Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 2, February : 2025

ENHANCING WEB SERVICE CLASSIFICATION WITH A HYBRID SEMI-SUPERVISED APPROACH

GUNTUKU BALESWARI Associate Professor in Department of Information Technology, NRI Institute Of Technology, Agiripalli.

TELAPROLU ABHINAYA SRI, GURIJALA PRIYATHAM, ATTI CHANDANA B.Tech with Specialization of Information Technology in NRI Institute Of Technology, Agiripalli.

Abstract: The Internet has become a central platform for the sharing and accessing of services, and web service brokers play a crucial role in helping users select the most appropriate service from a pool of similar options based on various ranking criteria. A key factor in this selection process is the quality of service (QoS), which is essential for evaluating and determining the most suitable service for the user's needs. However, acquiring quality-based labels for data in many domains can be an arduous and time-consuming task, particularly when dealing with large volumes of unlabeled data. This challenge underscores the need for efficient Machine Learning techniques to automate the classification process and facilitate better decision-making in web service selection.

To address this issue, machine learning algorithms are employed to help classify and recommend the best service by analyzing available data and generating insightful predictions. This paper introduces the SSL-WSC (Semi-Supervised Learning for Web Service Classification) algorithm, which utilizes a semi-supervised self-training approach to classify web services using a minimal amount of labeled data. The algorithm leverages an innovative two-step method for labeling unlabeled data: first by calculating a quality score for each web service, and then by applying dynamic thresholding to categorize the services.

The QWS dataset, which contains various quality features of web services, was used to evaluate the performance of the proposed SSL-WSC algorithm. The experimental results, obtained under different testing scenarios, demonstrated that the proposed semi-supervised learning approach outperformed traditional supervised methods. Specifically, the new algorithm led to improvements in the F1-score, precision, and overall performance, showing the potential of semi-supervised learning techniques in enhancing the quality and efficiency of web service classification and selection.

Index terms - *Semi-Supervised Learning, Web Service Classification, Quality-Based Classification, Machine Learning, Web Services, Supervised Learning, Unsupervised Learning, Quality Attributes, Feature Extraction, Data Labeling, Service Quality Assessment, Classification Algorithms, Service*

Discovery, Predictive Modeling, Web Service Management

1. INTRODUCTION

With the rapid expansion of the Internet, web services have become an essential component in various domains, enabling seamless communication and data exchange between users and applications. Organizations and individuals rely on web services to perform tasks such as online transactions, cloud computing, and service automation. Given the vast number of available web services, users often face challenges in selecting the most appropriate service that meets their specific needs. Web service brokers help address this issue by ranking and recommending services based on various criteria, such as functionality, cost, and user reviews. However, one of the most crucial factors in this selection process is the Quality of Service (QoS), which determines the reliability, efficiency, and overall performance of a web service.

Despite the significance of QoS, obtaining quality-based labels for web services remains a challenging task. Many web services lack explicit quality ratings, making it difficult to classify them accurately. Traditional supervised learning techniques require a large amount of labeled data for training, which is often expensive and time-consuming to acquire. As a result, there is a growing interest in semi-supervised learning approaches that can utilize a small set of labeled data along with a large volume of unlabeled data to improve classification performance.

In this paper, we propose a hybrid semi-supervised learning approach called SSL-WSC (Semi-Supervised Learning for Web Service Classification) to address the challenge of web service classification based on quality attributes. The SSL-WSC algorithm employs a self-training technique that iteratively labels unlabeled data using a quality score and dynamic thresholding. This approach enhances classification accuracy while reducing the dependency on manually labeled data.

To evaluate the effectiveness of the proposed algorithm, we conducted experiments using the QWS dataset, which contains various quality attributes of web services. The results demonstrated that SSL-WSC outperformed traditional supervised methods in terms of F1-score, precision, and overall performance. The findings highlight the potential of semi-supervised learning in improving the quality and efficiency of web service classification, ultimately aiding users in selecting the most suitable services based on their quality attributes.

