UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Posgrados

Unsupervised Anomaly Detection with OC-SVM for Predictive Maintenance in Rotating Machinery

Proyecto de Titulación

Lenin Fernando Salinas Tamayo

José David Vega, Ph.D.

Director de Trabajo de Titulación

Trabajo de titulación de posgrado presentado como requisito para la obtención del título de Magíster en Ciencia de Datos

Quito, mayo 2025

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ
COLEGIO DE POSGRADOS

HOJA DE APROBACIÓN DE TRABAJO DE TITULACIÓN

Título Trabajo de Titulación

Nombre del estudiante

Nombre del Director del Programa:	Felipe Grijalva
Título académico:	Ph.D. en Ingeniería Eléctrica
Director del programa de:	Ciencia de Datos

Nombre del Decano del colegio Académico:	Eduardo Alba
Título académico:	Doctor en Ciencias Matemáticas
Decano del Colegio:	Ciencias e Ingenierías

Nombre del Decano del Colegio de Posgrados:	Dario Niebieskikwiat
Título académico:	Doctor en Física

Quito, mayo 2025

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombre del estudiante:	Lenin Fernando Salinas Tamayo
Código de estudiante:	00338525
C.I.:	1721350070
Lugar y fecha:	Quito, 12 de mayo de 2025

ACLARACIÓN PARA PUBLICACIÓN

UNPUBLISHED DOCUMENT

DEDICATORIA

A mi papá, por ser mi ejemplo de esfuerzo, perseverancia y sabiduría. A mi mamá, por su amor incondicional, su apoyo constante y su fe en mí incluso en los momentos más difíciles. A mi hermana, por su alegría, compañía y motivación, que me impulsaron a seguir adelante cada día.

A ustedes, que con su amor y respaldo han sido el pilar de este logro, les dedico cada página de este trabajo. Gracias por ser mi fuerza y mi inspiración.

AGRADECIMIENTOS

Agradezco de manera especial a Ándres López, quien trabajó conmigo en la extracción de la base de datos, demostrando compromiso y dedicación en cada etapa del proceso. Su apoyo técnico fue fundamental para la obtención de información precisa y oportuna.

De igual manera, expreso mi sincero agradecimiento a David Vega, quien, con su conocimiento y experiencia, logró orientar la directriz del proyecto de manera acertada, permitiendo encontrar soluciones eficaces a las problemáticas que surgieron durante su ejecución. Su asesoría y guía fueron clave para el desarrollo y fortalecimiento de esta investigación.

A ambos, mi profunda gratitud por su acompañamiento, paciencia y valiosas contribuciones.

RESUMEN

Este trabajo propone un enfoque basado en un aprendizaje no supervisado para la detección de anomalías en maquinaria rotativa industrial, utilizando Máquinas (OC-SVM). El estudio se basa en datos reales recopilados por un dispositivo de monitoreo Dynamox instalado en un motor que cuenta con una unidad reductora. El conjunto de datos incluye mediciones de temperatura, velocidad y aceleración, estos datos han sido recolectados en tres ejes. Tras aplicar procesos estándar de preprocesamiento y escalado de características, se entrenó un modelo OC-SVM sobre datos no etiquetados, utilizando etiquetas sintéticas de valores atípicos con fines de análisis. Se realizó un estudio de sensibilidad sobre los hiperparámetros clave (`nu`, `gamma`), identificando compensaciones claras entre las métricas de *recall* y precisión. Gracias a la reducción de dimensionalidad mediante t-SNE fue posible visualizar tanto los valores atípicos como la separación entre clústeres, mejorando la interpretabilidad del modelo. Los resultados muestran que OC-SVM es una excelente alternativa para la detección de anomalías en entornos con abundancia de sensores y ausencia de datos etiquetados sobre fallas. Los hallazgos respaldan su integración en sistemas de mantenimiento predictivo, donde la detección temprana de fallas es fundamental.

Palabras clave: Detección de anomalías, OC-SVM, aprendizaje no supervisado, mantenimiento predictivo, t-SNE, datos de sensores industriales, hiperparámetros de OC-SVM.

