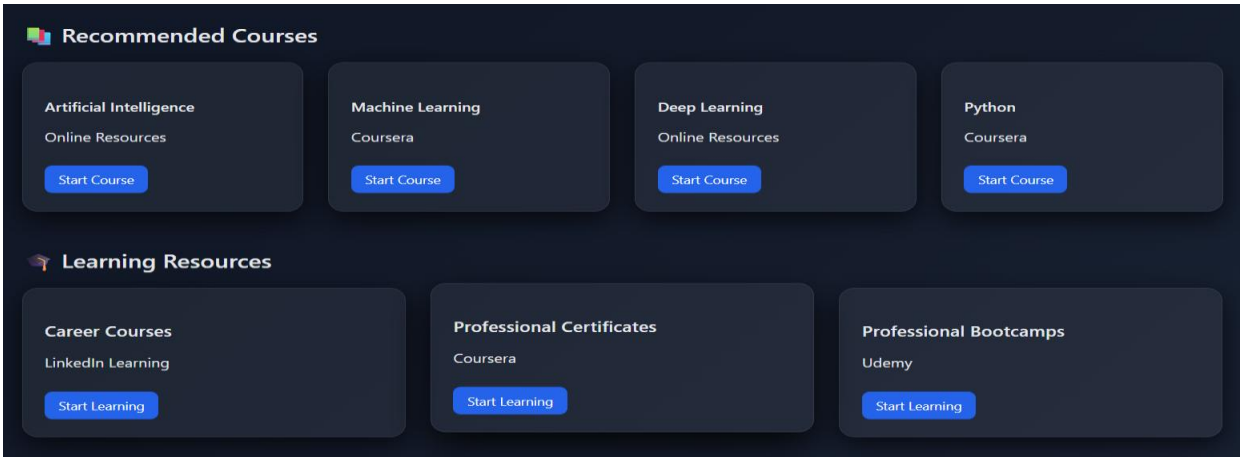

Fig .5. Recommended Courses

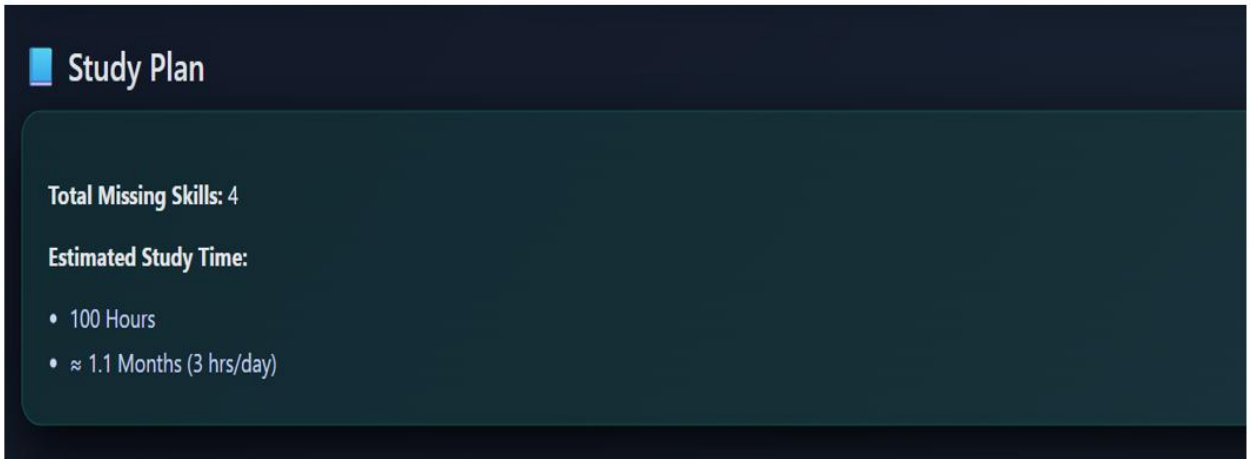

Fig .6. Study plan

V. CONCLUSION

This project presented an AI-Based Web-Based Syllabus to Skill Mapping System designed to evaluate how effectively academic syllabus content aligns with industry skill requirements. The system addresses the common gap between traditional academic curricula and the rapidly evolving demands of modern industries. By analyzing syllabus documents and identifying the skills covered within them, the system helps determine the level of industry readiness provided by educational programs. The proposed system utilizes Natural Language Processing (NLP) techniques to extract relevant skills from syllabus documents and compare them with a predefined database of industry-required skills. Based on this comparison, the system calculates an industry readiness score and identifies both covered and missing skills. This analysis provides valuable insights for students, enabling them to understand their current skill level and recognize areas where additional learning is required. The implementation of the system using Python and the Flask web framework allows the development of an efficient and scalable web-based platform. Through an interactive dashboard, the system presents analysis results using visual representations such as charts and graphs, making the information easy to interpret. Additionally, the system offers career role suggestions and learning recommendations based on the identified skill patterns. Although the current system relies primarily on keyword-based skill extraction, future enhancements can incorporate advanced machine learning and semantic analysis techniques to improve accuracy. Expanding the industry skill database, integrating real-time job market data, and adding features such as resume analysis and course recommendation modules can further strengthen the system. With these improvements, the proposed platform has the potential to significantly support students, educators, and institutions in bridging the gap between academic learning and industry expectations.

