IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 8, August 2026

DOI: 10.17148/IJIREEICE.2026.14805

Predictive Maintenance using Python with Silhouette Analysis

VINEETA SAHU[1], Ms. DOLLY VERMA[2], DR. ASHISH TAMRAKAR[3]

Student, Computer Science and Engineering, RSR Rungta College of Engineering and Technology, Bhilai, Chhattisgarh, India[1]

Assistant Professor, Computer Science and Engineering, RSR Rungta College of Engineering and Technology, Bhilai, Chhattisgarh, India[2]

Assistant Professor, Computer Science and Engineering, RSR Rungta College of Engineering and Technology, Bhilai, Chhattisgarh, India[3]

Abstract: Predictive maintenance (PdM) has emerged as a key application of artificial intelligence, machine learning, and data analytics in modern industry. Unlike preventive maintenance, predictive maintenance utilizes historical and real-time sensor data to identify equipment degradation and estimate potential failures before they occur. Python has become one of the most widely adopted programming languages for predictive maintenance due to its extensive ecosystem of scientific and machine learning libraries such as NumPy, Pandas, Scikit-learn, Matplotlib, and Seaborn. Among clustering validation techniques, silhouette analysis plays a significant role in determining the optimal number of clusters and evaluating the quality of clustering results. This review paper presents a comprehensive overview of predictive maintenance using Python with silhouette analysis. It discusses maintenance strategies, machine learning approaches, clustering techniques, silhouette coefficient interpretation, Python-based implementation frameworks, publicly available datasets, and recent research contributions. The paper also highlights advantages, limitations, and future research directions in predictive maintenance systems employing unsupervised learning and silhouette-based cluster validation.

Keywords: Predictive Maintenance, Python, Silhouette Analysis, Machine Learning, K-Means Clustering, Unsupervised Learning, Industrial Analytics, Condition Monitoring.

I. INTRODUCTION

Industries such as manufacturing, automotive, aerospace, energy, transportation, and process control heavily depend on the continuous operation of machinery and equipment. Unexpected machine failures result in production downtime, maintenance costs, safety risks, and reduced operational efficiency. Traditional maintenance strategies such as reactive maintenance (repair after failure) and preventive maintenance (scheduled servicing) often lead to unnecessary maintenance actions or catastrophic failures.

Predictive maintenance (PdM) addresses these limitations by continuously monitoring machine conditions through sensors and data acquisition systems. Parameters such as vibration, temperature, pressure, humidity, voltage, current, and acoustic emissions are analyzed to predict future failures. Machine learning algorithms can identify hidden patterns in sensor data and detect abnormal operating conditions before equipment breakdown occurs.

Python provides an ideal environment for predictive maintenance because of its open-source libraries for data preprocessing, visualization, clustering, classification, regression, and anomaly detection. In unsupervised predictive maintenance, clustering algorithms are frequently used to separate healthy and faulty operating conditions. Silhouette analysis evaluates clustering performance by measuring how well each data point belongs to its assigned cluster compared to neighboring clusters.This review focuses on the integration of Python and silhouette analysis for predictive maintenance applications.

II. PREDICTIVE MAINTENANCE OVERVIEW

A. Maintenance Strategies

Maintenance strategies can be classified into four major categories.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.021 ※ Peer-reviewed & Refereed journal ※ Vol. 14, Issue 8, August 2026
DOI: 10.17148/IJIREEICE.2026.14805

1) Reactive Maintenance: Equipment is repaired only after failure occurs. This approach has low planning requirements but results in high downtime and repair costs.

2) Preventive Maintenance: Maintenance activities are performed at predefined time intervals regardless of equipment condition. Although more reliable than reactive maintenance, it often replaces components prematurely.

3) Condition-Based Maintenance: Sensor measurements are periodically monitored, and maintenance decisions are based on equipment condition.

4) Predictive Maintenance: Advanced analytics and machine learning predict future failures using historical and real-time data, enabling maintenance only when necessary.