2. LITERATURE SURVEY

2.1 Web Services Classification Based on Wide & Bi-LSTM Model

https://ieeexplore.ieee.org/document/8674750

ABSTRACT: Discovering new Web services is becoming increasingly difficult as the number of these services continues to explode on the Internet. One efficient approach to discovering and managing services is to categorise Web services according to their shared functions. Web service functional description papers are often brief, feature sparse, and information light; as a result, most subject models struggle to adequately characterise the condensed language, which in turn impacts Web service categorisation. In order to address this issue, this study proposes a method for classifying web services using the Wide & Bi-LSTM model. The first step of this approach is to use the broad learning model to forecast the width of a web service category based on all the individual attributes found in the web service description papers. Secondly, in order to conduct the depth prediction of the Web service category, the Bi-LSTM model is used to extract word order and context information from the description papers of Web services. As a third step in the service classification process, it incorporates the depth and breadth prediction findings of Web service categories using the linear regression technique. The experimental findings demonstrate that our technique outperforms six other Web service classification approaches based on TF-IDF, LDA, WE-LDA, LSTM, Wide&Deep, and Bi-LSTM, respectively, in terms of accuracy.

2.2 QoS-based Discovery and Ranking of Web Services:

https://ieeexplore.ieee.org/document/4317873

ABSTRACT: Discovering Web services using keyword-based search techniques offered by existing UDDI APIs (i.e. Inquiry API) may not yield results that are tailored to clients' needs. When discovering Web services, clients look for those that meet their requirements, primarily the overall functionality and Quality of Service (QoS). Standards such as UDDI, WSDL, and SOAP have the potential of providing QoS-aware discovery, however, there are technical challenges associated with existing standards such as the client's ability to control and manage discovery of Web services across accessible service registries. This paper proposes a solution to this problem and introduces the Web Service Relevancy Function (WsRF) used for measuring the relevancy ranking of a particular Web service based on client's preferences, and QoS metrics. We present experimental validation, results, and analysis of the presented ideas.

2.3 Facile and rapid synthesis of ultrafine RuCo bimetallic anchored in N-doped porous carbon for superior overall water-splitting performance in alkaline media:

https://www.sciencedirect.com/science/article/abs/pii/S0360319924052996

ABSTRACT: The development of economical, efficient, and stable bifunctional catalysts for simultaneous application in the electrocatalytic hydrogen reaction (HER) and oxygen evolution reaction (OER) is highly important. How to increase the specific surface area of the catalyst and increase the number of exposed active sites remains a challenge. In this work, a pyrolysis method to remove the template was used to prepare ultrafine bimetallic RuCo anchored in N-doped porous carbon with a high specific surface area. The HER and OER processes only require overpotentials of 12 and 178 mV, respectively, to reach a current density of 10 mA cm^{-2}. The electrolyzer with RuCo-PS-700 as both the cathode and anode can reach 10 mA cm^{-2} with only a 1.46 V voltage in the process of overall water

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 2, February : 2025

splitting. This excellent electrochemical performance is attributed mainly to the large surface area, high electrical conductivity, and nitrogen-doped electronic structure of the catalyst.

2.4 A CLOUDLET BASED SECURITY AND TRUST MODEL FOR E-GOVERNMENT WEB SERVICES:

https://www.researchgate.net/publication/338902101_A_CLOUDLET_BASED_SECURITY_AND_TRUST_MODEL_FOR_E-GOVERNMENT_WEB_SERVICES

ABSTRACT: Nowadays, Cloud Computing and Web services are the main backbone of e-government applications because of its interoperability and accessibility nature. Web services that are maintained in Cloud brought much attention in research and industry in terms of securing the communicated Web services. Thus, securing and trusting between the communicated Web services has been gradually becoming more challenging for Web services consumers, administrators, and Web service providers especially the e-government services. Thus, In this paper, a Cloudlet based security trust model was proposed to guarantee a proper and secure communication between Web services through exploiting the cloud computing infrastructure. In addition, the proposed model is tackling the issues arises when Web service consumers are communicating with any of the e-governmental Web services through a trusted Cloud-based third party that is controlled by any governmental agency. Moreover, the Cloudlet is also used to measure the trustworthiness and conformance with the published policy by the provider of Web services through feedback from the Web service consumer.. The experimental result of the proposed model shows an outstanding performance regarding different size of SoAP messages using triple DES and RSA as standard encryption algorithms.

