Predictive Maintenance and Condition-Based Monitoring (CBM) for Maritime Drive Systems Using Machine Learning Techniques

Introduction

Objective: To demonstrate the application of machine learning techniques for predictive maintenance and CBM in maritime drive systems.

Problem Statement

- This dataset records operational performance metrics of a ship's gas turbine propulsion system. Each row represents a set of measurements taken under specific conditions, likely at different times or operational states. The parameters captured include various torque measurements, rates of revolutions, temperatures, pressures, fuel flow, and state coefficients. These metrics are crucial for monitoring and analyzing the propulsion system's performance, efficiency, and maintenance needs.
- The data can be used for predictive maintenance, performance optimization, and anomaly detection. By analyzing the relationship between these parameters, one can identify patterns that indicate potential issues or areas for improvement in the propulsion system.

Data Collection

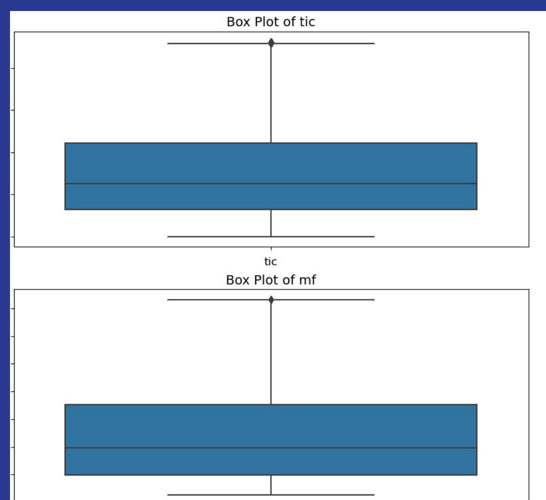
- Data Sources and Features:
- Data collected from sensors on maritime drive systems
- Features include Lever position (RPM), Ship speed (knots),Gas Turbine (GT) shaft torque (kN m), and more
- Target variable: GT Turbine decay state coefficient

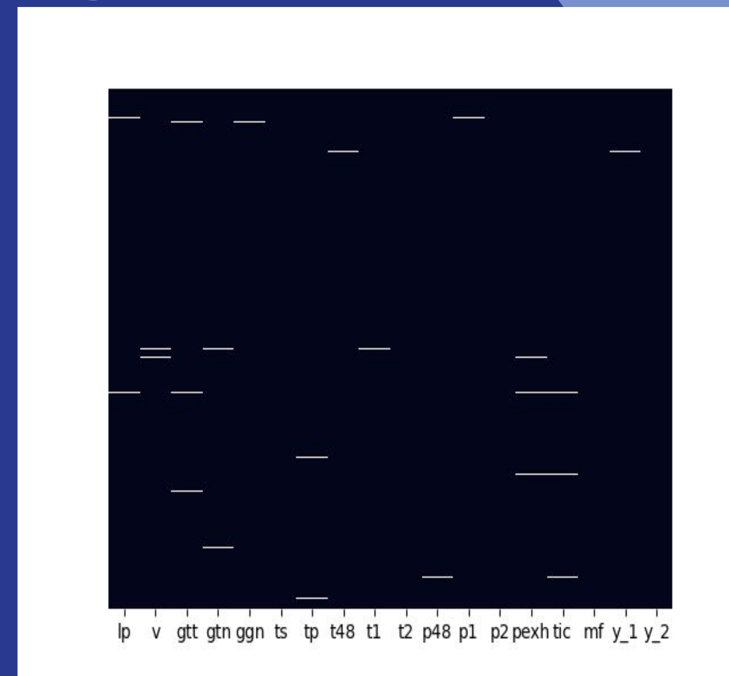
Data Cleaning

- Data appeared to be in the form of object data type
- Missing Values Imputation(KNN Imputation)
- Handling Duplicate Values

