

Balancing the Imbalance: Advanced Anomaly Detection for Predictive Maintenance

Ch.Srivatsa Alivelu Mangatayi¹ Vanam Pavan² Sri Ratna³ G.Bharath⁴ S.Narasimha⁵ ¹ Assistant Professor, ACE Engineering College Hyderabad, India ²Student, ACE Engineering College Hyderabad, India ⁴Student, ACE Engineering College Hyderabad, India ⁵Student, ACE Engineering College Hyderabad, India ⁵Student, ACE Engineering College Hyderabad, India Email : <u>¹srivatsajava@gmail.com</u> ²pavanvanam890@gmail.com ³srirathnatirunagari@gmail.com ⁴bharathgurram443@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 realtime 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,

© 2025, IJSREM | <u>www.ijsrem.com</u>

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

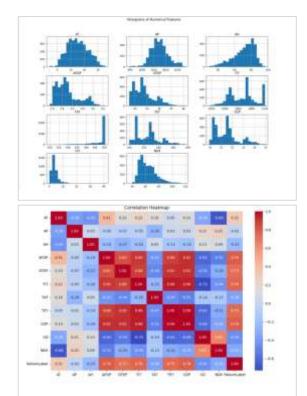


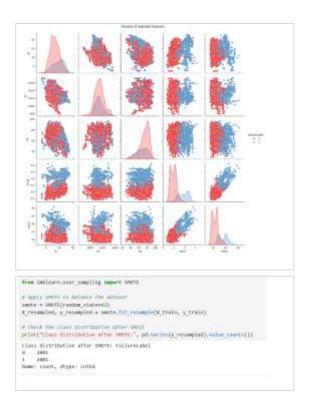
sensor monitoring.

VIII. OUTPUT SCREENS

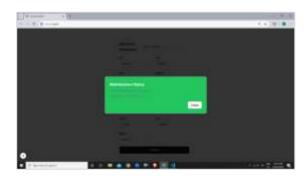
144	ward join	640. mit pr	63								
		17.84.19									
16	Gert April	continue.	eglet #	411							
	ware said	NUMBER OF	**								
	in nath :		22231	17 1745							
	ta v joi.										
1	10.00	Contraction of the	CONTRACTOR OF								
-	and the second second	the stress	1000								
* Printing the first for own artistidate bend())											
	10		- 20	APP				117		201	
	35,6018	10015-2	10.000	3 7715	34,347	10.00	1244.00	184.65	10.1017		
1	12,4088	3015-3	85,921	2,7719	11.367	1945.7	550,10	188.99	18.387		
ì	12,4588	1013-3 1013-8	\$3,921 \$5,971	2,7719	11.357	1945.7 1941.4	550,10 546,51	188.00	18.357 18.271		
1	12.4588 14.5278 34.5178	1015-3 1015-8 1844-8	83,323 35,577 12,784	2,1719 1,5410 1,7525	15.357	1945.7 1941.4 1813.8	558,08 548,01 548,01	186.00 186.00 1311.11	18.387 18.271 18.188		
1	12,4588	1015-3 1015-8 1844-8	83,323 35,577 12,784	2,1719 1,5410 1,7525	15.357	1945.7 1941.4 1813.8	558,08 548,01 548,01	186.00 186.00 1311.11	18.387 18.271 18.188		
1	12,4558 14,5978 34,5978 9,6718	0013-3 0013-8 10013-8 10017-5	83,323 35,577 12,784	2,7719 1,5410 1,7525 3,9403	15.357	1945.7 1941.4 1813.8	558,08 548,01 548,01	186.00 186.00 1311.11	18.387 18.271 18.188		
1111	12,4588 14,5978 34,5978 9,6718	9015-3 9015-9 1015-9 10168-9 1017-5 900	83,923 35,877 82,764 87,548 Falloret	2,7719 1,5410 1,7525 3,9403	15.357	1945.7 1941.4 1813.8	558,08 548,01 548,01	186.00 186.00 1311.11	18.387 18.271 18.188		
	12,4688 14,5078 34,5078 9,4758 (0) 1,7854 4,7755	0013-3 1013-9 1013-9 10177-3 10177-3 10077-3 10077-3 10077-3	85,921 35,977 82,784 87,548 Fallscrat	2,7719 1,5410 1,7525 3,9403	15.357	1945.7 1941.4 1813.8	558,08 548,01 548,01	186.00 186.00 1311.11	18.387 18.271 18.188		
	12,4689 14,5079 34,5178 9,4718 60 1,7844 4,7793 4,7795	9015-3 1015-9 58144-9 9017-5 403 43.871 95.899 71,100	85,921 35,977 82,784 87,548 Fallscrat	2,7719 1,5410 1,7525 3,9403	15.357	1945.7 1941.4 1813.8	558,08 548,01 548,01	186.00 186.00 1311.11	18.387 18.271 18.188		
	12,4688 14,5078 34,5078 9,4758 (0) 1,7854 4,7755	1015-3 1015-8 5814-8 1017-3 403 13.871 66.899 11.100 63.877	85,921 35,977 82,784 87,548 Fallscrat	2,7719 1,5410 1,7525 3,9403	15.357	1945.7 1941.4 1813.8	558,08 548,01 548,01	186.00 186.00 1311.11	18.387 18.271 18.188		