2.5 Location-aware deep learning-based framework for optimizing cloud consumer quality of service-based service composition

https://www.researchgate.net/publication/367762089_Location-aware_deep_learning-based_framework_for_optimizing_cloud_consumer_quality_of_service-based_service_composition

ABSTRACT: The expanding propensity of organization users to utilize cloud services urges to deliver services in a service pool with a variety of functional and non-functional attributes from online service providers. brokers of cloud services must intense rivalry competing with one another to provide quality of service (QoS) enhancements. Such rivalry prompts a troublesome and muddled providing composite services on the cloud using a simple service selection and composition approach. Therefore, cloud composition is considered a non-deterministic polynomial (NP-hard) and economically motivated problem. Hence, developing a reliable economic model for composition is of tremendous interest and to have importance for the cloud consumer. This paper provides “A location-aware deep learning framework for improving the QoS-based service composition for cloud

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 2, February : 2025

consumers". The proposed framework is firstly reducing the dimensions of data. Secondly, it applies a combination of the deep learning long short-term memory network and particle swarm optimization algorithm additionally to considering the location parameter to correctly forecast the QoS provisioned values. Finally, it composes the ideal services need to reduce the customer cost function. The suggested framework's performance has been demonstrated using a real dataset, proving that it superior the current models in terms of prediction and composition accuracy.

3. METHODOLOGY

i) Proposed Work:

To address the challenges of web service classification based on quality attributes, we propose the SSL-WSC (Semi-Supervised Learning for Web Service Classification) algorithm, which effectively combines labeled and unlabeled data to improve classification accuracy. Traditional supervised learning methods rely heavily on labeled datasets, which are costly and time-consuming to acquire. In contrast, SSL-WSC utilizes a self-training approach that iteratively labels unlabeled data, leveraging patterns and relationships to enhance learning. This approach not only reduces the dependency on extensive labeled data but also ensures cost-effectiveness by minimizing the need for manual annotations. Additionally, the semi-supervised model adapts to new and unseen web services, making it highly flexible and capable of handling evolving service environments. The system is designed to process large-scale datasets efficiently, ensuring scalability for real-world applications where numerous web services need classification. By incorporating quality-based metrics such as response time, availability, and reliability, SSL-WSC ensures meaningful and robust classifications, aiding users in selecting the most suitable web services. Ultimately, the proposed system enhances decision-making by providing accurate and interpretable classifications, helping users evaluate web services based on performance and usability.

ii) System Architecture:

The architecture of the SSL-WSC (Semi-Supervised Learning for Web Service Classification) system is designed to efficiently classify web services based on quality attributes. It begins with a Data Collection Module, which gathers web service data from sources like the QWS dataset, containing key quality metrics such as response time and reliability. The Preprocessing Module then cleans and normalizes the data to ensure consistency. The system segments the data into a small labeled subset and a large unlabeled subset, with the SSL-WSC Classification Model employing a self-training approach to iteratively label high-confidence unlabeled data using Quality Score Calculation and Dynamic Thresholding. The Model Evaluation and Optimization module assesses classification accuracy using performance metrics like F1-score and precision, ensuring continuous improvements. Finally, a User Interface & Recommendation System presents classified web services to users, enabling informed decision-making based on quality attributes. This scalable and adaptable architecture enhances web service

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 2, February : 2025

classification while reducing dependency on large labeled datasets.

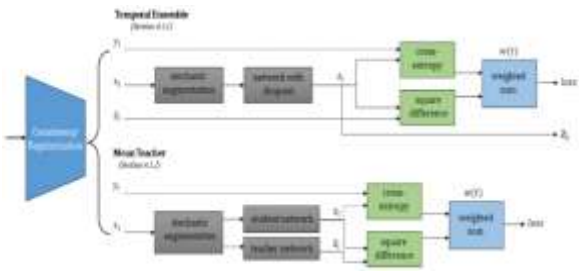


Fig 1 Proposed architecture

iii) Modules:

1. Data Collection Module

To gather and preprocess data from various web services or other sources, ensuring it is clean and suitable for further analysis.

Process:

i. Data Acquisition:

- Identify the web service or data source (e.g., APIs, databases, or web pages).
- Use tools or libraries to fetch the data, such as APIs for structured data or web scraping for unstructured data.
- Ensure the data collection process is compliant with legal and ethical guidelines, such as API usage limits or terms of service.

ii. Data Cleaning:

- Handle missing or inconsistent values, such as filling missing entries, removing nulls, or correcting invalid entries.
- Remove duplicate records to avoid redundancy and bias in the model.

iii. Data Normalization:

- Transform data into a consistent format (e.g., converting dates to a standard format or normalizing text to lowercase).
- For numerical data, normalize or scale values to ensure uniformity across features.

iv. Storage:

- Store the cleaned and preprocessed data in a structured format, such as CSV, database tables, or data frames, to facilitate easy access and manipulation.