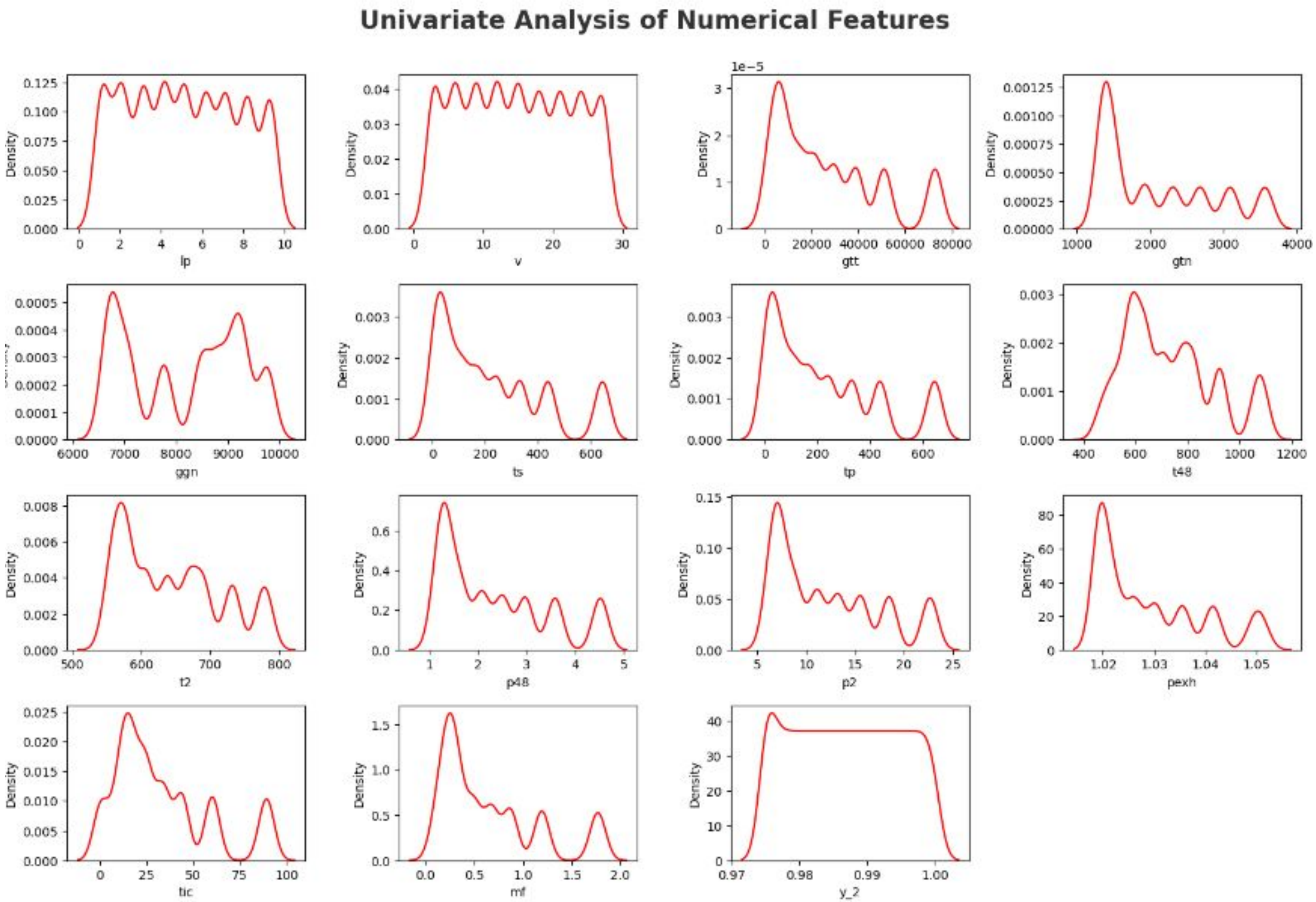
```
df.duplicated().sum()

43
```

- Handling outliers





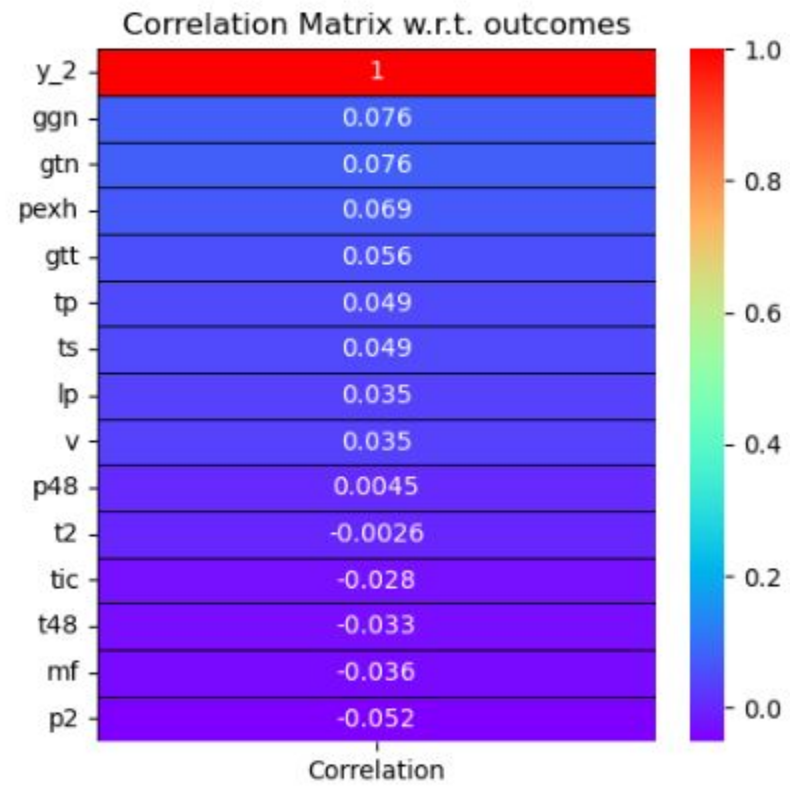
Exploratory Data Analysis



Exploratory Data Analysis

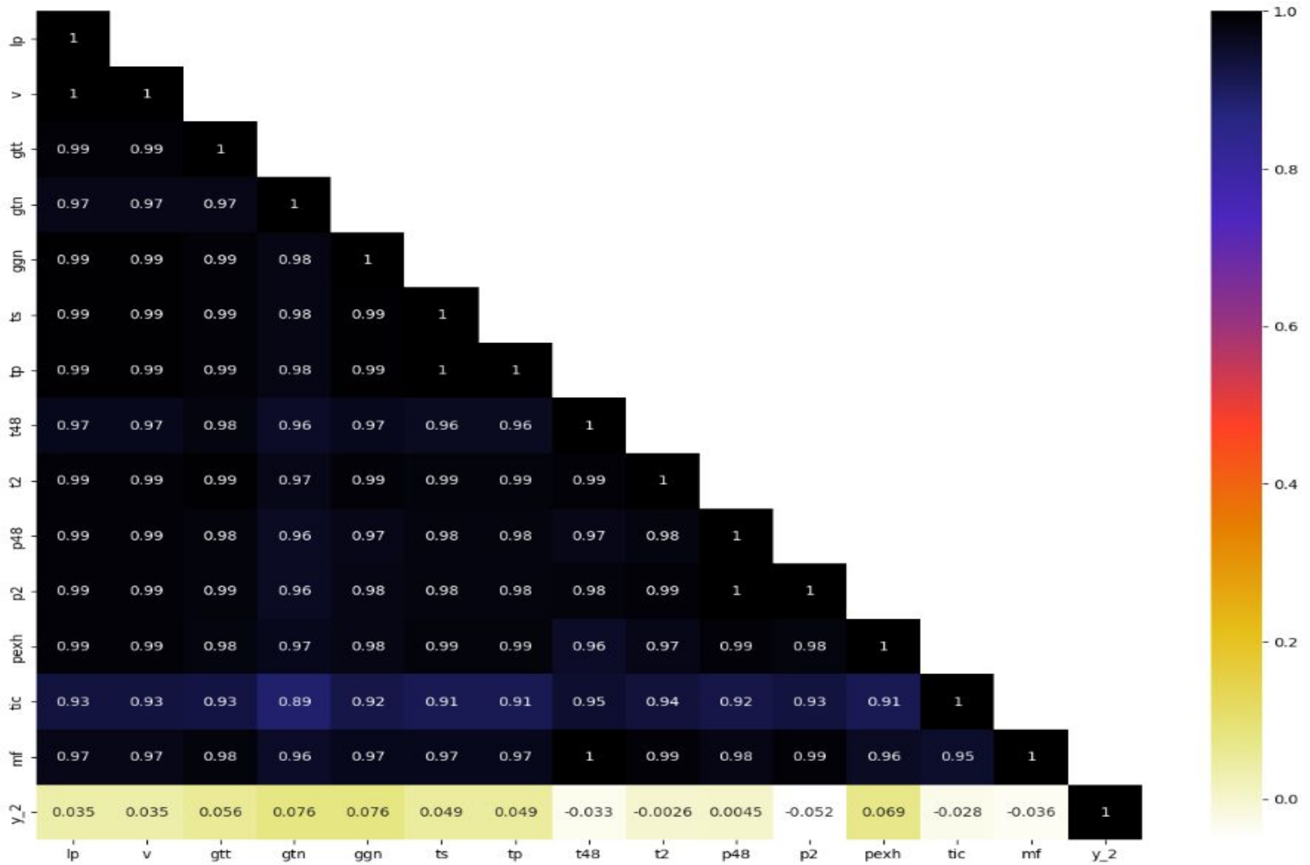
Feature Selection

There is Multicollinearity between some Features And all the features are weakly correlated with the target variable