TPD 0 15 0 14 0 27 0 28 0 28 0 29 0 20 0 21 0 21 0 21 0 21 0 21 0 21 0 21 0 22 0 23 0 24 0 25 0 26 0 27 0 26 0 27 0 28 0 29 0 20 0 20 0 29 0 20 0 20 0 20 0 20 0 20 0 20 0 20 0 20 </th <th></th> <th></th> <th></th> <th></th>				
450° 0 110° 0 121° 0				
450° 0 170° 0 171° 0 171° 0 272° 0	iii i			
450° 0 170° 0 171° 0 171° 0 274° 0 275° 0	Ad .			
TPD 0 11 0 141 0 151 0 152 0 154 0 155 0 154 0 155 0 154 0 155 0 155 0 156 0 157 0 156 0 156 0 157 0 158 0 158 0.00 158 0.00 158 0.00 158 0.00 158 0.00 158 0.00 158 0.00 158 0.00 158 0.00 158 0.00 159 0.00 159 0.00 159 0.00 159 0.00 159 0.00 150 0.00	14:54			
11 0 13 0 15 0 16 0 17 0 18 0 19 0 10 0 11 0 12 0 13 0 14 0 15 0 15 0 15 0 16 0 17 0 18 0 19 0 19 0 10 0 11 0 12 0 13 0 14 0 15 0 16 0 17 0 18 0 17 0 18 0 17 0 18 0 19 0 107 0				
bit 0 15% 0 20% 0 21%	tit.			
15 0 200 0 11 0 12 0 12 0 12 0 12 0 12 0 12 0 12 0 12 0 12 0 12 0 12 0 12 0 13 0 14 0 15 0 16 0 17 0 18 0 19 0 10 0 10 0 11 0 12 0 12 0 12 0 12 0 12 0 12 0 12 0 12 0 12 0 12 0	Tal.			
20 0 20 20 20 20 20 20 20 20 21 20 20 20 21 20 22 20 24 20 25 20 26 20 27 20 28 20 29 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 200 20 200 20 200 20 200 20 200 20 200	100		G	
N B Statustical B Statustical B Statustical B Statustical B Statustical B B M B M B M B M B M Statustical Statustical Statustical Statustical A Statustical B Statustical B Statustical B Statustical B Statustical B Statustical B Statustical <tr< td=""><td>-</td><td></td><td></td><td></td></tr<>	-			
Bit Bit Stank.kej			-	
Allowinksi B Class Semistic cores Semistic core	in i			
Name: Date: Control Starting Starting Starting Starting Starting Starting Starting Starting </td <td></td> <td></td> <td></td> <td></td>				
Line Section cons. Argument's equilation: 100 evices, 0 exc. 100 evices, 1				
Separation DBM setup: State			10000	
bit bit<				
6 Color Non-Nacl Court Propriet 8 Al Non-Nacl Court Final Piel 9 Non-Nacl Court Final Piel Final Piel 9 Non-Nacl Court				P
0 A/I and momental Clarité 1 A/I 3000 merculi. Clarité 2 A/II 3000 merculi. Clarité 3 A/II 3000 merculi. Clarité 4 A/III 3000 merculi. Clarité 4 111 3000 merculi. Clarité 4 112 3000 merculi. Clarité 1 113 3000 merculi. Clarité 1 113 3000 merculi. Clarité 1 114 3000 merculi. Clarité 1 115 3000 merculi. Clarité 1 116 3000 merculi. Clarité 1 116 3000 merculi. Clarité 10 3000 merculi. Clarité 116 11 116 3000 merculi. Clarité 12 Fairté 116 3000 116 13 116 116 116 300 14 116 11				
θ 0.000 mm =-0.11 Cartrié 4.07 0.000 mm =-0.11 Cartrié 5.48 0.000 mm =-0.11 Cartrié 5.48 0.000 mm =-0.11 Cartrié 6.11 0.000 mm =-0.11 Cartrié 6.11 0.000 mm =-0.11 Cartrié 7 1.11 0.000 mm =-0.11 Cartrié 7 1.000 mm =-0.11 Cartrié 0 9 1.29 0.000 mm =-0.11 Cartrié 9 1.29 0.000 mm =-0.11 Cartré 1.29 <				
4 3000 mm wall Γματιδά 9 400 3000 mm wall Γματιδά 9 4000 3000 mm wall Γματιδά 9 4000 3000 mm wall Γματιδά 8 112 3000 mm wall Γματιδά 8 112 3000 mm wall Γματιδά 9 100 3000 mm wall Γματιδά 9 100 3000 mm wall Γματιδά 10 3000 mm wall Γματιδά 1 10 3000 mm wall 1 1 10 3000 mm w				
2 44 964 executil Start4 5 449 864 executil Start4 8 259 864 executil Start4 8 259 864 executil Start4 8 259 864 executil Start4 8 101 866 executil Start4 9 107 966 executil Start4 9 107 966 executil Start4 9 107 966 executil Start4 9 107 966 executil Start4 9 109 109 1096 executil Start4 9 109 109 1096 ex				
All MODE Control Control A ATTN MODE Non-Noil Flagstab A TTN MODE Non-Noil Flagstab A TTN MODE Non-Noil Flagstab B MODE MODE Non-Noil Flagstab B MODE MODE Non-Noil Flagstab B MODE MODE MODE MODE B MODE MODE	10.0			
III BBH mervail Faurth III BBH mervail Faurth A 101 BBH mervail Faurth III Faurth BBH mervail III Faurth BBH mervail BBH mervail MBH mervail Bath III Faurth Bath BBH mervail BBH mervail Bath III Faurth Bath	2			
N 101 Mole marrial: Finantia V 101 Mole marrial: Finantia V 107 Mole marrial: Finantia V 109 Mole marrial: Finantia V 100 Mole marrial: Finantia V 100 Mole marrial: Finantia				
A UNI MOME manuality Classifie 9 UNIV MOME manuality Classifie 9 2009 MOME manuality Classifie 9 2009 MOME manuality Classifie 10 MOME manuality Classifie MOME manuality 10 MOME manuality MOME manuality MOME manuality				
UN 2004 mm-mail Family 9 129 2008 mm-mail Family 9 129 2008 mm-mail Family 9 129 2008 mm-mail Family 12 2008 mm-mail Family Family 12 5008 mm-mail Family Family 12 Family 2008 mm-mail Family 13 Family 2008 mm-mail Family 14 Family 2008 mm-mail Family 15 Family 2008 mm-mail Family 16 2008 mm-mail Family 2008 mm-mail 17 Family 2008 mm-mail Family 16 2008 mm-mail 2008 mm-mail 2008 mm-mail 17 Family 2008 mm-mail 2008 mm-mail 16 2008 mm-mail 2008 mm-mail 2008 mm-mail 17 Family 2008 mm-mail 2008 mm-mail 16 2008 mm-mail 2008 mm-mail 2008 mm-mail 16	×.	.127		
127 108 109 101 102 100 1			BUR not tall	Tion 184
127 108 109 101 102 100 1		184	3008 mm mult	First#4
9 D3 3000 severall Charted 10 Man 3000 severall Playted T fulleyiaded NUM several sets figure (footbell), istebil) severa state: (30.4.400		1.25	1908 meretall	Finness
of Add 2000 unreall Florid II fulleriador WHE mereall with Taper florid-life (enter) were stage 104 Add		110	3000 mm +ull	Giated
11 Falleriakol WHW moreall with Egge: ficatielli, istaeli) antro sing: 335.4.00	100			Plost64
Higger: flourd-Hilli, inte4(s) among sings: 3th A MU	10	Fallsreig		
enery stage: 3th & #5				
	_			





