Balancing the Imbalance: Advanced Anomaly Detection for Predictive Maintenance

Ch.Srivatsa Alivelu Mangatayi[1] Vanam Pavan[2]
Sri Ratna[3] G.Bharath[4] S.Narasimha[5]
[1] Assistant Professor, ACE Engineering College Hyderabad, India
[2]Student, ACE Engineering College Hyderabad, India [3]Student, ACE Engineering College Hyderabad, India [4]Student, ACE Engineering College Hyderabad, India
[5]Student, ACE Engineering College Hyderabad, India
Email : [1]srivatsajava@gmail.com [2]pavanvanam890@gmail.com
[3]srirathnatirunagari@gmail.com [4]bharathgurram443@gmail.com
[5] sangishettinarasimha@gmail.com

ABSTRACT:

Predictive maintenance is a crucial application in industrial systems, ensuring timely interventions to prevent failures and reduce downtime. However, the highly imbalanced nature of failure datasets poses a significant challenge to traditional machine learning models. This paper proposes an advanced anomaly detection framework utilizing synthetic data generation, oversampling techniques (SMOTE, ADASYN), and cost-sensitive learning to enhance failure detection accuracy. The system integrates Random Forest and XGBoost classifiers, optimized through hyperparameter tuning, to improve predictive maintenance efficiency. The proposed method demonstrates improved recall and precision, reducing false negatives while maintaining robustness in detecting rare failure events.

I. INTRODUCTION:

In industries reliant on machinery, unexpected failures lead to substantial financial losses and operational delays. Predictive maintenance, leveraging sensor data, aids in forecasting potential failures. However, the rarity of failure cases within datasets skews model performance, favoring normal conditions over anomalies. Addressing this issue requires advanced techniques in data preprocessing and machine learning model training. This paper introduces an innovative approach that enhances anomaly detection through synthetic data generation, resampling techniques, and cost-sensitive learning.

II. OBJECTIVES:

- Develop a robust predictive maintenance model for gas turbine systems.
- Address the issue of imbalanced datasets using SMOTE, ADASYN, and cost-sensitive learning.
- Train and optimize Random Forest and XGBoost models for anomaly detection.
- Deploy a real-time monitoring system for predictive maintenance.

III. PROBLEM STATEMENT

Predictive maintenance is a critical component of industrial asset management, aiming to mitigate unplanned downtimes and optimize operational efficiency. However, the rarity of failure events in sensor datasets leads to a highly imbalanced data distribution, which significantly hampers the effectiveness of traditional anomaly detection models. The system is designed to work with gas turbine Machine models such as GE 7F.05 and Siemens SGT-800. Conventional machine learning approaches often exhibit bias towards the majority class, leading to suboptimal recall rates for failure detection. This challenge necessitates the development of an advanced anomaly detection framework that leverages synthetic data augmentation, cost-sensitive learning, and ensemble-based models to improve predictive accuracy. By integrating techniques such as SMOTE, ADASYN, and hyperparameter-optimized classifiers like Random Forest and XGBoost, this study proposes a robust methodology to enhance failure prediction capabilities, ensuring proactive maintenance interventions and reducing operational risks in industrial settings.

IV. PROPOSED SYSTEM

The system is designed to work with **gas turbine Machine** models such as **GE 7F.05** and **Siemens SGT-800**, utilizing sensor data to detect anomalies and predict failures.

The proposed system incorporates:

Data Collection: Utilizing real-world sensor datasets such as the NASA CMAPSS dataset and synthetic data generation.

Data Preprocessing: Handling missing values, feature engineering, and normalization.

Handling Imbalance: Implementing SMOTE, ADASYN, and cost-sensitive learning to balance datasets.

Model Training & Optimization: Utilizing Random Forest and XGBoost, fine-tuned via GridSearchCV.

Deployment: Integrating the trained model into a real-time monitoring system.

V. SOFTWARE REQUIREMENTS

➢ Platform : Jupyter Notebook,VS code

➢ Technologies :

- **Programming**: Python,JavaScript

- **Libraries**: Pandas, NumPy, Scikit-learn, Imbalanced-learn, XGBoost

- **Frameworks**: Flask (for deployment), Next.js
- **Database**: MongoDB (for storing sensor data)
- **Visualization**: Plotly Dash / Streamlit (for monitoring dashboard)

➢ Testing :

- Postman

VI. TECHNOLOGY DESCRIPTION

Machine and Model:

- **Gas Turbine Models**: GE 7F.05, Siemens SGT-800
- **Machine Learning Models**: Random Forest, XGBoost

Machine Learning Techniques:

1. **Random Forest Classifier**: A robust ensemble model trained with class-weight adjustments.
2. **XGBoost**: A gradient boosting model optimized for imbalance handling.
3. **SMOTE & ADASYN**: Synthetic oversampling methods to balance dataset distribution.
4. **Hyperparameter Tuning**: Optimizing models via GridSearchCV.

Deployment Framework:

- A **Flask API** for model inference.
- A **Streamlit dashboard** for real-time anomaly monitoring.
- A **MongoDB database** to store historical sensor data.

