Research Article GU J Sci, Part A, 12(3): 858-872 (2025) 10.54287/gujsa.1655744

Gazi University
Journal of Science
PART A: ENGINEERING AND INNOVATION
http://dergipark.org.tr/gujsa

A Decision Tree-Based Approach for Classifying and Characterizing the Failure Type of White Goods

Çağlar SOLMAZ[1] Serhat PEKER[2*] Onur DOĞAN[2]

[1] Vestel Trade, Istanbul, Türkiye
[2] Izmir Bakircay University, Department of Management Information Systems, Izmir, Türkiye

Keywords

Data Mining
Classification
Decision Tree Analysis
Rule-Based Approach
Predictive Maintenance
Business Analytics

Abstract

After-sales service plays a vital role in the white goods industry, significantly affecting both customer satisfaction and operational performance. This paper presents a decision tree-based approach for classifying and characterizing failure types in white goods, using after-sales service data from a white goods manufacturer. We employ the Classification and Regression Tree (CART) algorithm to identify patterns in failure occurrences based on product category, region, usage duration, and brand. The model generates interpretable decision rules, providing insights into the factors contributing to failures. The results reveal that product category and region are the most significant factors influencing product failures. These findings support manufacturers and service providers in optimizing maintenance strategies and improving service operations. The proposed approach enhances decision-making processes in after-sales service, leading to higher customer satisfaction and extended product life cycles.

Cite

Solmaz, Ç., Peker, S., & Doğan, O. (2025). A Decision Tree-Based Approach for Classifying and Characterizing the Failure Type of White Goods. *GU J Sci, Part A*, *12*(3), 858-872. doi:10.54287/gujsa.1655744

Author ID (ORCID Number)

0000-0003-0031-1944	Çağlar SOLMAZ
0000-0002-6876-3982	Serhat PEKER
0000-0003-3543-4012	Onur DOĞAN

Article Process

Submission Date	12.03.2025
Revision Date	25.03.2025
Accepted Date	22.07.2025
Published Date	30.09.2025

1. INTRODUCTION

After-sales services encompass all technical support provided to customers post-purchase, including warranties, insurance, maintenance/repair, and product updates. According to Gartner's AMR research, businesses derive 45% of their gross profits and 24% of their revenues from the aftersales market (Gartner 1999). Furthermore, Cohen and Agrawal (2006) emphasize that customers do not necessarily demand flawless products; rather, they place greater value on timeliness and effectiveness of solutions to problems when they arise, thereby underscoring the pivotal role of after-sales services in maintaining customer satisfaction and loyalty. After-sales service also plays a crucial role in the white goods industry, directly influencing customer satisfaction, brand loyalty, and operational efficiency.

Traditional after-sales service models often rely on reactive maintenance strategies, where failures are addressed only after they occur. However, this approach can lead to increased operational costs, prolonged downtimes, and reduced customer satisfaction. A shift toward predictive and proactive maintenance can

*Corresponding Author, e-mail: serhat.peker@bakircay.edu.tr

This work is licensed under Creative Commons Attribution-ShareAlike 4.0 International.

significantly enhance service efficiency. By analyzing historical maintenance records, companies can anticipate potential issues, allocate appropriate resources, and reduce unnecessary service visits. Machine learning is widely used in various fields, such as retail, finance, banking, security, astronomy, and behavioral ecology (Kantardzic, 2003). Data mining and machine learning techniques have been successfully applied in after-sales service prediction and modeling, particularly in forecasting spare parts demand, optimizing service logistics, and identifying failure patterns (Ko et al., 2017; Rohaan et al., 2022; Liu et al., 2024).

In the white goods industry, common failure types include compressor malfunctions in refrigerators, motor issues in washing machines, and thermostat failures in ovens. These failures are broadly categorized as electrical (e.g., wiring faults, control board failures) or functional (e.g., mechanical wear, component degradation). Current industry practices predominantly rely on reactive maintenance, leading to delayed responses and higher costs. However, manufacturers are increasingly recognizing the potential of predictive analytics to transition toward proactive strategies tailored to product categories and regional usage patterns. For example, humidity levels in coastal regions may accelerate corrosion in electrical components, while heavy usage in urban households could lead to mechanical wear in washing machines.

Recent studies underscore notable gaps in the predictive maintenance literature for white goods. Unlike industrial machinery or critical infrastructure, household appliances have seen relatively sparse research attention in failure prediction. A primary challenge is the scarcity of domain-specific data. Papaioannou et al. (2024) highlights a significant lack of publicly available malfunction datasets for appliances, which hampers the development of effective predictive models. The limited availability of large, labeled datasets is cited as a major barrier to advancing appliance fault diagnosis. This data gap is only beginning to be addressed. For example, Fonseca, et al. (2023) recently released a real-world dataset of refrigerator and washing machine performance (current and vibration readings) to facilitate AI-driven predictive maintenance research. The creation of such datasets is crucial, as realistic data is “paramount” for training and validating predictive algorithms in this domain.