ABSTRACT

This paper approaches an unsupervised anomaly detection framework for industrial rotating machinery, One-Class Support Vector Machines (OC- SVM). this study is supported on real sensor data collected from a Dynamox monitoring device installed on a motor with a gear-reducing unit, the dataset includes temperature, velocity, and acceleration measurements collected across three axes. After standard preprocess- ing and feature scaling algorithms, the OC-SVM model was trained on unlabeled data with synthetic outlier labels used for analysis purposes. We performed a sen- sitivity study on key hyperparameter (nu, gamma) it was found distinct trade-offs between recall and precision. The reduction in dimensionality with t-SNE enabled the visualization of outliers and cluster separability, enhancing the interpretability of the model. The results showed that OC-SVM is a great alternative for anomaly detection in sensor-rich environments without labeled failure data. The results obtained support its integration into predictive maintenance pipelines where early fault detection is critical.

Key words: Anomaly detection, One-Class SVM, unsupervised learning, predictive maintenance, t-SNE, industrial sensor data, OC-SVM hyperparameters.

TABLA DE CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

1

Unsupervised Anomaly Detection with OC-SVM for Predictive Maintenance in Rotating Machinery

Lenin Salinas, David Vega

***Abstract*—This paper approaches an unsupervised anomaly detection framework for industrial rotating machinery, One-Class Support Vector Machines (OC-SVM). this study is supported on real sensor data collected from a Dynamox monitoring device installed on a motor with a gear-reducing unit, the dataset includes temperature, velocity, and acceleration measurements collected across three axes. After standard preprocessing and feature scaling algorithms, the OC-SVM model was trained on unlabeled data with synthetic outlier labels used for analysis purposes. We performed a sensitivity study on key hyperparameter (`nu`, `gamma`) it was found distinct trade-offs between recall and precision. The reduction in dimensionality with t-SNE enabled the visualization of outliers and cluster separability, enhancing the interpretability of the model. The results showed that OC-SVM is a great alternative for anomaly detection in sensor-rich environments without labeled failure data. The results obtained support its integration into predictive maintenance pipelines where early fault detection is critical.**

***Index Terms*—Anomaly detection, One-Class SVM, unsupervised learning, predictive maintenance, t-SNE, industrial sensor data, OC-SVM hyperparameters.**

I. Introduction

ONE of the major challenges in industrial, an engineering environments is the maintenance of machinery. The continuous operation of production systems *relies* on the good functioning of mechanical components. If we were capable of detecting early signs of malfunction and predicting potential failures, we could improve operational efficiency and minimize production delays. Most of the time, failures in critical systems, such as electric motors or gear reducers, are preceded by unusual patterns in sensor readings, including but not limited to increases in temperature or irregular vibration. Detecting these early changes is key to implementing effective predictive maintenance strategies. Thus, improving industry production.

This study explores an unsupervised approach to detecting early-stage anomalies in sensor data collected from a single industrial motor equipped with a gear reducing unit. The motor is instrumented with a Dynamox sensor that integrates accelerometer, velocimeter, and a thermocouple, enabling continuous measurement of acceleration, velocity, and temperature. This sensors was installed at two distinct observation points on the same motor: The input (LA) and output (LOA) sides of the shaft. This dual location setup allowed us to see and capture variations in mechanical and thermal behavior across different propagation paths in the same system. The final dataset, was created with over 176,000 time-stamped entries and serves as the foundation for the anomaly detection framework presented in this work.

As we mention before the primary objective is to identify outliers: data points that deviate from nominal operating conditions, and interpret them as early indicators of potential mechanical issues. For example, analysis of the dataset shows that temperatures above 60 °C fall within the top 5% of all records, and values exceeding 67 °C are in the top 1%, this may suggest overheating. Plus, most regular operation falls between 24 °C and 44 °C, based on the inter quartile range.

To identify such anomalies, this work uses One-Class Support Vector Machines (OC-SVM), a non-parametric method well suited for unsupervised anomaly detection this study works perfectly without the need of labeled data, this advantage makes it the best model for our investigation . OC-SVM constructs a decision boundary around normal instances in a high-dimensional feature space, enabling the detection of outliers as deviations from the learned normal profile [1]. This makes it particularly accurate for industrial settings where failure events are rare and unlabeled, where preserving model simplicity and interpretability is necessary. To facilitate interpretation, the sensor data is also transformed with a dimensionality reduction using t-distributed Stochastic Neighbor Embedding (t-SNE), which enables visual inspection of latent structure and deviations in a two-dimensional space.