Future Scope

The proposed Web-Based Syllabus to Skill Mapping and Skill Gap Analysis System provides an effective solution for analyzing academic syllabi and identifying missing industry-relevant skills. Although the system successfully performs syllabus analysis, skill extraction, skill gap identification, and recommendation generation, there are several opportunities for future improvements that can enhance the system's functionality and applicability. One of the major future enhancements of the system is the integration of advanced machine learning and deep learning techniques. Currently, the system primarily relies on text processing and keyword-based skill extraction methods. In the future, more sophisticated Natural Language Processing (NLP) models such as contextual embeddings and transformer-based models can be incorporated to improve the accuracy of skill detection and semantic understanding of syllabus content. Another important area of future development is the integration with real-time industry datasets and job portals. By connecting the system with job market

platforms and industry skill databases, the platform can dynamically update the list of required skills based on current market trends. This would allow the system to provide more accurate and up-to-date skill gap analysis.

REFERENCES

[1] Mogili, U., Ampolu, K. V., Rajasekharam, B., & Timothy, M. J. AI-Driven Interaction in AR Environments, in Journal of Digital Economy, 2024, Volume 3, Issue 1, pp. 228-234.

[2] Timothy, M. J., Rajasekharam, B., Ampolu, K. V., & Mogili, U. Threat Detection Using AI in Cybersecurity Systems, in IJIS, 2023, Volume 7, Issue 1, pp. 1-7.

[3] Ampolu, K.V., Mogili, U., Timothy, M. J., & Rajasekharam, B. Machine Learning Models for Predictive Maintenance, in IJIS, 2022, Volume 6, Issue 4, pp. 1-7.

[4] Rajasekharam, B., Timothy, M. J., Mogili, U., Ampolu, K.V., Machine Learning Models for Predictive Maintenance, in JDE, 2023, Volume 2, Issue 2, pp. 95-101.

[5] Soujania, B., Ampolu, K. V., Timothy, M. J., & Mogili, U. (2025) Classifying Disease Information Forums through Semantic Similarity-Based Machine Learning, Science, Technology and Development Journal, Volume XIV, Issue II, pp 67-75.

[6] B Satish Kumar, Kavitha C., Mogili, U.R., S. Pallam Shetty (2022). "Application of Machine Learning To Enhance the Performance of The Prophet Routing Protocol For Delay Tolerant Networks". Journal for Basic Sciences, Volume 23, Issue 5, 2107-2116, DOI:10.37896/JBSV23.5/2278.

[7] I. Sree Geeta, Umamaheswararao Mogili. (2022), "Use of Several Machine Learning Algorithms for Effective Prediction of Cyberbullying", International Journal of Creative Research Thoughts, Volume 10, Issue 6, pp 17.

[8] Mogili, U., & Mohamed, A. (2023, November). Artificial intelligence and machine learning in the fields of education, medical, and smart phones. In AIP conference proceedings (Vol. 2917, No. 1, p. 050012). AIP Publishing LLC.

[9] Esram, R., Deepak, B. B. V. L., Mogili, U. R., & Syam Sundar, P. (2022). Agribots concepts and operations—a review. Applications of Computational Methods in Manufacturing and Product Design: Select Proceedings of IPDIMS 2020, 31-40.

[10] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," Proc. International Conference on Learning Representations (ICLR), 2013.

[11] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," Information Processing and Management, vol. 24, no. 5, pp. 513–523, 1988.

[12] J. Ramos, "Using TF-IDF to determine word relevance in document queries," Proc. First Instructional Conference on Machine Learning, pp. 133–142, 2003.

[13] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python, O'Reilly Media, 2009.

[14] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008.

[15] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2012.

[16] T. Joachims, "Text categorization with support vector machines," Proc. European Conference on Machine Learning, pp. 137–142, 1998.

[17] B. Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, 2012.

[18] A. Rajaraman and J. Ullman, Mining of Massive Datasets, Cambridge University Press, 2011.

[19] R. Feldman and J. Sanger, The Text Mining Handbook, Cambridge University Press, 2007.

[20] I. Sommerville, Software Engineering, 10th ed., Pearson Education, 2015.

[21] R. Pressman and B. Maxim, Software Engineering: A Practitioner's Approach, McGraw-Hill, 2019.

[22] T. Gruber, "A translation approach to portable ontology specifications," Knowledge Acquisition, vol. 5, no. 2, pp. 199–220, 1993.

[23] M. Porter, "An algorithm for suffix stripping," Program, vol. 14, no. 3, pp. 130–137, 1980.

[24] F. Sebastiani, "Machine learning in automated text categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.

[25] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[26] T. Mitchell, Machine Learning, McGraw-Hill, 1997.

[27] D. Jurafsky and J. H. Martin, Speech and Language Processing, Pearson, 2020.
[28] K. P. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012.
[29] P. Resnick and H. Varian, "Recommender systems," Communications of the ACM, vol. 40, no. 3, pp. 56–58, 1997.
[30] S.S.D.K. Maha Lakshmi, Umamaheswararao Mogili, Sravya Eluri, Dogga Ramachandra Rao. (2023), "Online Dynamic Out Patient Queue System for Automated Token Generation in Hospitals", Science, Technology and Development Journal, Volume XII, Issue VII, pp 71-78, DOI:23.18001.STD.2023.V12I07.23.37707.