Feature Selection

Multicollinearity?

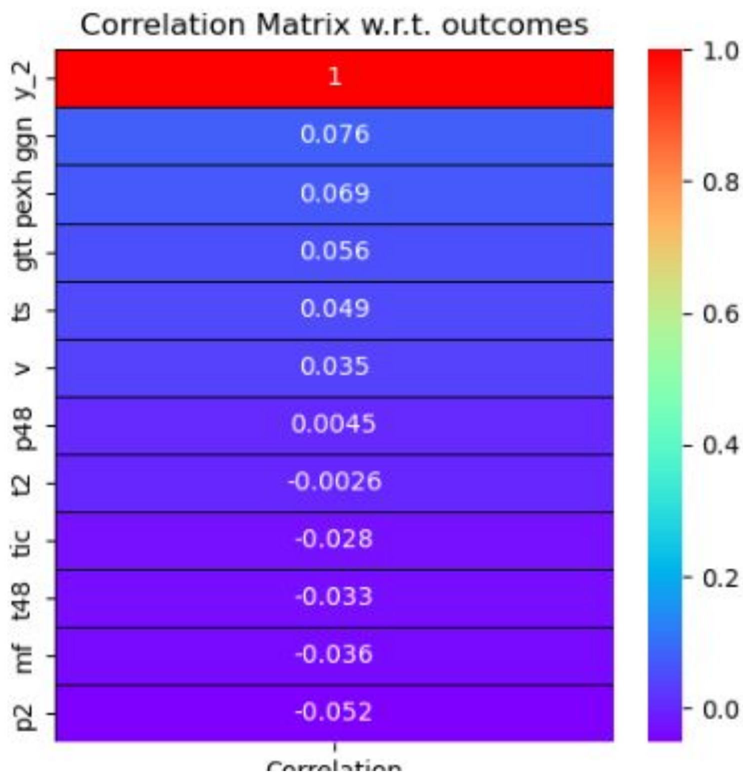


	Variable	VIF
0	lp	4.182897e+04
1	v	3.593633e+04
2	gtt	3.594344e+04
3	gtn	2.076585e+03
4	ggn	4.933077e+02
5	ts	inf
6	tp	inf
7	t48	6.017449e+03
8	t2	1.530424e+03
9	p48	3.334107e+04
10	p2	2.214146e+04
11	pexh	1.291398e+03
12	tic	7.234065e+01
13	mf	1.730938e+04
14	y_2	1.125156e+01

Feature Selection

Approach 1 To handle Multicollinearity

Dropping features Such as lp, tp & gtn



	Variable	VIF
0	v	518.690501
1	gtt	17980.056854
2	ggn	316.566362
3	ts	21413.640865
4	t48	4461.480612
5	t2	954.085556
6	p48	26771.524813
7	p2	18379.939562
8	pexh	1166.209645
9	tic	63.761759
10	mf	10904.957020
11	y_2	8.028856