We also carried out a performance evaluation using classification metrics based on simulated ground-truth scenarios. By testing different OC-SVM configurations, we explored how changing hyperparameters impacts the model's ability to correctly detect potential faults—measured through precision, recall, and the F1-score, which is the harmonic mean of precision and recall, is especially useful when the dataset is imbalanced and false positives or false negatives carry different costs [2]. Basically is a value that balances how many faults the model can catch (recall) and how many of its predictions are actually correct (precision). It's especially useful in studies like this were we don't have many faults in the data, which is often the case in real machines

Finally this work presents a modular and interpretable anomaly detection pipeline developed by using real sensor data obtained from industrial equipment. The pipeline is designed to operate without the need for labeled failure instances, making it suitable for practical deployment in environments where such labels are rarely available. In order to increase the effectiveness of the proposed approach, the study explores different configurations of the OC-SVM model and evaluates how each affects anomaly detection performance. Although the data itself is real, synthetic labels are introduced with evaluation purposes, allowing the computational metrics of standard performance data such as precision, recall, and F1-score. The resulting framework enables reliable outlier detection under real-world operating conditions and supports proactive, data-driven maintenance planning in complex mechanical systems.

Lenin Salinas

April, 2024

II. Prior Works

Anomaly detection has emerged as a central topic in predictive maintenance strategies, to be more specific in industrial contexts where labeled failure data is small or unavailable. The One-Class Support Vector Machine (OC-SVM) is widely used in this setting for its ability to model normal operational behavior and flag deviations as anomalies.

The foundational idea of OC-SVM builds upon the Support Vector Data Description (SVDD) introduced by Tax and Duin [1], which formulates the anomaly detection task as enclosing the majority of data points within a small hyper sphere in a transformed feature space. This kernel-based approach enables flexible modeling of complex distributions, even in high-dimensional settings.

In recent years, researchers have made efforts to improve the sensitivity of OC-SVM to local variations within the data. Traditional OC-SVM models often assume that data is distributed uniformly, this can lead to inaccurate boundaries and higher false positive rates with real data validations where this assumption doesn't hold. To address this, Mittal et al. [3] emphasized the growing importance of using machine learning techniques, such as OC-SVM, to perform anomaly detection in sensor-based industrial environments. They explained that OC-SVM maintain a valuable baseline due to its simplicity and effectiveness in identifying deviations in high-dimensional datasets where labeled failures are scarce. Their study also highlights the need of complementary strategies, for example: feature scaling and dimensionality reduction, to enhance OC-SVM's interpretability when struggling with noisy and heterogeneous sensor data.

On the other hand it has been found that hybrid methods have also gained traction. Lee et al. [4] demonstrated that combining deep learning architectures with OC-SVM improves anomaly detection in manufacturing environments. Similarly, Graß et al. [5] explored unsupervised methods for production line monitoring, showing the viability of anomaly based alerts in cyber physical systems.

Recent studies reinforced the potential of mixing deep learning with traditional anomaly detection models to enhance better detection in complex environments. For instance, the hybrid framework proposed in [6] couples autoencoder-derived embeddings with classical detectors such as OC-SVM, achieving greater resilience against noise and improving interpretability. Plus, recent advances in unsupervised deep learning, as explained in [7], show that end-to-end models can autonomously extract informative features from multivariate time-series sensor data, making possible high-precision anomaly detection without the need for labeled instances of failure. Both methods illustrate the growing tendency to combine deep neural networks with statistical models in order to address the challenges of industrial anomaly detection more effectively.

However, in a recent applied study, Mateo et al. [8] developed an ensemble system for anomaly auditing in packaging machines using OC-SVM and Minimum Covariance Determinant (MCD) classifiers. Their approach achieved improvements of up to 933% in F1-score compared to baseline models, underscoring the practical utility of OC-SVM in real-world deployments with no labeled faults.