	tastint be		the based	t-meril					
1.0	4.05	1.00	1.86	- 100					
	1.00	1.00							
			1.53	104					
internet alle	4.00	8.00	8.00	10					
			2.0						
tes adapte av	antin ins	-		-10171er					
C. NOREL & BARR	ontinue()	ai. Histories	-	re-11					
			9						
a la constante de la constante	irran Lines			10 a.					
hanna furset c	THE DESIGNATION OF THE DESIGNATI			Name?					
1	1.00	1.00	1.0	1000 111					
motoresia and documents and documents	1.8	1.49	1,98	111. 111					
	1.140	1.49	1.046	111					
and the set	- 1.00	1.00	1.00	111					
(+	- 16	1.4	1.00						
	- 16	1.0	1.00	911			-		
(+	- 16	1.0	1.00	ta Matter Fasking	50 Tores			- 00	
(+	- 16	1.0		Martine Facilities	-			14	
(+	- 16	1.00		Martine Facilities				100	
(+	- 16	1.4	1.00		-			03	
(+	- 16	1.4	1.00	Martin 2.11 1.11 1.11 1.11	-	2		3	
(+	- 16	1.4	1.00		-	1			
(+	- 16	1.4	1.00	Mattan Antisa An	1 1 1	1			
(+	- 16		1.00	Mathie Nachdal a an a an a an a an a an a an a an	-			3	
(+	- 16		1.00	Matter Ma	· · · · · · · · · · · · · · · · · · ·				
(+	- 16		1.00		11111			**3	
(+	- 16		1.00		· · · · · · · · · · · · · · · · · · ·				
(+	- 16		1.00		· · · · · · · · · · · · · · · · · · ·				
4 (1000)	- 16				· · · · · · · · · · · · · · · · · · ·				





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.