Feature Extraction

Approach 2 To handle Multicollinearity(Dimensionality reduction)

- Creating new features with the already present ones
- Using PCA for Dimensionality Reduction
- PCA transformed 14 features into 7 principal components

	0	1	2	3	4	5	6
0	4.205703	-0.074201	-0.010264	0.070737	0.037790	0.201319	-0.064949
1	4.089529	-0.104002	-0.196696	-0.077565	0.038060	0.151875	0.023093
2	7.111566	0.790785	-0.042792	-0.087382	-0.087373	-0.106686	0.060332
3	-3.223670	0.047824	0.086142	-0.072109	-0.190482	0.072546	-0.029383
4	-3.987746	0.267118	-0.364208	0.113781	-0.170464	0.028041	0.001007
...	...	...	...	...	...	...	...
9551	2.188874	-0.488066	-0.185049	-0.072165	0.098363	0.009454	0.037211
9552	-3.254282	0.072633	0.061709	-0.103414	-0.209112	0.088293	-0.016324
9553	2.238284	-0.451433	-0.149043	-0.044428	0.097472	0.030337	0.059906
9554	4.109063	-0.103519	-0.171409	-0.053098	0.038768	0.158711	0.008325
9555	4.085011	-0.116102	-0.191976	-0.069625	0.037996	0.150225	0.006372

Model Evaluation

Models Implemented

- Decision Tree Regression(base model)
- Random Forest Regressor
- Gradient Boosting
- AdaBoost Regressor
- XGBoost Regressor
- CatBoost Regressor
- KNN Regressor
- Random Forest Regressor(Tuned)
- KNN Regressor(Tuned)
- XGBoost Regressor(Tuned)
- CatBoost Regressor(Tuned)

Model Selection

Model Evaluation

MODEL	TRAINING SCORE	TESTING SCORE
Decision Tree Regression	1.000	0.9816
RandomForest Regressor	0.9988	0.9897
XGBoost regressor	0.988685	0.947674
Gradient Boost	0.8328	0.8190
AdaBoost Regressor	0.963432	0.929529
KNN Regressor	0.9786	0.9642
XGBoost Regressor	0.9928	0.9838
CatBoost Regressor	0.9933	0.9884

Model Selection

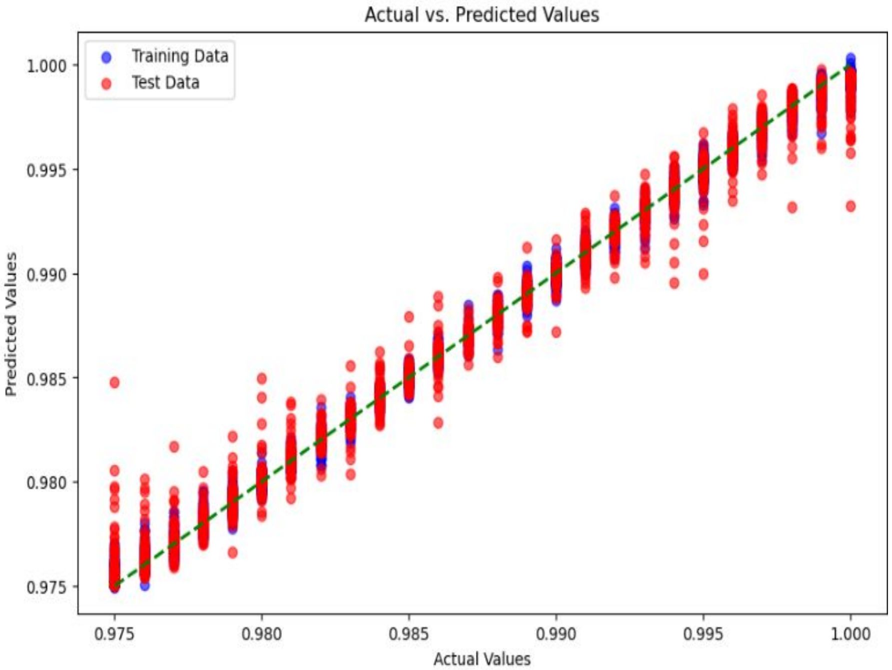
Model Evaluation(Tuned)