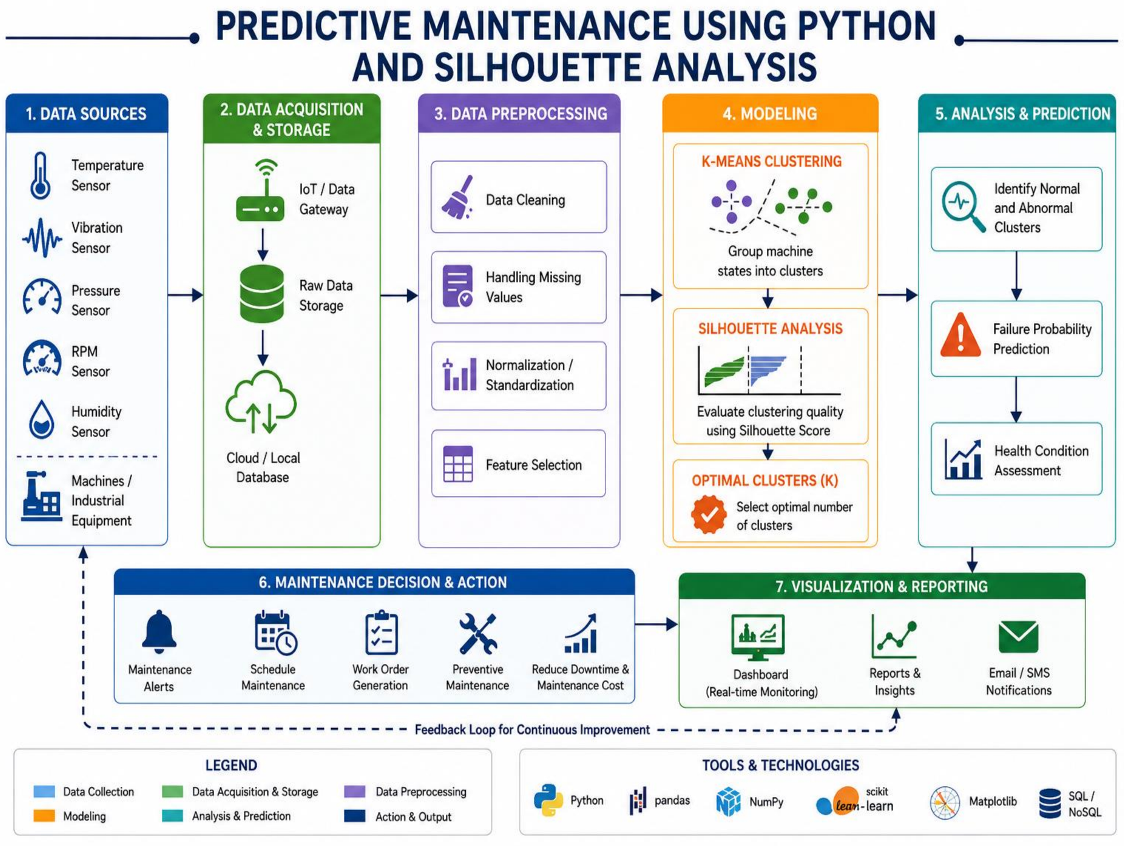
B. Predictive Maintenance Workflow

A typical predictive maintenance system includes:

- Sensor data acquisition
- Data preprocessing
- Feature extraction
- Clustering or classification
- Failure prediction
- Maintenance scheduling

The workflow can be implemented efficiently in Python using libraries such as Pandas, NumPy, Scikit-learn, and Matplotlib.

C. Block Diagram



International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 8, August 2026
DOI: 10.17148/IJIREEICE.2026.14805

III. PYTHON FOR PREDICTIVE MAINTENANCE

Python has become the dominant programming language for industrial analytics because of its simplicity, portability, and extensive library ecosystem.

A. Common Python Libraries

Library	Purpose
NumPy	Numerical computation
Pandas	Data manipulation
Matplotlib	Visualization
Seaborn	Statistical plots
Scikit-learn	Machine learning algorithms
SciPy	Scientific computing
TensorFlow / PyTorch	Deep learning

B. Data Preprocessing

Predictive maintenance datasets typically require preprocessing before machine learning analysis. Common preprocessing operations include:

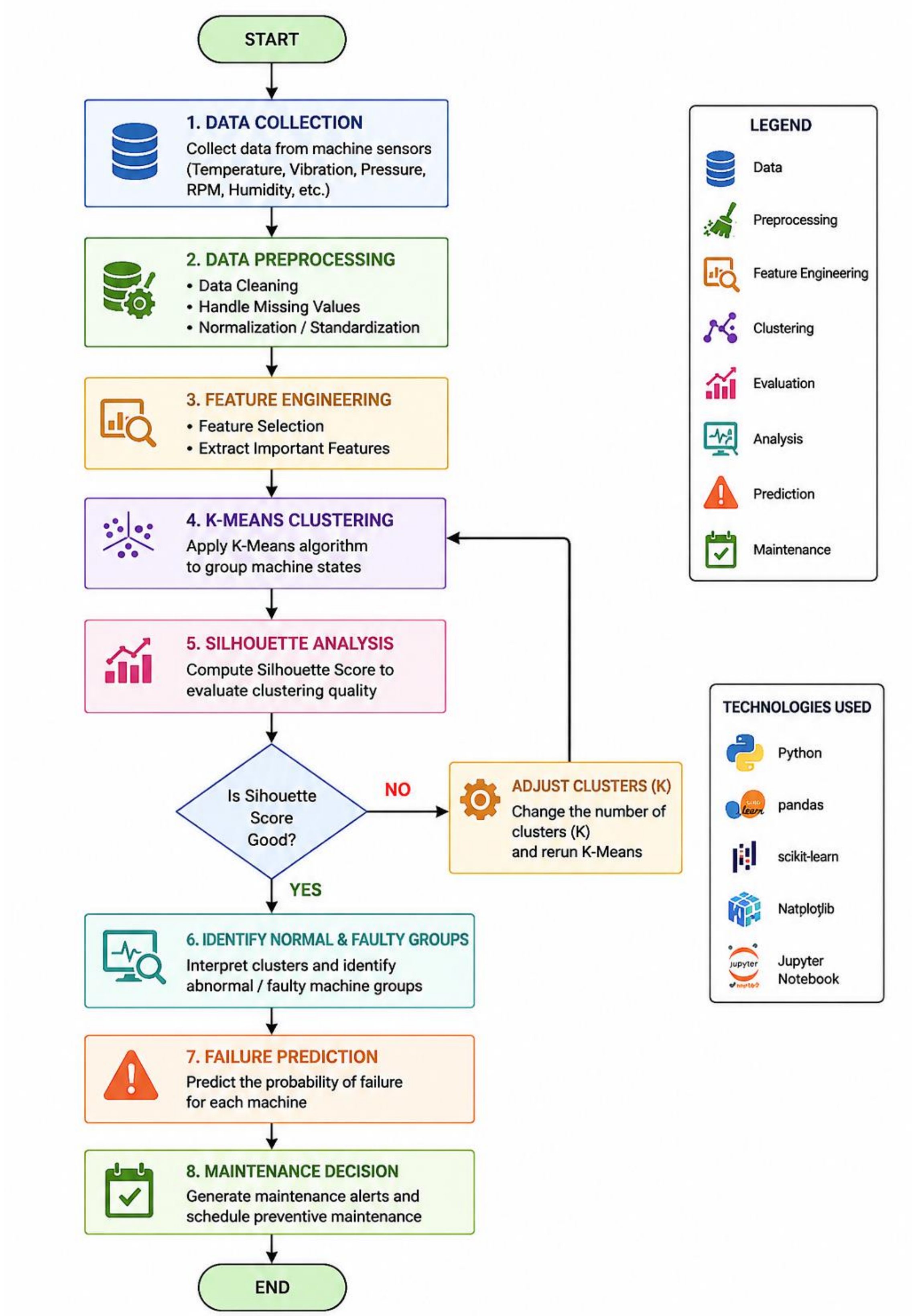
- Missing value handling
- Outlier removal
- Feature scaling
- Normalization
- Standardization
- Noise filtering

Example Python preprocessing:

```
from sklearn.preprocessing import StandardScaler scaler = StandardScaler() X_scaled = scaler.fit_transform(data)
```

IJIREEICE
ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 8, August 2026
DOI: 10.17148/IJIREEICE.2026.14805

C. Flowchart



ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.021 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 8, August 2026
DOI: 10.17148/IJIREEICE.2026.14805

IV. CLUSTERING IN PREDICTIVE MAINTENANCE

Clustering is an unsupervised learning technique that groups similar observations without requiring labeled failure data.

A. K-Means Clustering

K-Means is one of the most widely used clustering algorithms for predictive maintenance due to its simplicity and computational efficiency.
Algorithm steps:

1. Select K cluster centers
2. Assign each data point to the nearest center
3. Update cluster centers
4. Repeat until convergence

Python implementation:

```
from sklearn.cluster import KMeans model = KMeans(n_clusters=3, random_state=42) labels =
model.fit_predict(X_scaled)
```

B. Applications in Predictive Maintenance

Clustering can identify:

- Healthy operating conditions
- Early degradation
- Severe fault conditions
- Unknown abnormal patterns
- Equipment operating modes

V. SILHOUETTE ANALYSIS

A. Definition

Silhouette analysis evaluates clustering quality by comparing intra-cluster similarity and inter-cluster separation.
For a sample i:
$s(i) = (b(i) - a(i)) / \max(a(i), b(i))$
where:

- a(i): average distance within the same cluster
- b(i): average distance to the nearest neighboring cluster

The silhouette coefficient ranges from -1 to +1.