Another gap is the limited integration of IoT sensors in legacy appliances, which makes data collection difficult. Most in-service white goods were not built with smart monitoring in mind; as one study notes, “most of devices do not provide any IoT interface,” forcing researchers to gather data via external sensor retrofits (Ferreira, et al., 2021). This lack of built-in instrumentation has historically kept white goods largely excluded from predictive maintenance implementations, in contrast to heavily instrumented industrial equipment. Moreover, white goods present unique complexity. They combine mechanical systems (motors, compressors, pumps) with electronic controls. This dual nature means failures can arise from mechanical wear or electronic faults (or interactions of both), requiring specialized prediction approaches. Existing predictive maintenance studies often overlook this combination, focusing instead on single-domain systems. Together, these factors, such as data scarcity, low connectivity, and system complexity, explain why failure prediction for appliances remains underexplored. Bridging these gaps is an open research opportunity identified in recent literature.

This study aims to present a decision tree-based approach for classifying and characterizing failure types in white goods, focusing on both electrical and functional issues. Using real-world after-sales service data from a white goods manufacturer, we employ the CART (Classification and Regression Trees) algorithm to uncover patterns in failure occurrences based on factors such as product category, region, usage duration, and brand. The model generates interpretable decision rules that provide insights into the most critical factors contributing to failures. By applying the proposed approach to after-sales service data, companies can identify key factors influencing failure occurrences and improve predictive maintenance strategies. The contributions of this study are twofold. First, it provides a practical, rule-based framework that helps service personnel predict the type of failure before visiting a customer, leading to more efficient resource management and scheduling. Second, the insights derived from this model can aid manufacturers and service companies in optimizing their after-sales strategies by targeting specific product categories and regions, potentially enhancing customer satisfaction and brand reputation.

2. BACKGROUND

2.1. Knowledge Discovery in Databases and the Cross-Industry Standard Process for Data Mining

Knowledge Discovery in Databases (KDD) is a process for extracting useful and meaningful information from data (Witten et al., 2011). This method reveals hidden information in databases and provides significant business insights (Fayyad, 2001). Businesses can use this information in market analysis, product development, and customer segmentation. Moreover, data-driven decision-making enables enterprises to gain a competitive advantage. Especially with the increasing data intensity, the data mining process is crucial for businesses.

Data mining, a step in the KDD process, includes techniques widely used in various sectors, such as e-commerce, retail, services, banking, and healthcare. The data mining process is carried out using various methods, and the most commonly used methodology in literature is the Cross-Industry Standard Process for Data Mining (CRISP-DM) (Abbasi et al., 2016). This methodology comprises six stages for planning and executing data analytics and mining projects. These stages are business understanding, data understanding, data preparation, modeling, evaluation, and deployment (Chapman et al., 1999). In the business understanding stage, the project's goals and objectives are defined. In the data understanding stage, data sources and data quality are examined. In the data preparation stage, data cleaning and preprocessing are performed. Machine learning algorithms are applied to the data during the modeling stage. The evaluation stage assesses the model’s performance, and in the deployment stage, the resulting insights are presented to the end-users.

2.2. Decision Tree Algorithms

Classification algorithms are capable of processing large volumes of information. By leveraging categorical class labels and assumptions derived from training datasets, these algorithms enable systematic categorization of data and facilitate the classification of newly acquired data in a similar manner (Nikam, 2015). Among the various approaches, decision trees represent a particularly powerful and interpretable technique, widely applied across fields such as machine learning, image processing, and model identification (Stein et al., 2005). Decision trees efficiently and consistently combine a series of basic tests, where each test compares a numerical attribute against a threshold value (Damanik et al., 2019). The conceptual rules created by decision trees are much easier to interpret than the numerical weights in neural networks (Barros et al., 2012). Unlike black-box models such as deep learning, decision trees allow policy makers to extract human-readable insights that facilitate better decision-making. Decision trees are predominantly used for clustering purposes and are a frequently employed classification model in data mining (Gavankar & Sawarkar, 2017).