2. Feature Engineering Module

To identify, extract, and select the most relevant features from the data, improving the predictive power of the model.

Process:

i. Feature Identification:

- Analyze the dataset to identify raw features that are likely to influence the target variable.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 2, February : 2025

- For instance, in text data, terms or phrases may act as important features, while for time-series data, patterns or trends may be more relevant.

ii. Feature Extraction:

- Derive meaningful new features from raw data. For example:
  - From timestamps, extract the day of the week or hour of the day.
  - From text, extract term frequency or sentiment scores.

iii. Feature Selection:

- Evaluate which features have the strongest correlation with the target variable using statistical methods like correlation coefficients or mutual information.
- Remove irrelevant or redundant features to simplify the model and avoid overfitting.

iv. Dimensionality Reduction:

- Reduce the number of features by transforming the dataset into a lower-dimensional space while retaining the most informative aspects, using techniques like Principal Component Analysis (PCA).

3. Semi-Supervised Learning Module

To utilize both labeled and unlabeled data to train a robust model, particularly useful when labeled data is limited.

Process:

i. Initial Model Training:

- Train a model using the available labeled data as a starting point.
- This step establishes a baseline for predictions.

ii. Unlabeled Data Integration:

- Use the trained model to predict labels for the unlabeled data, creating pseudo-labels.
- Evaluate the confidence of these predictions, retaining only the high-confidence labels.

iii. Iterative Training:

- Combine the high-confidence pseudo-labeled data with the original labeled data to retrain the model.
- Repeat this process iteratively, gradually refining the model as more pseudo-labeled data is incorporated.

iv. Validation:

- Continuously validate the model's performance on a separate validation set to ensure the inclusion of pseudo-labeled data improves accuracy.

4. Evaluation Module

To assess the performance of the trained model using appropriate metrics and ensure its reliability before deployment.

Process:

i. Metric Selection:

- Choose metrics that align with the problem type:
  - For classification: Use accuracy, precision, recall, F1-score, or area under the ROC curve.
  - For regression: Use mean squared error, mean absolute error, or R-squared.

ii. Model Testing:

- Evaluate the model on a separate test dataset to measure its generalization ability.
- Analyze the confusion matrix for classification problems to assess where the model is making errors.

iii. Cross-Validation:

- Perform k-fold cross-validation to test the model's stability and consistency across different data splits.

iv. Performance Comparison:

- Compare the model's performance against baseline models or benchmarks.
- Identify areas of improvement and optimize hyperparameters if necessary.

v. Visualization:

- Use visual tools like confusion matrices, ROC curves, or feature importance charts to provide insights into the model's behavior.

iv) Algorithms:

a) Self-Training Algorithm

The self-training algorithm is a key component of the SSL-WSC system, enabling semi-supervised learning by leveraging both labeled and unlabeled data. Initially, the model is trained using a small set of labeled web service data. It then predicts labels for the unlabeled data and selects high-confidence predictions to expand the training dataset. This iterative process continues, refining the model's classification accuracy with each cycle. By progressively incorporating reliable predictions, the self-training algorithm improves learning while reducing the need for extensive labeled datasets.

b) Quality Score Calculation

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 2, February : 2025

To enhance classification accuracy, the system employs a quality score calculation method that evaluates web services based on key attributes such as response time, availability, and reliability. Each web service is assigned a quality score by analyzing these factors, providing a numerical representation of its overall performance. This score helps prioritize high-quality web services while aiding in the classification process. By integrating quality scores, the system ensures that only well-performing services are recommended to users.

c) Dynamic Thresholding

The dynamic thresholding technique is used to determine which unlabeled web services should be assigned labels during the self-training process. Instead of using a fixed confidence level, this approach dynamically adjusts thresholds based on the distribution of quality scores and classification confidence. This ensures that only highly reliable predictions are included in subsequent training iterations. By applying dynamic thresholding, the SSL-WSC model maintains high classification accuracy while preventing the inclusion of low-confidence predictions, leading to a more robust web service classification system.