B. Interpretation

Silhouette Value	Interpretation
+1	Excellent clustering
0.5–0.9	Good clustering
0.25–0.5	Weak clustering
Less than 0	Incorrect clustering

C. Python Implementation

```
from sklearn.metrics import silhouette_score score = silhouette_score(X_scaled, labels) print(score)
```

D. Selecting Optimal K

Silhouette analysis is commonly used to determine the optimal number of clusters.

DOI: 10.17148/IJIREEICE.2026.14805

```
scores = [] for k in range(2, 8):     labels = KMeans(n_clusters=k).fit_predict(X_scaled)
scores.append(silhouette_score(X_scaled, labels))
```

The value of K with the highest silhouette score is considered optimal.

VI. LITERATURE REVIEW

A. Review of Recent Studies

Predictive maintenance (PdM) has become a significant research area in Industry 4.0 because it enables industries to detect equipment degradation before failure occurs, thereby reducing maintenance costs, downtime, and production losses. Modern predictive maintenance systems rely on sensor data, machine learning algorithms, and data analytics to continuously monitor the health of industrial equipment.

Jardine *et al.* [1] presented a comprehensive review of machinery diagnostics and prognostics for condition-based maintenance systems. Their work established the foundation for predictive maintenance by emphasizing continuous monitoring of equipment parameters such as vibration, temperature, and pressure. The authors showed that data-driven maintenance strategies improve equipment reliability and operational efficiency.

Lei *et al.* [2] conducted a systematic review of machinery health prognostics and remaining useful life (RUL) prediction techniques. They highlighted the importance of data preprocessing, feature extraction, and machine learning methods in predictive maintenance systems. Their study demonstrated that data-driven approaches are highly effective for analyzing complex industrial equipment.

Python has become one of the most widely used programming languages for predictive maintenance because of its powerful machine learning ecosystem. Pedregosa *et al.* [3] introduced the Scikit-learn library, which provides efficient implementations of clustering algorithms, preprocessing methods, and silhouette score evaluation. Python libraries such as NumPy, Pandas, and Matplotlib enable researchers to perform complete predictive maintenance workflows, including data analysis, clustering, and visualization.

Clustering techniques are particularly useful in predictive maintenance because industrial datasets often contain unlabeled operational data. MacQueen [4] proposed the K-Means clustering algorithm, which groups similar observations into clusters based on feature similarity. In predictive maintenance applications, K-Means is commonly used to identify healthy, degraded, and faulty machine operating conditions without requiring predefined fault labels.

A major challenge in K-Means clustering is selecting the appropriate number of clusters. Rousseeuw [5] introduced silhouette analysis as a method for validating clustering quality. The silhouette coefficient evaluates both intra-cluster cohesion and inter-cluster separation, making it an effective technique for determining the optimal number of clusters. This approach improves the accuracy and reliability of predictive maintenance systems by ensuring well-defined operational groups.

Public benchmark datasets have significantly supported predictive maintenance research. The AI4I 2020 Predictive Maintenance Dataset from the UCI Machine Learning Repository [6] contains temperature, rotational speed, torque, tool wear, and machine failure information for industrial equipment. This dataset is widely used for evaluating clustering algorithms and predictive maintenance models implemented in Python.

Hastie, Tibshirani, and Friedman [7] discussed clustering and statistical learning techniques for large numerical datasets, while Mitchell [8] provided the theoretical foundation of machine learning methods used for predictive analytics. These studies support the integration of K-Means clustering and silhouette analysis for identifying hidden operational patterns in industrial sensor data.

Based on the reviewed literature, it is evident that predictive maintenance systems benefit significantly from Python-based machine learning frameworks and clustering validation techniques. However, many existing studies focus primarily on supervised learning methods, while fewer studies investigate unsupervised clustering with silhouette-based optimization. Therefore, this review focuses on the application of Python, K-Means clustering, and silhouette analysis for efficient and interpretable predictive maintenance of industrial equipment.

DOI: 10.17148/IJIREEICE.2026.14805

C. Silhouette Analysis in Predictive Maintenance

Rousseeuw [6] introduced silhouette analysis as a graphical aid for cluster interpretation and validation. The method has been widely adopted for determining optimal cluster numbers and evaluating clustering quality. In predictive maintenance applications, silhouette analysis serves multiple purposes: determining the optimal number of operating states or fault conditions, evaluating the quality of clustering results, improving anomaly detection by identifying well-separated clusters, and reducing false maintenance alarms through reliable cluster validation.