III. Methodology

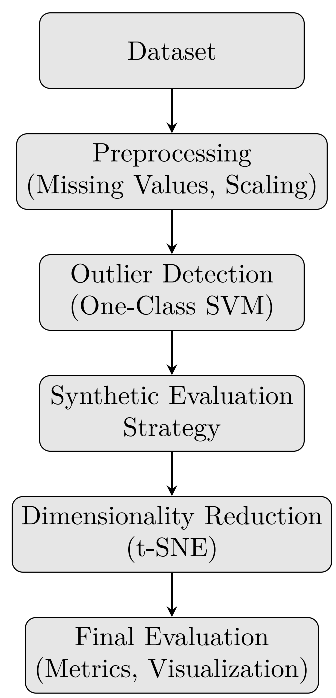

Figure 1. Overview of the anomaly detection pipeline.

A. Dataset Description

The dataset used in this study was collected from a single industrial motor equipped with a gear-reducing unit. A Dynamox sensor was installed on the motor to monitor its operational behavior. This sensor integrates multiple sensing capabilities, including a triaxial accelerometer, velocimeter, and a thermocouple, enabling the simultaneous recording of mechanical and thermal conditions from a single observation point.

This device recorded multiple execution cycles of the machine, resulting in a diverse dataset.

Each data point contains seven numerical features obtained from the sensor's measurements: A collection of temperature, acceleration and velocity components across the three orthogonal axes (X, Y, Z). In total, the dataset comprises 176,510 time-stamped entries. Each record also includes metadata such as equipment type, subsystem, and monitoring point. The monitoring point is identified by the values `LA` and `LOA`, which refer to the shaft's input and output sides, respectively. These labels indicate the physical location of the sensor on either side of the motor, providing insight into the direction and propagation of mechanical stress or heat during operation. These variables form the base of the dataset and represent the measurable physical quantities used for anomaly detection:

- **TEMPERATURE**: temperature in degrees Celsius.
- **VELOCITY X, VELOCITY Y, VELOCITY Z**: velocity components in the X, Y, and Z axes.
- **ACCELERATION X, ACCELERATION Y, ACCELERATION Z**: acceleration components in the X, Y, and Z axes.

These sensor readings provide a comprehensive view of the motor's dynamic behavior and serve as the foundation for the unsupervised anomaly detection models developed in this study.

Preliminary analysis demonstrated that temperature values range from $-46.16\,°C$ to $128.86\,°C$, with typical operation occurring between $24\,°C$ and $44\,°C$. A small percentage of extreme values suggests the presence of anomalies or measurement errors.

B. Data Preprocessing

Table I
Summary of input variables in the *BombaXLX* dataset

Variable	Description	Unit
TEMPERATURE	Temperature reading	°C
VELOCITY X	Velocity in X axis	mm/s
VELOCITY Y	Velocity in Y axis	mm/s
VELOCITY Z	Velocity in Z axis	mm/s
ACCELERATION X	Acceleration in X axis	m/s^2
ACCELERATION Y	Acceleration in Y axis	m/s^2
ACCELERATION Z	Acceleration in Z axis	m/s^2

The first step during the data pre-processing was to define which portions of the raw data were useful for modeling. The original dataset included contextual metadata such as equipment type and monitoring location. However, only the numerical sensor readings—temperature, acceleration, and velocity—were relevant for the anomaly detection task. Therefore, the scope was narrowed to include only the seven key numerical variables that reflect the motor's physical behavior.

Once the relevant data was selected, the exploratory data analysis (EDA) phase began. String entries such as `"NAN"` were converted to proper `NaN` representations to correct handling of missing values. Next, rows containing missing data in the selected variables were removed to keep consistency and avoid potential bias during model training.