MODEL	TRAINING SCORE	TESTING SCORE
RandomForest Regressor	0.9988	0.9897
KNN Regressor	0.9931	0.9757
XGBoost Regressor	0.9943	0.9847
CatBoost Regressor	0.9984	0.9931

Model Selection

Model Performance and Insights:

XGBoost Regressor was chosen Which had MSE of $3.21484730740419 \times 10^{-7}$ and $7.799225102719444 \times 10^{-7}$ for train and test respectively and mean Accuracy of 98.45% after Cross-Validation



Model Selection

Model Evaluation(PCA)

MODEL	TRAINING SCORE	TESTING SCORE
Decision Tree Regression	1.000	0.8016
RandomForest Regressor	0.9980	0.9103
XGBoost regressor	0.988685	0.947674
Gradient Boost	0.8081	0.7256
AdaBoost Regressor	0.3060	0.2712
KNN Regressor	0.9741	0.9261
XGBoost Regressor	0.9916	0.9112
CatBoost Regressor	0.9945	0.9427

Model Selection

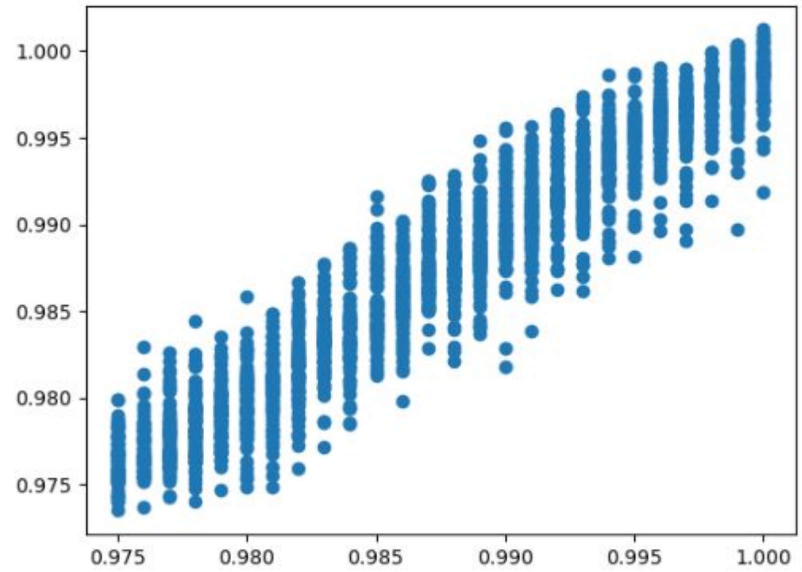
Model Evaluation PCA(Tuned)

MODEL	TRAINING SCORE	TESTING SCORE
RandomForest Regressor	0.7470	0.6843
KNN Regressor	0.9931	0.9757
XGBoost Regressor	0.9893	0.9178
CatBoost Regressor	0.9977	0.9445

Model Selection

Model Performance and Insights:

XGBoost Regressor was chosen Which had MSE of 1.3266507980553283 × 10-7 and 2.9824599748076083 × 10-6 for train and test respectively and mean Accuracy of 98.16% after Cross-Validation



Conclusion

- PCA effectively reduced the dimensionality of our model
- This reduction did not significantly impact the model's performance.
- Machine learning techniques improved predictive maintenance
- Training R^2 score: 0.9951
- Test R^2 score: 0.9872

Dimensionality Reduction

- Using PCA for Dimensionality Reduction:
- PCA transformed 14 features into 7 principal components
- Selected components explain over 95% of the variance
- Benefits: Reduced multicollinearity, improved computational efficiency

Results

- Model Performance and Insights:
- High predictive accuracy with reduced features
- Effective for predictive maintenance and CBM
- Potential to reduce downtime and maintenance costs

Conclusion

- Summary and Future Work:
- PCA effectively reduced the dimensionality of our model
- Machine learning techniques improved predictive maintenance
- Future work: Explore other dimensionality reduction techniques, further analyze model interpretability

Q&A

- Questions and Discussion