Project Title: Detection of Pipeline Inspection Gauge (PIG) Using Machine Learning and Acoustic Analysis

PRATIKSHYA TIWARI, Graduate Student
Department of Mechanical Engineering, Cullen College of Engineering

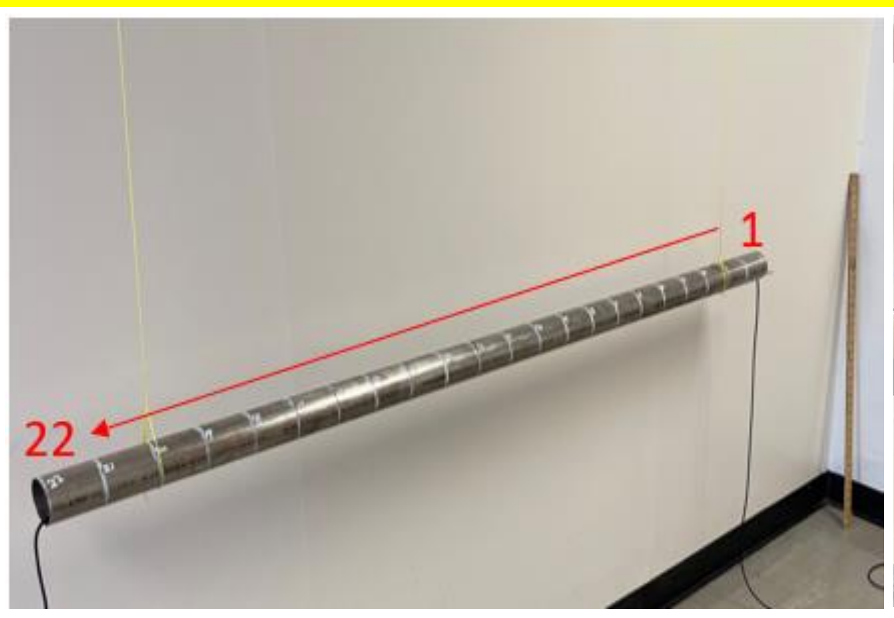
Problem Statement

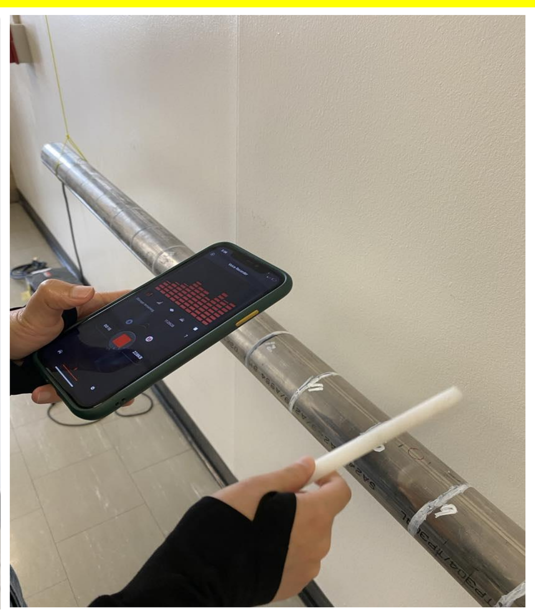
In the pipeline industry, pinpointing the location of a "pig" without accessing the pipe is challenging. This project aims to enhance this process by utilizing sound-based machine learning algorithms for automated, non-intrusive detection. This method is expected to be faster and more accurate than manual detection method, revolutionizing inspection practices, improving efficiency and reducing costs.

Brief Literature Review

Recent advancements in acoustic signal processing highlight the potential of machine learning to improve pipeline monitoring. Ensemble methods, combining outputs from multiple models, have shown superior performance. Zhou et al. demonstrated this approach's effectiveness over traditional methods [1], emphasizing the utility of complex model frameworks. Data imbalance remains a significant challenge in such systems. A study [2] addressed class imbalance in educational data by evaluating techniques like oversampling and undersampling, finding that SMOTEENN with ensemble classifiers performed well, suggesting its applicability in other areas like pipeline inspection.