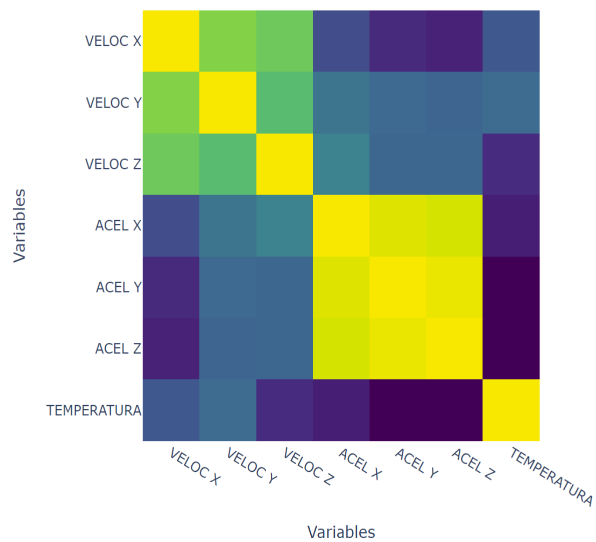

Figure 2. Correlation heatmap among physical variables measured by Dynamox sensors. High correlations between certain pairs, such as `VELOC Y` and `VELOC Z`, or `ACEL Y` and `ACEL Z`, suggest informational redundancy. On the other hand, the variable `TEMPERATURA` shows low correlation with the others, potentially providing complementary information relevant for anomaly detection.

The correlation heatmap shown in Figure 2 reveals high dependencies between several variables, particularly between the velocity and acceleration components on the Y and Z axes. Such relationships suggest potential redundancy that could affect model interpretability or induce multicollinearity in supervised learning settings. Interestingly, temperature exhibits low correlation with all other variables, reinforcing its value as an independent signal that may capture distinct types of anomalies—such as thermal drift or overheating—not directly observable through vibrational data alone.

Then, all numerical variables were standardized using the `StandardScaler` from the `scikit-learn` library. This standardization process was necessary to ensure that the One-Class SVM model could learn the underlying structure of the data without being biased by differences in scale. Plus without this normalization, variables with larger magnitudes could disproportionately influence the model, ending in biased or suboptimal results.

Lastly, for evaluation purposes, a synthetic binary label named `true_outlier` was introduced to emulate ground truth in the absence of real failure annotations. This label was generated by randomly assigning a small fraction (e.g., 10%) of the samples as outliers, under the assumption that anomalies are rare in industrial monitoring data. While not reflecting actual faults, this approach enables quantitative performance comparisons using classification metrics such as precision, recall, and F1-score.

To better understand the spatial distribution of anomalies detected by the OC-SVM, we applied a dimensionality reduction: the high-dimensional data was projected into two dimensions using *t-distributed Stochastic Neighbor Embedding* (t-SNE). This non-linear dimensionality reduction technique is well-suited for visualizing complex datasets where clusters and outliers may not be linearly separable.

The input to t-SNE consisted of the scaled feature matrix. The resulting 2D coordinates were combined with the OC-SVM outlier predictions and temperature values for visualization. Each point in the plot represents a single data instance, where:

- Marker **shape** encodes whether the instance was detected as an outlier or not.
- Marker **color** corresponds to temperature, providing understanding into thermal variation.

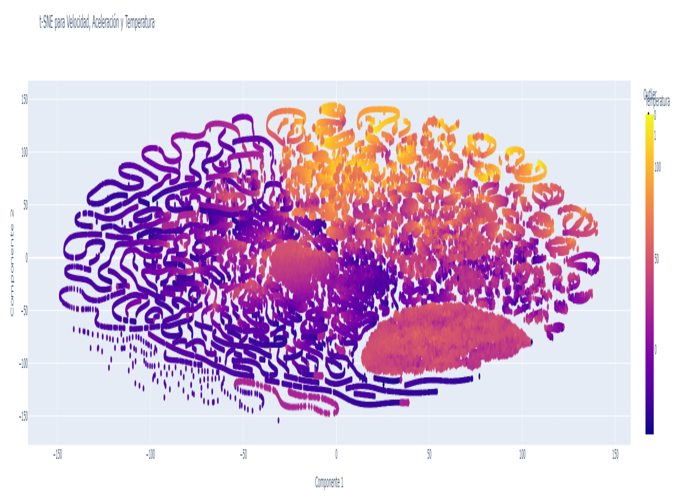

Figure 3. t-SNE projection of temperature, velocity, and acceleration features. Outliers are indicated by symbols, and temperature is captured by color. This visualization highlights cluster separation and thermal gradients for visual interpretability.