Recent studies have demonstrated the practical application of silhouette analysis in industrial settings. A master's thesis investigating predictive maintenance for welding robots employed silhouette analysis extensively to validate clustering results. The study applied custom-distance clustering algorithms to weld current and voltage curves, using silhouette scores to select optimal numbers of clusters. Silhouette scores above 0.50 indicated good clustering performance, while lower scores suggested ambiguous cluster boundaries. The research highlighted the importance of correlating clustering results with metadata for meaningful interpretation.

In the aviation domain, Silhouette scores were used as unsupervised evaluation metrics alongside supervised metrics such as the Rand Index for clustering brake wear patterns . The dual evaluation approach provided comprehensive validation of clustering results.

VII. PUBLIC DATASETS FOR PREDICTIVE MAINTENANCE

Researchers frequently use publicly available datasets for experimentation.

A. NASA Turbofan Engine Dataset

Contains engine degradation data with multiple sensor measurements and remaining useful life labels.

B. UCI Predictive Maintenance Dataset

Includes:

- Temperature
- Rotational speed
- Torque
- Tool wear
- Failure labels

C. Bearing Fault Datasets

Common datasets include:

- CWRU Bearing Dataset
- IMS Bearing Dataset
- FEMTO Bearing Dataset

These datasets are suitable for clustering and silhouette-based analysis.

VIII. ADVANTAGES OF SILHOUETTE-BASED PREDICTIVE MAINTENANCE

A. Benefits

- Determines optimal number of clusters
- Evaluates clustering quality objectively
- Improves anomaly detection
- Reduces false maintenance alarms
- Supports unsupervised fault discovery
- Easy implementation in Python

Impact Factor 8.021 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 8, August 2026

DOI: 10.17148/IJIREEICE.2026.14805

B. Industrial Advantages

- Reduced downtime
- Lower maintenance costs
- Increased equipment availability
- Better resource planning
- Improved operational safety

IX. CHALLENGES AND LIMITATIONS

Despite significant advantages, several challenges remain.

A. Data Challenges

- Missing sensor values
- Noise
- Class imbalance
- High-dimensional features

B. Algorithmic Limitations

- K-Means assumes spherical clusters
- Sensitive to initialization
- Silhouette computation becomes expensive for very large datasets
- Performance decreases with overlapping clusters

C. Industrial Deployment Issues

- Sensor calibration
- Real-time processing
- Edge deployment constraints
- Cybersecurity
- Model interpretability

X. FUTURE RESEARCH DIRECTIONS

Future predictive maintenance research may include:

- Deep clustering algorithms
- Autoencoder-based feature extraction
- Graph neural networks
- Federated predictive maintenance
- Explainable silhouette-based clustering
- IoT and edge AI integration
- Remaining Useful Life estimation using hybrid clustering-classification models

Combining silhouette analysis with advanced unsupervised learning techniques can significantly improve predictive maintenance systems in smart manufacturing environments.

XI. CONCLUSION

Predictive maintenance has become an essential component of Industry 4.0 and intelligent manufacturing systems. Python provides a flexible and powerful platform for implementing predictive maintenance workflows involving data

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.021 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 8, August 2026
DOI: 10.17148/IJIREEICE.2026.14805

preprocessing, clustering, visualisation, and machine learning. Silhouette analysis is an effective cluster validation technique that helps determine the optimal number of clusters and improves the reliability of unsupervised fault detection models. The integration of K-Means clustering and silhouette analysis enables industries to identify equipment degradation patterns without requiring labeled failure data. This review highlights current methodologies, datasets, Python tools, advantages, limitations, and future opportunities in silhouette-based predictive maintenance research. The approach offers a cost-effective, scalable, and interpretable solution for modern industrial maintenance systems.

REFERENCES

[1] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
[2] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, 2009.
[3] M. H. Dunham, *Data Mining: Introductory and Advanced Topics*. Upper Saddle River, NJ, USA: Pearson, 2003.
[4] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2022.
[5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[6] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. New York, NY, USA: Springer, 2009.
[7] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
[8] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006.
[9] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, 2015.
[10] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," *Informatica*, vol. 31, no. 3, pp. 249–268, 2007.
[11] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mechanical Systems and Signal Processing*, vol. 104, pp. 799–834, 2018.
[12] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, USA, 1967, pp. 281–297.
[13] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.
[14] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.
[15] NASA Prognostics Center of Excellence, "Turbofan Engine Degradation Simulation Dataset," NASA Ames Research Center, Moffett Field, CA, USA, 2008.
[16] F. Pedregosa *et al*., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
[17] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.
[18] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Hoboken, NJ, USA: Pearson, 2021.
[19] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.
[20] UCI Machine Learning Repository, "AI4I 2020 Predictive Maintenance Dataset," University of California, Irvine, CA, USA, 2020.