A decision tree consists of four fundamental structures: the root node, nodes, branches, and leaves. Each node represents the features to be classified within a category, and each subset defines a value the node can take (Mahesh, 2018). Due to their simplicity and accuracy in analyzing multiple data forms, decision trees have been employed in many different application areas (Mrva et al., 2019). In the context of white goods, decision trees are particularly advantageous due to their ability to handle categorical variables (e.g., product categories, regions) and numerical data (e.g., usage duration) simultaneously. This aligns with the heterogeneous nature of after-sales data, where factors like regional climate variations (e.g., coastal vs. arid regions) and product-specific usage patterns (e.g., frequent vs. intermittent use) must be analyzed cohesively.

2.3. Overview of Predictive Maintenance Research in Consumer Electronics and Home Appliances

Recent advancements in predictive maintenance have extended beyond industrial machinery into the realm of consumer electronics and household appliances. The reviewed studies demonstrate a growing interest in applying data-driven maintenance strategies to systems such as washing machines, refrigerators, boilers, and HVAC units. Table 1 presents a concise overview of selected peer-reviewed journal articles published within the last decade, focusing on predictive maintenance implementations in these consumer-facing domains.

The selected studies highlight that washing machines and residential heating systems are among the most investigated appliances. These systems are frequently used, subject to wear, and relatively accessible for data collection, which makes them suitable candidates for predictive modeling. In particular, appliances that exhibit both mechanical and electrical components have become central targets due to their higher risk of functional or electrical failure. This reflects an industry-wide shift toward using condition-based monitoring and proactive fault detection in everyday products.

A variety of technical approaches have been applied through the studies. Several contributions rely on Internet of Things (IoT) devices to collect real-time sensor data such as current, vibration, and temperature. This data is then processed using machine learning algorithms to predict faults before they occur. For example, Fonseca et al. (2023) introduced a high-resolution labeled dataset collected from actual appliance repair cases, enabling the development of accurate predictive maintenance models. Similarly, Fernandes et al. (2020) demonstrated how an IoT-based monitoring system can be used to forecast faults in domestic boilers, helping to avoid unnecessary energy waste and improve reliability. Other studies applied advanced deep learning techniques such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) to analyze smart plug data, as seen in the work of Casagrande et al. (2021) on washing machines. This method allows for non-intrusive monitoring, eliminating the need for internal sensors by interpreting energy consumption signatures. Furthermore, several papers employed data mining techniques, including sequential pattern recognition and association rules, to predict future maintenance activities based on historical service records. These methods offer practical solutions for optimizing maintenance schedules and reducing downtime.

One of the emerging challenges in predictive maintenance research is the scarcity of labeled fault data, especially for consumer appliances. To address this, recent work by Papaioannou et al. (2024) proposed a simulation-based approach to generate synthetic anomalies in appliance power consumption profiles. By injecting realistic malfunction signatures into normal usage data, the authors created training datasets for machine learning models, thereby overcoming the limitations of real-world data availability. This type of contribution is particularly important for scaling predictive maintenance applications in domains where failure data are difficult or costly to collect.

In addition to technical implementations, comprehensive review papers (e.g., Carvalho et al., 2019; Cinar et al., 2020) provide valuable insight into the broader landscape of predictive maintenance research. These reviews classify machine learning algorithms commonly used in predictive maintenance, identify application trends, and discuss ongoing challenges such as data imbalance, sensor integration, and model interpretability. They underscore the significance of combining low-cost sensor technology with data analytics to enable scalable maintenance solutions across different sectors, including smart homes and industrial systems.

Some studies, such as that of Falatouri et al. (2023), go a step further by considering spatial and temporal patterns in maintenance data across geographic regions. Their use of spatial–temporal networks demonstrate the potential of predictive analytics in optimizing large-scale service operations, such as technician dispatch and spare part logistics. Similarly, other works from national databases (e.g., TRDizin) illustrate practical deployments of machine learning models in real industrial settings, further supporting the feasibility of these approaches in both research and commercial environments.

3. METHODOLOGY

In this study, a systematic framework called the Cross-Industry Standard Process for Data Mining (CRISP-DM) was employed to develop a decision tree-based model for predicting failure types in white goods. The details are provided in the following subsections.