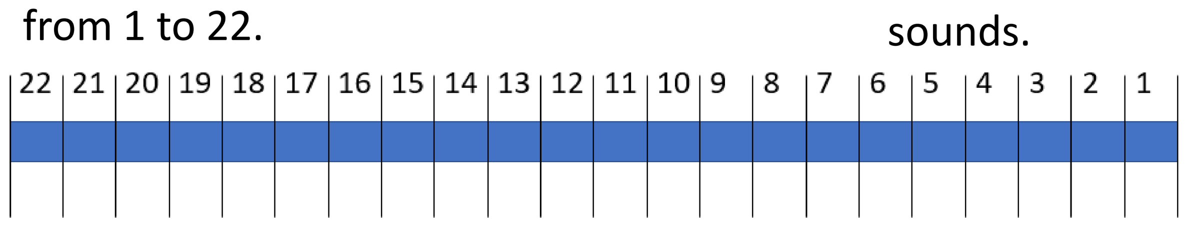
Experimental Setup and Collection of Data





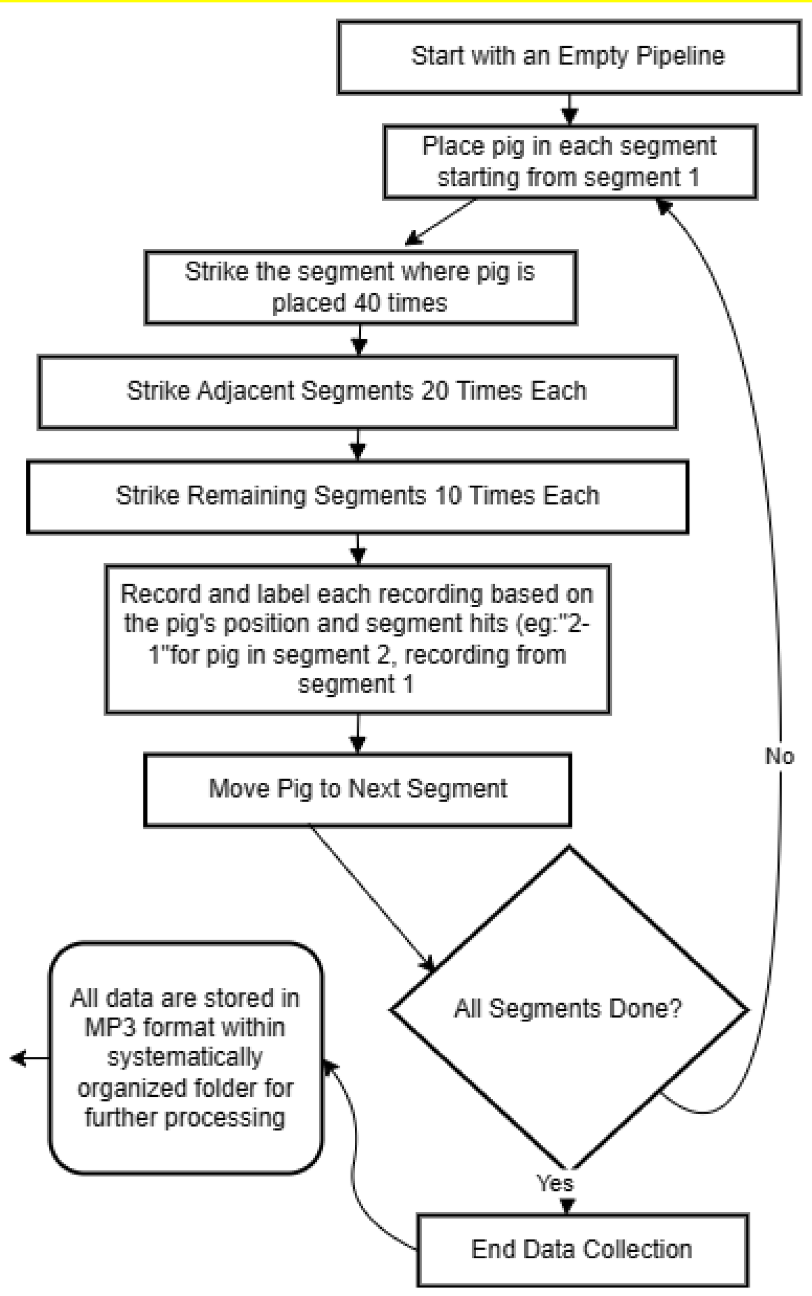
Pipeline Configuration: Horizontally suspended pipeline divided into 22 segments. Each segment in figure below is 3 inches apart and labeled from 1 to 22.