VII. ALGORITHM

Step 1: Data Preprocessing

- Load sensor datasets from real-world or synthetic sources.
- Handle missing values using forward-fill and interpolation techniques.
- Feature engineering: Calculate rate-of-change metrics and rolling averages.
- Normalize features using StandardScaler.

Step 2: Handling Imbalanced Data

- Apply **SMOTE and ADASYN** to generate synthetic failure cases.
- Implement **cost-sensitive learning** by adjusting class weights in classifiers.

Step 3: Model Training & Evaluation

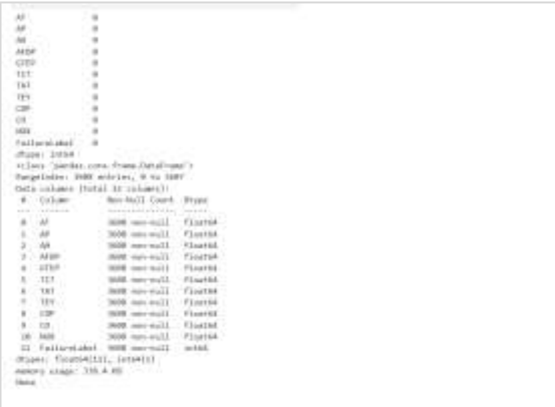
- Train **Random Forest and XGBoost** models on resampled data.
- Evaluate performance using classification report and AUC-ROC scores.
- Optimize hyperparameters via **GridSearchCV**.

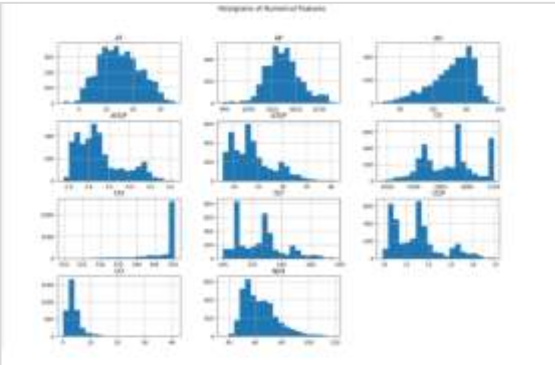
Step 4: Deployment & Real-Time Prediction

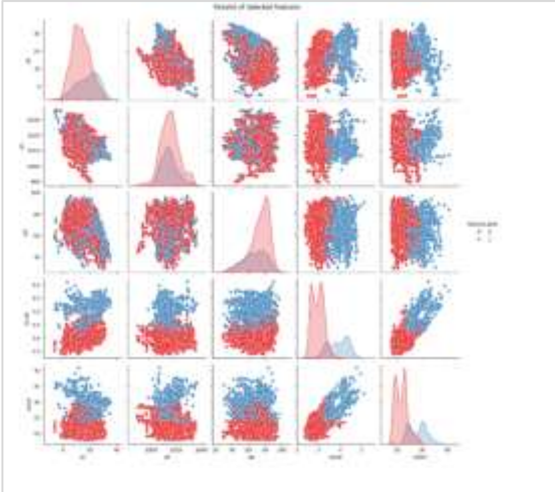
- Save trained models using **joblib**.
- Develop a **Flask API** for failure prediction.
- Implement **Streamlit-based visualization** for

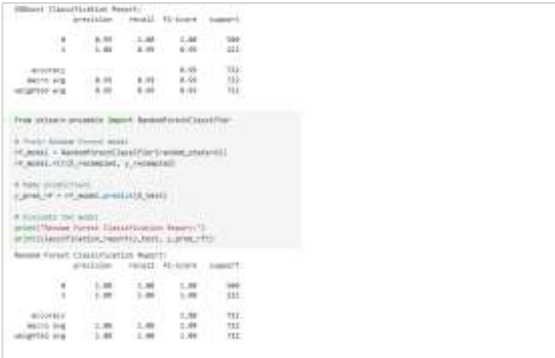
sensor monitoring.

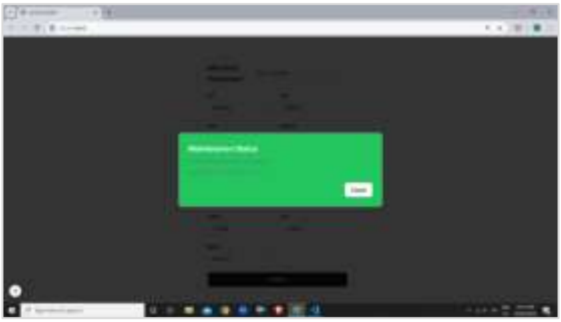
VIII. OUTPUT SCREENS











IX. CONCLUSION

This paper presents an advanced anomaly detection framework for predictive maintenance, addressing class imbalance through synthetic oversampling and SMOTE and ADASYN Balancing Techniques. The proposed approach enhances failure detection accuracy, ensuring reliable and timely maintenance interventions. Future work includes integrating IoT-based real-time data collection and XAI Frameworks learning techniques for adaptive anomaly detection.

X. REFERENCES

[1] Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). Learning from Imbalanced Data Sets. Springer.

[2] Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. Neural Networks, 106, 249-259.

[3] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[4] Saxena, A., & Goebel, K. (2008). Turbofan Engine Degradation Simulation Data Set. NASA Ames Prognostics Data Repository.