Table 1. *Summary of relevant predictive maintenance research*

Study	**Appliance/System**	**Method/Approach**	**Main Contribution/Findings**
Fonseca et al. (2023)	Refrigerators & Washing Machines	IoT sensors + Data Collection (current, vibration)	Provides a labeled dataset of current and vibration signals from real washer and fridge repair cases, supporting AI model training for appliance fault detection.
Es-sakali et al. (2022)	HVAC (Heating, Ventilation, A/C)	Literature Review (algorithms for HVAC predictive maintenance)	Systematically reviews HVAC predictive maintenance algorithms, emphasizing the importance of combining IoT sensors with data-driven methods for effective fault prediction.
Fernandes et al. (2020)	Home Boilers (Heating System)	IoT Data + ML Models (Failure Forecasting)	Develops an IoT-based system for predicting boiler faults using machine learning, reducing energy waste and unplanned maintenance.
Casagrande et al. (2021)	Washing Machines	Deep Learning (CNN & LSTM on power signals)	Uses deep learning on smart-plug data to detect internal faults in washing machines, enabling low-cost, sensor-free predictive maintenance.
Al-Refaie et al. (2023)	Washing Machine (maintenance records)	Data Mining (GSP sequences + Assoc. Rules)	Applies data mining to washing machine maintenance logs to predict upcoming maintenance needs, improving planning and uptime.
Papaioannou et al. (2024)	Various Home Appliances (fridge, washer, dryer, etc.)	Synthetic Data Generation (anomaly simulation)	Simulates power anomalies in appliances to generate synthetic training data, aiding development of predictive maintenance models.
Carvalho et al. (2019)	*Various Equipment* (survey)	Review (ML techniques in predictive maintenance)	Reviews 160+ studies applying ML in predictive maintenance, highlighting common methods, challenges, and their application across different equipment.
Cinar et al. (2020)	*Industrial Systems* (review)	Review (applications of ML in predictive maintenance)	Reviews machine learning-based predictive maintenance in Industry 4.0, stressing the role of connected sensors and AI in reducing equipment downtime.
Falatouri et al. (2023)	Home Heating Appliances (distributed)	Spatial–Temporal ML Model (demand forecast)	Uses spatial-temporal modeling to predict maintenance for heating appliances across locations, improving service logistics and accuracy.

3.1. Business Understanding

This stage involved defining the research objective and identifying key factors influencing failure types in household appliances. The goal was to develop a model that predicts whether a failure is functional or electrical, based on product category, region, usage duration, and brand. Understanding these patterns can help manufacturers optimize maintenance strategies and improve product reliability.

3.2. Data Understanding and Preparation

The dataset for this research was sourced from the maintenance and after-sales service records of a white goods manufacturer in Manisa, Turkey. Failure records from various white goods product categories were extracted from the CRM system, utilizing SAP HANA, covering the period between 2018 and 2022. The dataset comprises information on various appliances, and Table 2 outlines the selected key features related to product failures.

Table 2. *Feature descriptions*

Feature	**Description**
Product Category	Groups of white goods products
Brand	Brands of white goods products
Region	The region where customers submit requests
Usage Duration	Time in months between installation and breakdown
Failure Type	Type of failure occurring (Electrical or Functional)

During data preparation, the “Product Category” variable was categorized into eight groups, numbered 1 to 8, and the “Brand” variable was divided into six groups, numbered 1 to 6. The “Region” variable was created by dividing the provinces into seven geographical regions, numbered 1 to 7. The target class variable, “Failure Type,” consists of two classes: Electrical Failure (1) and Functional Failure (2).

Tables 3 and 4 summarize the distribution and statistics for the categorical and numerical features in the final dataset.

3.3. Implementation of CART algorithm

Decision trees are widely recognized for their balance between interpretability and predictive capability, making them suitable for practical, field-level decision-making. In this study, the Classification and Regression Tree (CART) algorithm was selected due to its ability to generate clear, interpretable rules, which are especially valuable for technicians and service managers who require transparent, actionable insights during after-sales service operations. For example, CART enables rules such as: "IF Product = Refrigerator AND Region = Black Sea, THEN Functional Failure", which directly align with the diagnostic logic used by field personnel.

GU J Sci, Part A 12(3) 858-872 (2025) 10.54287/gujsa.1655744

Table 3. *Distribution of Categorical Features*

Feature	**Value**	**Code**	**Distribution (%)**
Product Category	Washing Machine	1	31
	Dishwasher	2	22
	Refrigerator	3	20
	Air Conditioner	4	10
	Oven	5	6
	Deep Freezer	6	5
	Stove	7	4
	Dryer	8	2
Brand	VL	1	56
	RL	2	20
	SG	3	11
	FX	4	6
	VT	5	4
	WR	6	3
Region	Mediterranean	1	15
	Eastern Anatolia	2	9
	Aegean	3	13
	Southeastern Anatolia	4	16
	Central Anatolia	5	13
	Black Sea	6	15
	Marmara	7	19
Failure Type	Electrical Failure	1	48
	Functional Failure	2	52

Table 4. *Statistical Details of Numerical Variables*

Feature	Min	Max	Median	Mean	Std. Dev.
Usage Duration	0	84	24	27.97	21.60

Compared to black-box models such as Random Forest and Support Vector Machines (SVM), the CART algorithm provides direct interpretability through its rule-based structure (Breiman et al., 1984; Molnar, 2020), which is particularly valuable for real-time decision-making in operational settings (Rudin, 2019). Although other machine learning methods—such as Random Forest and SVM—may achieve higher predictive performance, they often require extensive hyperparameter tuning, involve greater implementation complexity, and lack inherent transparency (Hastie et al., 2009). Therefore, CART was favored for its transparency, ease of implementation, and its ability to generate intuitive rules that can be readily interpreted by non-technical personnel in the field. The trade-off in model complexity versus interpretability was carefully considered, with priority given to explainability to support practical application in after-sales service environments.