Equipment Used: iPhone, which records the sounds, and a stick to strike the pipeline to generate sounds.



4-4
4-5
4-6
4-7
4-8
4-9
4-10
4-11

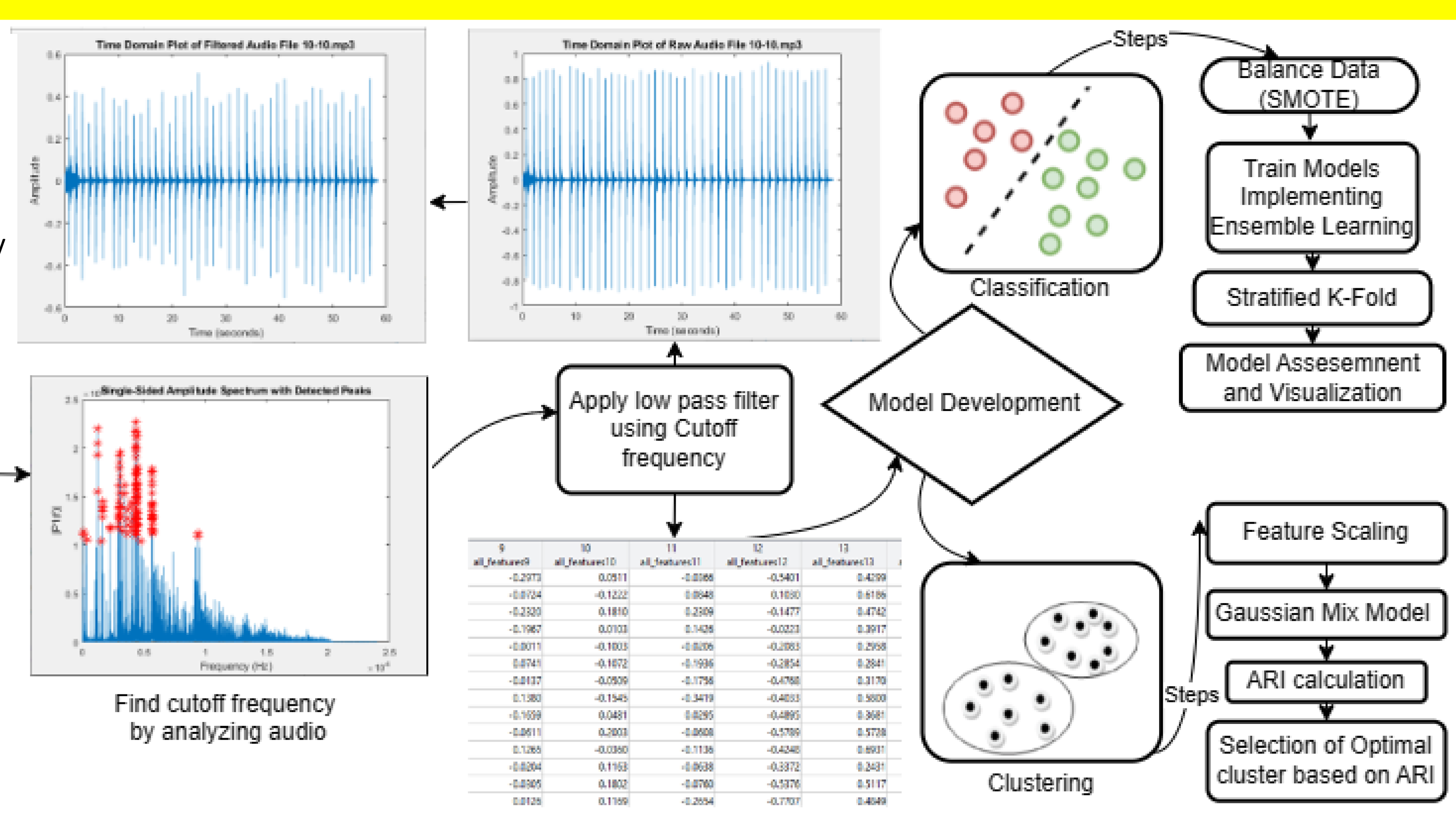
General Naming Guidelines: When the pig is present in a segment, recordings are labeled "Pig Segment-Recording Segment," indicating both the pig's location and the recording source.The convention simplifies distinguishing between pig-present and pig-absent recordings based on segment numbers.