As illustrated in Figure 3, several dense regions arise, corresponding to normal operational clusters. Outliers tend to appear in low-density areas and are often associated with temperature values outside the normal range. This reinforces the hypothesis that thermal variation may be a leading indicator of anomalous machine behavior in this dataset.

C. Synthetic Evaluation Strategy

In the absence of ground-truth failure labels, this study used synthetic labels to simulate a binary classification scenario. A column named `true_outlier` was introduced, with values randomly assigned to mimic a plausible 90%–10% split between normal and anomalous data. While such labeling does not reflect real-world failures, it provides a controlled environment for testing the sensitivity and robustness of the One-Class SVM (OC-SVM) under various parameter settings.

The dataset was split into training and validation subsets using stratified sampling to preserve the class distribution. Several OC-SVM models were trained on the same scaled training data, each using a different configuration of the `nu` and `gamma` hyperparameters. The models were then evaluated on the validation set. To measure performance, standard classification metrics were computed:

- **Precision**: the fraction of predicted outliers that were correctly identified.
- **Recall**: the fraction of actual outliers correctly detected by the model.
- **F1-score**: the harmonic mean of precision and recall, providing a balanced metric.

Table II
Performance metrics for different OC-SVM configurations using synthetic labels.

Configuration	Recall	Precision	F1-score
$\nu = 0.05$, $\gamma =$ scale	0.35	0.92	0.50
$\nu = 0.15$, $\gamma =$ scale	0.95	0.70	0.80
$\nu = 0.10$, $\gamma = 0.001$	0.52	0.89	0.66
$\nu = 0.10$, $\gamma = 1.0$	0.75	0.41	0.53
$\nu = 0.20$, $\gamma =$ scale	0.98	0.61	0.75
$\nu = 0.05$, $\gamma = 0.01$	0.29	0.91	0.44

Plus, a second evaluation approach was employed to compare OC-SVM configurations against a baseline model ($\nu = 0.1$, $\gamma =$ scale) served as a reference for evaluation This reference model provided consistent pseudo-labels for benchmarking the relative effectiveness of other configurations.

Table III
Performance of OC-SVM models under different hyperparameter combinations (relative to baseline model).

Configuration	Recall	Precision	F1-score
$\nu = 0.15$, $\gamma =$ scale	1.000	0.667	0.800
$\nu = 0.20$, $\gamma =$ scale	1.000	0.500	0.666
$\nu = 0.10$, $\gamma = 1.0$	0.666	0.666	0.666
$\nu = 0.05$, $\gamma =$ scale	0.500	1.000	0.667
$\nu = 0.10$, $\gamma = 0.001$	0.454	0.454	0.454
$\nu = 0.05$, $\gamma = 0.01$	0.366	0.733	0.488

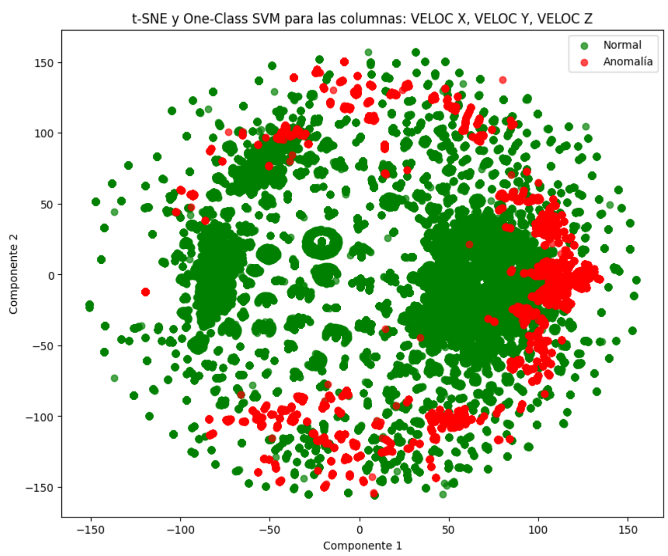

Figure 4. t-SNE projection of velocity components (VELOC X, VELOC Y, VELOC Z) using One-Class SVM. Red points represent anomalies detected by the model, while green points indicate normal behavior. This visualization reveals broad clusters with moderately defined separation.