The CART algorithm was implemented using the R programming language in RStudio, utilizing the "rpart" library for model training. To enhance model robustness and avoid overfitting, hyperparameter tuning was performed, including optimization of the complexity parameter (CP) and depth of the tree.

To build the decision tree model, the dataset was divided into a 70:30 ratio, with 70% used for training and 30% for testing. The decision tree generates rules in the form of simple IF-THEN statements, where the IF part (also called the antecedent) represents a condition, and the THEN part provides the prediction. Two key metrics are used to evaluate the usefulness of these decision rules: support (or coverage) and accuracy (or confidence) (Molnar, 2010). Support refers to the percentage of instances where the condition of a rule applies, while accuracy measures how often the rule correctly predicts the target class for those instances. Typically, there is a trade-off between support and accuracy: adding more features to the condition can increase accuracy but may reduce support. To balance this, specific thresholds were set for both metrics, allowing us to select and interpret the most effective decision rules generated by the algorithm. Additionally, the feature importance scores were analyzed to determine the most influential factors in predicting failure types.

4. RESULTS

The rule-based representation of the decision tree model built using the CART algorithm is summarized in Table 5. Due to the large number of rules generated, we focused on rules with accuracy greater than 60% and support above 5%, resulting in a total of 9 rules. These selected rules provide insights into how different combinations of product category, region, and usage duration influence the failure type (either electrical or functional).

Table 5. *The rules generated by CART algorithm*

Rules			Accuracy (%)	Support (%)
Rule #	**IF**	**THEN-Failure Type**		
1	ProductCategory=1,2,3,4,6,8 =>	2	61	89
2	ProductCategory=1,2,3,4,6,8 AND Region=5,6 =>	2	71	25
3	ProductCategory=3,6 AND Region=1,2,3,4,7 =>	2	61	18
4	ProductCategory=3,6 AND Region=1,2,3,4,7 AND UsageDuration>=2 =>	2	62	16
5	ProductCategory=1,3 AND Region=5,6 =>	2	73	14
6	ProductCategory=2,4,6,8 AND Region=5,6 =>	2	69	11
7	ProductCategory is 5,7 =>	1	78	11
8	ProductCategory=2,6,8 AND Region=5,6 =>	2	70	9
9	ProductCategory=5,7 AND Region=1,2,3,4,7 =>	1	80	8

In Table 5, each rule is represented as an IF-THEN statement. The IF part consists of conditions based on one or more attributes from identified four features, while the THEN part predicts the type of failure (1 for Electrical and 2 for Functional). The table also reports the accuracy (confidence) and support (coverage) of each rule, indicating how well the rule applies to the dataset.

The analysis reveals that functional failures are the most common type of failure, particularly in washing machines, dishwashers, refrigerators, air conditioners, deep freezers, and dryers (Product Categories 1, 2, 3, 4, 6, and 8). Rule 1 highlights that these product types frequently experience functional failures, with an accuracy of 61% and a support of 89%, covering a significant portion of the dataset. The impact of geographical location is evident in Rule 2, which refines this pattern by showing that when these products are located in Central Anatolia (Region 5) or the Black Sea region (Region 6), the accuracy of predicting functional failures increases to 71%, though the rule applies to a smaller portion of the dataset (support: 25%). This suggests that environmental or regional factors may contribute to higher rates of functional issues in these areas.

Product-specific patterns further support these findings. Rule 3 indicates that refrigerators and deep freezers (Product Categories 3 and 6) in regions 1, 2, 3, 4, and 7 have a 61% likelihood of experiencing functional failures, with 18% support. When usage duration is greater than or equal to two months (Rule 4), the accuracy of predicting functional failure rises slightly to 62%, suggesting that prolonged usage may contribute to higher failure rates. Additionally, washing machines and refrigerators in regions 5 and 6 (Rule 5) exhibit the highest accuracy (73%) for functional failures, emphasizing the regional dependency in failure patterns.