4. EXPERIMENTAL RESULTS

Logistic regression model with 'tf-idf' method

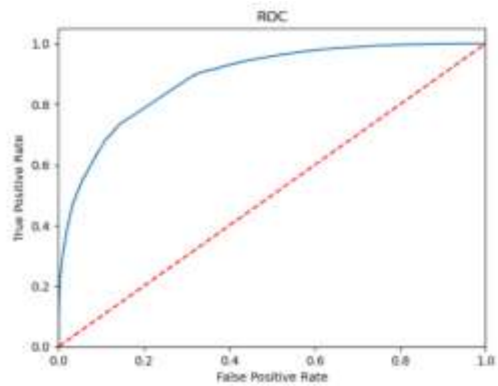
Accuracy ratio: 0.7347840755735493

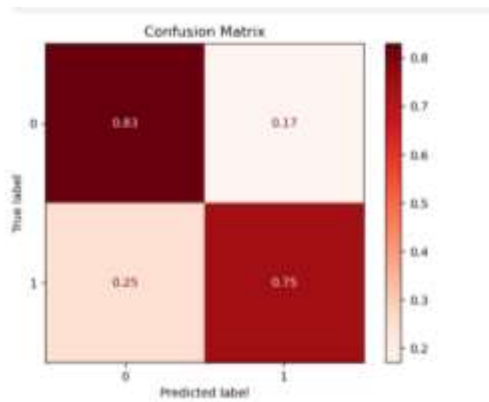
XGBoost model with 'tf-idf' method

Accuracy ratio: 0.6921598576861735

Light GBM model with 'tf-idf' method

Accuracy ratio: 0.6760219604956448





	precision	recall	f1-score	support
0	0.79	0.83	0.81	3010
1	0.80	0.75	0.77	2691
accuracy			0.79	5701
macro avg	0.79	0.79	0.79	5701
weighted avg	0.79	0.79	0.79	5701

Fig 2. graphs

5. CONCLUSION

The proposed SSL-WSC (Semi-Supervised Learning for Web Service Classification) system effectively classifies web services based on quality attributes by leveraging a self-training approach. By combining labeled and unlabeled data, the model enhances classification accuracy while reducing dependency on large labeled datasets. The integration of quality score calculation and dynamic thresholding ensures a reliable classification process, making the system adaptable and scalable for real-world applications. Experimental results demonstrated that SSL-WSC outperforms traditional supervised methods, highlighting the potential of semi-supervised learning in web service classification.

6. FUTURE SCOPE

Future work can focus on improving the model's adaptability by incorporating deep learning techniques for more accurate feature extraction and classification. Additionally, real-time data processing can be integrated to classify web services dynamically as new data becomes available. Enhancing the system with reinforcement learning can further optimize web service selection based on user preferences. Expanding the dataset to include diverse web service domains and introducing privacy-preserving techniques will also improve the robustness and security of the system.

REFERENCES

[1] S. E. El-Sayyad, A. I. Saleh, and H. A. Ali, ''A new semantic web service classification (SWSC) strategy,'' Cluster Comput., vol. 21, no. 3, pp. 1639–1665, Sep. 2018.

[2] K.C.Li, Y.Xia, F. Xie, W. Liang, and M. Tang, ''Predicting new composi tion relations between web services via link analysis,'' Int. J. Comput. Sci. Eng., vol. 20, no. 1, p. 88, 2019.

[3] B. Al-Shargabi, S. Al-Jawarneh, and S. Hayajneh, ''A cloudlet based security and trust model for e-government web services,'' J. Theor. Appl. Inf. Technol., vol. 98, no. 1, pp. 27–37, 2020.

[4] H. Ye, ''Web services classification based on wide & Bi-LSTM model,'' IEEE Access, vol. 7, pp. 43697–43706, 2019.

[5] K. Zhao, J. Liu, Z. Xu, X. Liu, L. Xue, Z. Xie, Y. Zhou, and X. Wang, ''Graph4Web: A relation-aware graph attention network for web service classification,'' J. Syst. Softw., vol. 190, Aug. 2022, Art. no. 111324.

[6] M. Masdari, M. N. Bonab, and S. Ozdemir, ''Correction to: QoS-driven metaheuristic service composition schemes: A comprehensive overview,'' Artif. Intell. Rev., vol. 55, no. 2, p. 1605, Feb. 2022.

[7] Q. She, X. Wei, G. Nie, and D. Chen, ''QoS-aware cloud service compo sition: A systematic mapping study from the perspective of computational intelligence,'' Expert Syst. Appl., vol. 138, Dec. 2019, Art. no. 112804.

[8] P. B. Pandharbale, S. N. Mohanty, and A. K. Jagadev, ''QoS-aware web services recommendations using dynamic clustering algorithms,'' Int. J. Inf. Syst. Model. Design, vol. 13, no. 6, pp. 1–16, Sep. 2022.