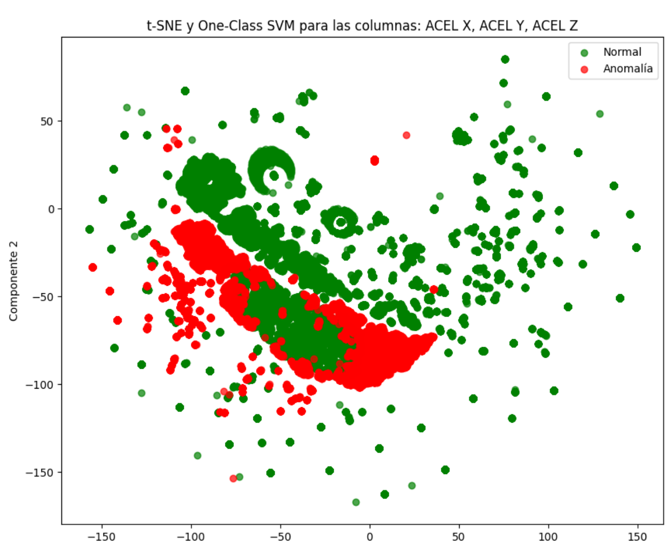

Figure 5. t-SNE projection of acceleration components (ACEL X, ACEL Y, ACEL Z) using One-Class SVM. A greater density of anomalies can be observed, suggesting higher sensitivity of acceleration signals to irregular behaviors.

The visualizations suggest that velocity-based features submit more distinct clustering between normal and anomalous data compared to acceleration. This shows us that velocity signals may capture clearer patterns of abnormal behavior in the motor's operation, while acceleration data contributes useful but more dispersed information.

Comparison Between Tables II and III. Table II presents performance against randomly generated synthetic labels, simulating a realistic, and artificial scenario in the absence of known failures. This approach evaluates the OC-SVM's general ability to separate a minority anomaly class under blur assumptions.

In contrast, Table III benchmarks each configuration against a fixed OC-SVM model ($\nu = 0.1$, $\gamma =$ scale), using it as a reference to generate pseudo labels. This relative comparison highlights how similar or divergent each model is in capturing the same decision boundaries, providing insight into the stability and variance of the learned anomalies.

Overall, these complementary tables enable a broader understanding of the OC-SVM's behavior under uncertainty and without real failure ground truth. They provide an absolute and a relative lens through which we can interpret model performance.

IV. Discussion and Insights

The results from the OC-SVM experiments revealed elegant trade-offs between sensitivity and specificity based on hyperparameter selection. It was found that configurations with high `nu` values, such as 0.15 or 0.20, resulted in excellent recall successfully capturing nearly all anomalous samples but at the cost of precision, which declined as more normal points were misclassified as outliers. Conversely, more conservative settings (e.g., `nu`=0.05) yielded better precision but missed a larger fraction of true outliers.

The effect of `gamma` was also fundamental: lower values produced smoother decision boundaries, resulting in higher generalization but lower detection sensitivity. In contrast, aggressive values such as `gamma=1.0` captured complex data patterns but increased false positives. This confirms the idea that `gamma` controls the model's local flexibility and that its interaction with `nu` must be carefully tuned.

Plus, the use of t-SNE visualizations provided interpretable insights into the clustering structure and temperature gradients. Observations flagged as outliers often appeared at the margins or limit zones in the t-SNE projection, reinforcing the model's capacity to detect subtle deviations in thermal and also mechanical behavior.

From a practical stance, these findings support the deployment of OC-SVM in early-warning systems for rotating machinery. Operators can choose hyperparameter settings depending on their risk tolerance, either favoring comprehensive fault detection or reducing false alarms in production environments. Future work could explore semi-supervised extensions or incorporate temporal dynamics for enhanced diagnostic improvement.