On the other hand, electrical failures are strongly associated with ovens and stoves (Product Categories 5 and 7). Rule 7 shows that these products have a high likelihood of electrical failures (78% accuracy), regardless of location. This pattern becomes even more pronounced in Rule 9, where electrical failure rates increase to 80% when these products are in regions 1, 2, 3, 4, or 7, though this rule applies to a smaller portion of the dataset (support: 8%). These findings indicate that product type is a primary determinant of failure type, while regional differences serve as a secondary factor that refines predictions.

The relative importance of the features in the decision tree model is visualized in Figure 1, which highlights their contributions to the model's decision-making process. Essentially, variables that frequently appear in the decision-making process of the model are assigned higher scores, indicating their stronger influence in predicting the target variable (failure type).

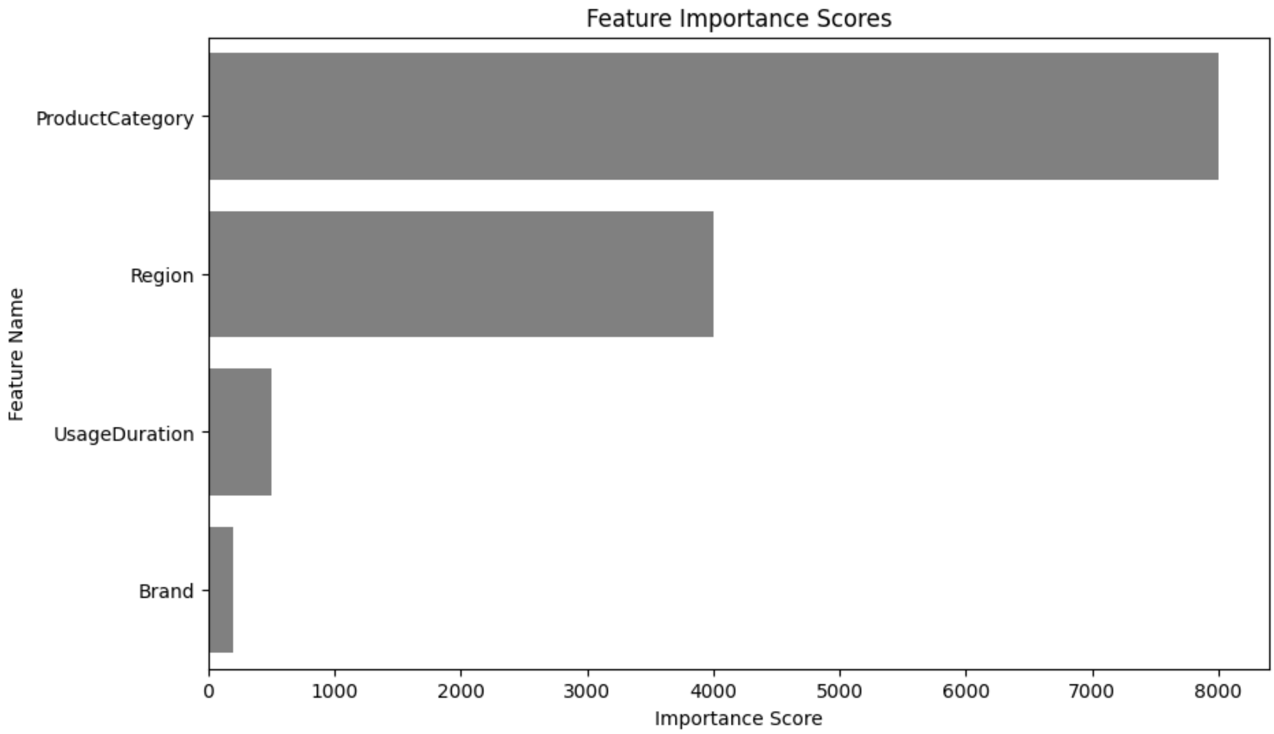

Figure 1. *Feature Importance scores*

Product category emerges as the most influential variable, with a score exceeding 8,000, reinforcing its dominant role in predicting failure type. This is expected, as product type directly determines the nature of failures—whether functional or electrical. The second most significant variable is region, with a score above 4,000, suggesting that geographical factors contribute notably to the failure patterns. The inclusion of regional factors in multiple rules further confirms that certain areas are more prone to specific types of failures.

In contrast, usage duration, with a score around 1,000, has a lower impact on failure prediction. While it appears in some rules (e.g., Rule 4), its influence is less significant compared to product type and region. Brand, with the lowest feature importance score, has minimal impact on failure classification, suggesting that failure type is not significantly affected by the manufacturer but rather by the product's category and location. These findings align with the rule-based results, confirming that product category and region are the primary factors influencing failure types, while other attributes play a lesser role.