Method(s)

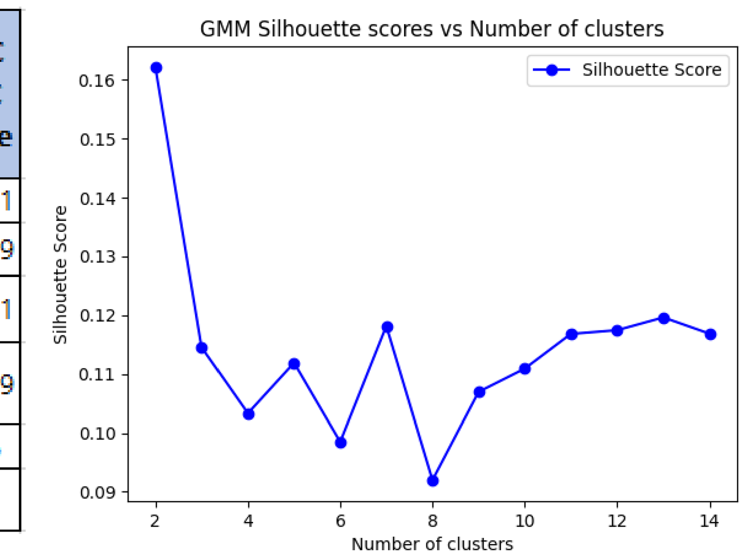
Signal Processing:
Developed a MATLAB script for acoustic signal Analysis. Identified significant frequency peaks in audio and apply low pass filter with tailored cutoff frequency for noise reduction.

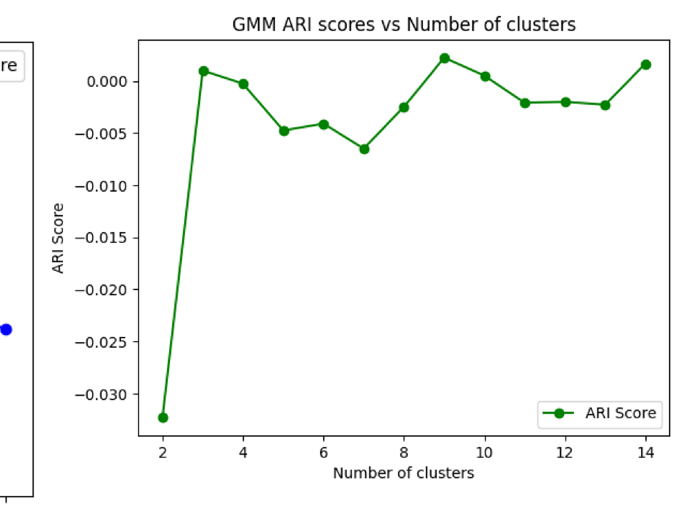
Creative Aspects:
Ensemble learning with soft voting, handling of imbalanced data with SMOTE and adaptive cutoff frequency calculations.