V. Conclusion

This study introduced an unsupervised approach for detecting anomalies in sensor data collected from industrial rotating machinery, improving the capabilities of One-Class Support Vector Machines (OC-SVM). The proposed methodology combined careful data preprocessing, dimensionality reduction, and a dual evaluation strategy to address the common challenge of missing labeled fault data in real-world applications.

Our results highlight how sensitive the OC-SVM model is to its hyperparameters particularly `nu` and `gamma` which directly affect the balance between recall and precision. Through visual exploration using t-SNE, we were able to understand how outliers relate to underlying thermal patterns, reinforcing the value of combining quantitative and visual diagnostics.

Overall, the OC-SVM is used as a effective tool for anomaly detection in environments where fault labels are hard to obtain or even unavailable. Its simplicity and interpretability make it perfect for deployment in a wide range of industrial settings, specially those aiming for proactive and cost-efficient maintenance.

Furthermore, this work goes beyond anomaly flagging by proposing a structured evaluation framework. Designing two complementary strategies: One using synthetic ground-truth labels and another pointing against a baseline model, we were able to test both the model's general detection capability and its consistency across parameter settings. This revealed meaningful trade-offs between sensitivity and specificity, showing that the OC-SVM can be fine-tuned depending on whether early detection or false alarm reduction is prioritized.

Recommendation for Future Work: To further enhance model performance, future studies should consider hybrid approaches that integrate temporal dynamics or expert-defined rules. Combining OC-SVM with time-series embedding methods, recurrent architectures, or ensemble voting mechanisms could improve robustness in highly variable industrial environments, and increase its applicability in fault prognosis.

References

[1] D. M. Tax and R. P. Duin, "Support vector data description," *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004. [Online]. Available: https://link.springer.com/article/10.1023/B:MACH.0000008084.60811.49

[2] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets," in *PloS One*, vol. 10, no. 3. Public Library of Science, 2015. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118432

[3] A. Mittal, A. Dave, D. Thakkar, D. Kanchetti, and R. Munirathnam, "Anomaly detection in sensor data with machine learning: Predictive maintenance for industrial systems," *Journal of Electrical Systems*, 2024. [Online]. Available: https://www.researchgate.net/publication/386461421_Anomaly_Detection_in_Sensor_Data_with_Machine_Learning_Predictive_Maintenance_for_Industrial_Systems

[4] K. Lee, S. Kim, and H.-W. Kim, "Enhanced anomaly detection in manufacturing processes through hybrid deep learning techniques," *IEEE Access*, vol. PP, pp. 1–1, 2023. [Online]. Available: https://www.researchgate.net/publication/373418134_Enhanced_Anomaly_Detection_in_Manufacturing_Processes_through_Hybrid_Deep_Learning_Techniques

[5] A. Graß, C. Beecks, and J. A. Carvajal Soto, "Unsupervised anomaly detection in production lines," in *Machine Learning for Cyber Physical Systems.* Springer, 2019, pp. 17–25. [Online]. Available: https://www.researchgate.net/publication/329756428_Unsupervised_Anomaly_Detection_in_Production_Lines_Technologies_for_Intelligent_Automation

[6] G. R. Ginni and S. L. Chakravarthy, "A hybrid framework for robust anomaly detection: Integrating unsupervised and supervised learning with advanced feature engineering," *International Journal of Computational and Experimental Science and Engineering*, vol. 11, no. 2, Apr. 2025. [Online]. Available: https://www.ijcesen.com/index.php/ijcesen/article/view/1383

[7] W. Liu, L. Yan, N. Ma, G. Wang, X. Ma, P. Liu, and R. Tang, "Unsupervised deep anomaly detection for industrial multivariate time series data," *Applied Sciences*, vol. 14, no. 2, 2024. [Online]. Available: https://www.mdpi.com/2076-3417/14/2/774

[8] F. Mateo, J. Vila-Francés, E. Soria-Olivas, M. Martínez-Sober, J. Gómez-Sanchis, and A.-J. Serrano-López, "Dynamic classifier auditing by unsupervised anomaly detection methods: an application in packaging industry predictive maintenance," *arXiv preprint*, vol. arXiv:2405.11960, 2024, version v1, 20 May 2024. [Online]. Available: https://arxiv.org/abs/2405.11960