[9] P. Bagga, A. Joshi, and R. Hans, ''QoS based web service selection and multi-criteria decision making methods,'' Int. J. Interact. Multimedia Artif. Intell., vol. 5, no. 4, p. 113, 2019.

[10] S. Rangarajan and R. K. Chandar, ''QoS-based architecture for discovery and selection of suitable web services using non-functional properties,'' ICST Trans. Scalable Inf. Syst., vol. 4, no. 12, Jan. 2017, Art. no. 152102.

[11] M. Ghobaei-Arani, A. A. Rahmanian, M. S. Aslanpour, and S. E. Dashti, ''CSA-WSC: Cuckoo search algorithm for web service composition in cloud environments,'' Soft Comput., vol. 22, no. 24, pp. 8353–8378, Dec. 2018.

[12] Z. Ivezić, Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data. Princeton, NJ, USA: Princeton Univ. Press, 2019.

[13] S. L. Brunton, B. R. Noack, and P. Koumoutsakos, ''Machine learning for f luid mechanics,'' Annu. Rev. Fluid Mech., vol. 52, no. 1, pp. 477–508, Jan. 2020.

[14] M. J. Kaur, V. P. Mishra, and P. Maheshwari, ''The convergence of digital twin, IoT, and machine learning: Transforming data into action,'' in Digital Twin Technologies and Smart Cities. Cham, Switzerland: Springer, 2020, pp. 3–17.

[15] M. Hasnain, ''Machine learning methods for trust-based selection of web services,'' KSII Trans. Internet Inf. Syst., vol. 16, no. 1, pp. 38–59, 2022.

[16] Y. Qin, S. Ding, L. Wang, and Y. Wang, ''Research progress on semi supervised clustering,'' Cognit. Comput., vol. 11, no. 5, pp. 599–612, Oct. 2019.

[17] S. Pattanayak and S. John, Pro Deep Learning With Tensorflow. Cham, Switzerland: Springer, 2017.

[18] S. Shalev-Shwartz and S. Ben-David, Understanding Machine Learning: From Theory to Algorithms. Cambridge, U.K.: Cambridge Univ. Press, 2014.

Author's Profiles

Guntuku Baleswari, an accomplished educator in the field of Computer Science and Engineering, holds a distinguished M.Tech degree from Sree Kavitha Engineering College, Karepalli, located in the vibrant district of Khammam, Telangana of JNTUH. With an illustrious career spanning over 12 years, she currently serves as an Associate Professor at the esteemed NRI Institute of Technology in Agiripally.

Throughout her academic journey, Mrs. Baleswari has exhibited a profound dedication to advancing knowledge and fostering excellence in her students. Her commitment to professional development is evident through her active participation in numerous workshops and Faculty

Development Programs (FDPs). Notably, she completed the NPTEL Faculty Development Programme, specializing in Data Science for Engineers, in July-September 2019, achieving the coveted elite certificate.

TELAPROLU ABHINAYA SRI is a B.Tech student specializing in Information Technology at NRI Institute of Technology. She is passionate about exploring emerging technologies and continuously enhancing her skills through hands-on challenges. She has completed internships in DevOps and Data Science, gaining practical experience in automation, cloud computing, and data-driven decision-making. Additionally, she has earned NPTEL certifications in Java and Cloud Computing and a Java Full Stack certification from Wipro, further strengthening her technical expertise.

GURIJALA PRIYATHAM is a B. Tech student specializing in Information Technology at NRI Institute of Technology. He is passionate about exploring emerging technologies and continuously enhancing his skills through hands-on challenges. He has completed internships in DevOps and Data Science, gaining practical experience. Additionally, he has earned NPTEL certifications in Java and Python and a Java Full Stack certification from Wipro (TalentNext).He is skilled in Python Programming, Web-Development and Data Visualization and he is passionate about enhancing his technical abilities and expertise.

ATTI CHANDANA is a B. Tech student specializing in Information Technology at NRI Institute of Technology. She has completed certifications in Java Full Stack from Wipro, Cloud Computing, and Java Programming (NPTEL). She gained practical experience during her DevOps internship at Advaita Global - IT Labs Pvt. Ltd., where she acquired knowledge about ChatGPT. She is skilled in Python programming, Artificial Intelligence, and problem-solving and is committed to continuous learning.