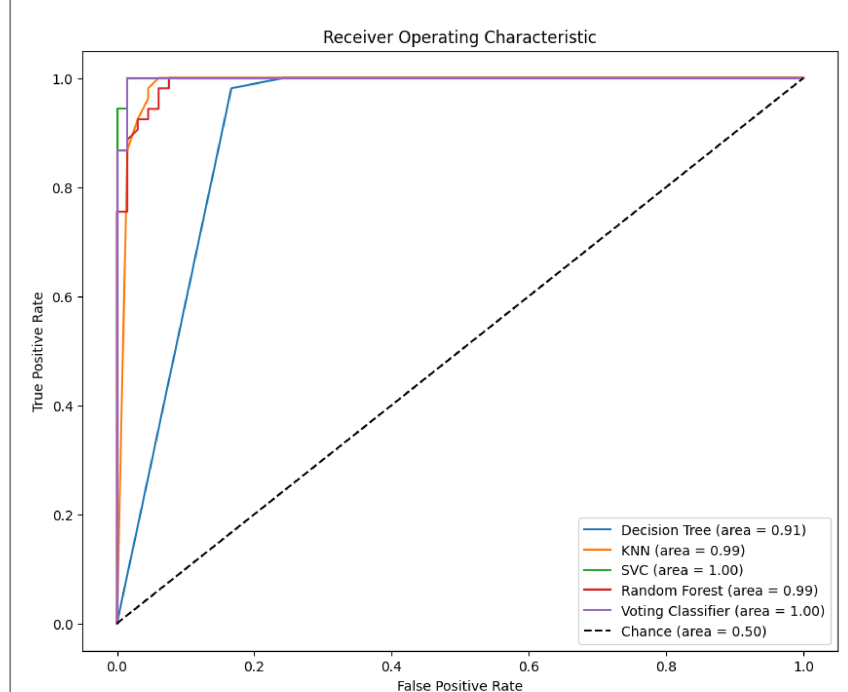


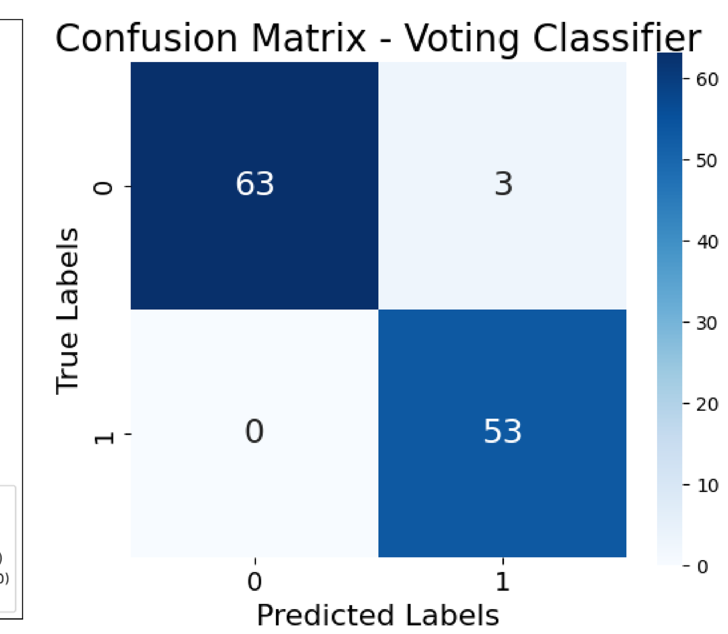
Results, Analysis and Discussion

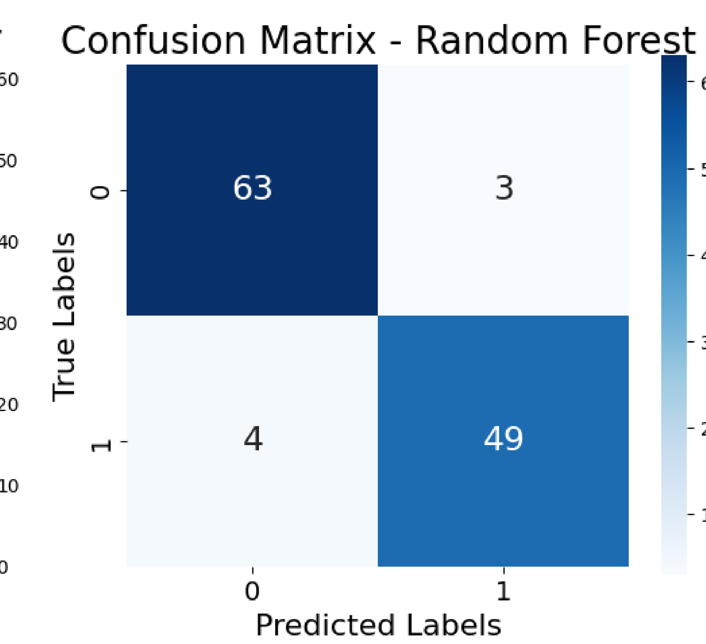
Models involved in Ensemble Learning	Training Accuracy	Testing Accuracy	F1 Score	ROC AUC Score
Decision Tree	0.93	0.87	0.87	0.91
KNN	0.92	0.97	0.96	0.99
SVC	0.98	0.98	0.98	1
Random Forest	0.96	0.94	0.93	0.99
Voting Classifier Mean Stratified Cross Validation Score- 0.95				
Independent Data Set Validation for Voting Classifier-0.93				











The supervised models (Decision Tree, KNN, SVC, Random Forest) show good performance, especially the SVC and Voting Classifier. In contrast, the unsupervised learning model, the Gaussian Mixture Model (GMM) for clustering, does not appear to perform as well. The low ARI and Silhouette scores suggest that there might not be a distinct or meaningful cluster structure in the data. This is particularly evident when the dataset is used for binary classification, where clustering might not be appropriate due to the natural grouping of the data into predefined classes rather than discovering new clusters. Overall, the supervised models are likely to be more suitable for this dataset, given their strong performance metrics, while the unsupervised GMM clustering approach may not be the right choice for this particular problem, possibly due to the absence of a natural cluster structure in the data or the binary nature of the classification task.

Conclusion