5. CONCLUSION

This research focused on improving after-sales service in the white goods industry by developing a decision tree-based approach to predict failure types (electrical or functional) using the Classification and Regression Tree (CART) algorithm. By analyzing key factors such as product category, region, usage duration, and brand, a model was developed using historical failure data from a white goods manufacturer in Turkey. The CART model was employed to extract interpretable rules that aid in predicting failure types based on these input features, facilitating better decision-making for service personnel.

The key findings of this study indicate that product category and region are the most influential factors in predicting failure types, whereas usage duration and brand have relatively minor impacts. The decision tree model revealed that functional failures are more prevalent in products such as washing machines and refrigerators, particularly in Central Anatolia and Black Sea regions. Conversely, electrical failures are more commonly associated with ovens and stoves. These insights underscore the importance of tailoring maintenance strategies and failure prevention policies based on both product type and geographical location, enabling more targeted and effective interventions in service management.

The practical implications of this study are significant for after-sales service operations in the white goods industry. By implementing the decision tree model, service centers can predict the type of failure (electrical or functional) before dispatching technicians, enabling more efficient workforce allocation and reducing unnecessary site visits. This leads to improved scheduling, better resource utilization, and a higher likelihood of first-time fix rates. Additionally, by identifying patterns related to product type and region, companies can tailor their maintenance strategies and training programs, ultimately enhancing service quality, reducing costs, and boosting customer satisfaction.

Despite the valuable insights gained from this study, a couple of limitations should be acknowledged. First, the analysis relies on a specific dataset, which may not fully represent all brands, product categories, or regional variations in failure patterns. Future research could expand the dataset to include a more diverse range of products and geographical regions to enhance generalizability. Further, this study focuses on key variables such as product category, region, usage duration, and brand. In future studies, it may be useful to incorporate additional contextual factors to provide a more comprehensive understanding of failure patterns.

AUTHOR CONTRIBUTIONS

Conceptualization, Ç.S. and S.P.; methodology, Ç.S. and S.P.; fieldwork and data preparation, Ç.S.; software and implementation, Ç.S.; investigation and literature review, Ç.S., S.P. and O.D.; manuscript-original draft, Ç.S., S.P. and O.D.; manuscript-review and editing, S.P. and O.D.; visualization, Ç.S. and S.P.; supervision, S.P.; project management, S.P. All authors have read and legally accepted the final version of the article published in the journal.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Abbasi, A., Sarker, S., & Chiang, R. H. L. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the Association for Information Systems*, *17*(2), 3. https://doi.org/10.17705/1jais.00423

Al-Refaie, A., Abu Hamdieh, B., & Lepkova, N. (2023). Prediction of maintenance activities using generalized sequential pattern and association rules in data mining. *Buildings*, *13*(4), 946. https://doi.org/10.3390/buildings13040946

Barros, R. C., Basgalupp, M. P., De Carvalho, A. C. P. L. F., & Freitas, A. A. (2012). A survey of evolutionary algorithms for decision-tree induction. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *42*(3), 291–312. https://doi.org/10.1109/TSMCC.2011.2157494

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. CRC Press.

Carvalho, T. P., Soares, F. A., Vita, R., Francisco, R. D. P., Basto, J. P., & Alcalá, S. G. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, *137*, 106024. https://doi.org/10.1016/j.cie.2019.106024

Casagrande, V., Fenu, G., Pellegrino, F. A., Pin, G., Salvato, E., & Zorzenon, D. (2021). Machine learning for computationally efficient electrical loads estimation in consumer washing machines. *Neural Computing and Applications*, *33*(22), 15159-15170. https://doi.org/10.1007/s00521-021-06138-9

Chapman P., Clinton J., Kerber R., Khabaza T., Reinartz T., & Shearer C., & W. (1999). Guía paso a paso de Minería de Datos CRISP-DM 1.0.

Cinar, Z. M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. *Sustainability*, *12*(19), 8211. https://doi.org/10.3390/su12198211

Cohen, M. A., Agrawal, N., & Agrawal, V. (2006). Winning in the aftermarket. *Harvard business review*, *84*(5), 129.