The project aimed to detect pipeline maintenance tools (pigs) using acoustic signals, processing audio data to extract features like Mel Frequency Cepstral Coefficients (MFCCs). We trained several machine learning models, including Decision Tree, KNN, SVM, RandomForest, a Feedforward Neural Network, and used Clustering. An innovative soft voting ensemble method was implemented to enhance predictive accuracy by combining model strengths. To address imbalanced datasets, we employed the Synthetic Minority Over-sampling Technique (SMOTE) to balance data. Model performance was evaluated using accuracy, F1 score, ROC AUC, and confusion matrices, with Stratified K-Fold Cross-Validation ensuring generalizability. The supervised models, particularly SVC and the Voting Classifier ensemble, performed well in detecting pigs in pipelines. In contrast, unsupervised Clustering was less effective due to the binary nature of the task. Future work could explore advanced neural networks.

Acknowledgements

The financial support from Midstream Integrity Services (MIS) and technical support from Smart Materials & Structures Lab (SMSL) and Artificial Intelligence Lab for Monitoring & Inspection (AILMI) at UH.

References (brief)

[1]S. Lee and B. Kim, “Machine Learning Model for Leak Detection Using Water Pipeline Vibration Sensor,” *Sensors (Basel)*, vol. 23, no. 21, Nov. 2023, doi: 10.3390/s23218935.
[2]H. Hassan, N. B. Ahmad, and S. Anuar, “Improved students’ performance prediction for multi-class imbalanced problems using hybrid and ensemble approach in educational data mining,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2020. doi: 10.1088/1742-6596/1529/5/052041.