Damanik, I. S., Windarto, A. P., Wanto, A., Poningsih, Andani, S. R., & Saputra, W. (2019). Decision Tree Optimization in C4.5 Algorithm Using Genetic Algorithm. *Journal of Physics: Conference Series*, *1255*(1), 012012. https://doi.org/10.1088/1742-6596/1255/1/012012

Es-sakali, N., Cherkaoui, M., Mghazli, M. O., & Naimi, Z. (2022). Review of predictive maintenance algorithms applied to HVAC systems. *Energy Reports*, *8*, 1003-1012. https://doi.org/10.1016/j.egyr.2022.07.130

Falatouri, T., Brandtner, P., Nasseri, M., & Darbanian, F. (2023). Maintenance forecasting model for geographically distributed home appliances using spatial-temporal networks. *Procedia Computer Science*, *219*, 495-503. https://doi.org/10.1016/j.procs.2023.01.317

Fayyad, U. (2001). Knowledge Discovery in Databases: An Overview. *Relational Data Mining*, 28–47. https://doi.org/10.1007/978-3-662-04599-2_2

Fernandes, S., Antunes, M., Santiago, A. R., Barraca, J. P., Gomes, D., & Aguiar, R. L. (2020). Forecasting appliances failures: A machine-learning approach to predictive maintenance. *Information*, *11*(4), 208. https://doi.org/10.3390/info11040208

Ferreira, L. L., Oliveira, A., Teixeira, N., Bulut, B., Landeck, J., Morgado, N., & Sousa, O. (2021, October). *Predictive maintenance of home appliances: Focus on washing machines*. In: IECON 2021–47th Annual Conference of the IEEE Industrial Electronics Society (pp. 1-6). IEEE.

Fonseca, T., Chaves, P., Ferreira, L. L., Gouveia, N., Costa, D., Oliveira, A., & Landeck, J. (2023). Dataset for identifying maintenance needs of home appliances using artificial intelligence. *Data in Brief*, *48*, 109068. https://doi.org/10.1016/j.dib.2023.109068

Gartner. (1999). Gartner AMR Research Report.

Gavankar, S. S., & Sawarkar, S. D. (2017, April 07-09). *Eager decision tree*. In: 2nd International Conference for Convergence in Technology, I2CT, (pp. 837–840). https://doi.org/10.1109/I2CT.2017.8226246

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Kantardzic, M. (2003). Data Mining: Concepts, Models, Methods, and Algorithms. (2003). *Technometrics*, *45*(3), 277. https://doi.org/10.1198/tech.2003.s785

Ko, T., Hyuk Lee, J., Cho, H., Cho, S., Lee, W., & Lee, M. (2017). Machine learning-based anomaly detection via integration of manufacturing, inspection and after-sales service data. *Industrial Management and Data Systems*, *117*(5), 927–945. https://doi.org/10.1108/IMDS-06-2016-0195/FULL/XML

Liu, Y., Hu, S., Zhang, H., Dong, Q., & Liu, W. (2024). Intelligent mining methodology of product field failure data by fusing deep learning and association rules for after-sales service text. *Engineering Applications of Artificial Intelligence*, *133*, 108303. https://doi.org/10.1016/J.ENGAPPAI.2024.108303

Mahesh, B. (2018). Machine Learning Algorithms-A Review. *International Journal of Science and Research*. https://doi.org/10.21275/ART20203995

Molnar, C. (2020). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). Lean Publishing.

Mrva, J., Neupauer, S., Hudec, L., Sevcech, J., & Kapec, P. (2019, November 21-23). *Decision support in medical data using 3D decision tree visualisation*. In: 7th E-Health and Bioengineering Conference (EHB). https://doi.org/10.1109/EHB47216.2019.8969926

Papaioannou, A., Dimara, A., Papaioannou, C., Papaioannou, I., Krinidis, S., Anagnostopoulos, C.-N., Korkas, C., Kosmatopoulos, E., Ioannidis, D., & Tzovaras, D. (2024). Simulation of Malfunctions in Home Appliances' Power Consumption. *Energies* (19961073), *17*(17). https://doi.org/10.3390/en17174529

Rohaan, D., Topan, E., & Groothuis-Oudshoorn, C. G. M. (2022). Using supervised machine learning for B2B sales forecasting: A case study of spare parts sales forecasting at an after-sales service provider. *Expert Systems with Applications*, *188*, 115925. https://doi.org/10.1016/J.ESWA.2021.115925

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Nikam S. S. (2015). A Comparative Study of Classification Techniques in Data Mining Algorithms | Oriental *Journal of Computer Science and Technology*. http://www.computerscijournal.org/?p=1592

Stein, G., Chen, B., Wu, A. S., & Hua, K. A. (2005). *Decision tree classifier for network intrusion detection with GA-based feature selection*. In: Proceedings of the 43rd annual ACM Southeast Conference (vol.2) (pp. 2136–2141). https://doi.org/10.1145/1167253.1167288

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*, (3rd ed.). Elsevier. https://doi.org/10.1016/C2009-0-19715-5