Research Article

International Journal of Data Science and Applications (JOINDATA) 7(1), 48-63, 2024

Received: 10-Oct-2024 Accepted: 21-Nov-2024

homepage: https://dergipark.org.tr/tr/pub/joindata

SAKARYA UYGULAMALI BİLİMLER ÜNİVERSİTESİ

Optimizing Energy Efficiency in Wood Processing Plants through Data Analysis: A Case Study on Value-Added Wood Products Manufacturing

Melike Nur İnce[1], Çağatay Taşdemir[2, 3, *]

[1] Department of Industrial Engineering (Grad. School), Bursa Technical University, Bursa 16310, Türkiye; 22435025008@ogrenci.btu.edu.tr

[2] Department of Industrial Engineering (Grad. School), Bursa Technical University, Bursa 16310, Türkiye; cagatay.tasdemir@btu.edu.tr

[3 *] Department of Forest Industry Engineering, Bursa Technical University, Bursa, Türkiye *Correspondence: Bursa Technical University, Bursa, Türkiye, 22435025008@ogrenci.btu.edu.tr and cagatay.tasdemir@btu.edu.tr

ABSTRACT

Energy consumption in value-added wood products manufacturing facilities has significant environmental and economic impacts. High energy usage increases costs and expands the carbon footprint, making it challenging to achieve sustainability goals. Inefficient energy management in wood processing plants elevates operational costs and exacerbates the environmental burden. Therefore, optimizing energy efficiency through data analysis techniques is critically important. This study analyzes energy consumption data to identify inefficiencies and propose effective optimization strategies. Historical data encompassing operational parameters, energy consumption, environmental conditions, and production output from five high-capacity wood processing machines in the wood products industry were collected daily over the past three years. The dataset includes ten categories: Date, Machine ID, Runtime Hours, Load Percentage, Electricity Usage, Gas Usage, Temperature, Humidity, Production Output, and Energy Efficiency. Initially, the data were loaded into a pandas Data Frame, converted to date and time format, and checked for missing and outlier values, followed by standardization of numerical features. Descriptive statistics were calculated for each feature, and data consistency was verified. The distributions of critical features were visualized with histograms, and the relationships between numerical features were illustrated using a correlation matrix heatmap. Trends and seasonal patterns in energy consumption and production output were analyzed by resampling the data monthly. Principal Component Analysis (PCA) was applied to reduce the dimensionality of the dataset while retaining significant information, and three clusters were formed using the K-Means algorithm. The clusters were visualized in the PCA-reduced feature space, and their characteristics were analyzed to prioritize machines for energy efficiency improvements. Cluster 2, characterized by an average energy usage of 209.79 kWh, an average gas usage of 107.90 m^3, an average production output of 1595.03 units, an average energy efficiency of 5.26 units/kWh, and an average load percentage of 75.35%, demonstrated low energy consumption and high production output, indicating highly efficient operations. Therefore, it is recommended that the best practices from this cluster be standardized and implemented across other clusters. Additionally, investing in technological advancements to enhance energy efficiency and conducting continuous improvement efforts to maintain and improve efficiency are suggested.

Keywords: Energy Efficiency, K-Means Clustering, Principal Component Analysis (PCA), Wood Products Industry .

1 Introduction

Furniture is a product that has existed throughout history and has become an integral part of our daily lives, with a continued presence in the future. As a consequence of the growth in production activities, the size of the furniture market has continued to expand from the past to the present. Several studies have been conducted to understand the market size. In one such study, the size of the global furniture market is estimated to be 579 billion US dollars in 2023. This figure is expected to increase annually, reaching approximately 873 billion US dollars by 2030 [1]. In a separate investigation, the global furniture market was evaluated at USD 516.66 billion in 2022. It is projected to experience growth from USD 541.52 billion in 2023 to USD 780.43 billion by 2030, exhibiting a compound annual growth rate (CAGR) of 5.36% over the projected period [2].

The wood furniture industry represents a significant contributor to the global growth potential of the furniture industry as a whole. Recent studies indicate that the market value of wood furniture in 2023 is estimated at USD 569.9 billion. It is projected to reach approximately USD 877.7 billion by 2032, representing a compound annual growth rate (CAGR) of 5% [3]. In a separate study, the global market for wood furniture reached a valuation of USD 275.7 billion in 2023 and is projected to reach USD 413.5 billion by 2032. These forecasts indicate a compound annual growth rate (CAGR) of 4.47% between 2024 and 2032 [4]. The two studies concur that the wood furniture sector is experiencing growth. This growth will consequently result in increased energy demand for the industry, necessitating the utilization of more energy for the machinery and equipment employed.

In this context, the issue of energy efficiency is becoming increasingly significant, particularly in light of the rising energy costs. In the long term, the rise in energy costs associated with production has rendered energy efficiency a necessity and a competitive factor in global markets. [5]. Because the manufacturing sector is responsible for 23% of global energy consumption, the combination of heightened environmental awareness, rising energy costs, and increasingly stringent regulations is compelling organizations to pursue operational excellence and effective management of energy efficiency in their production systems [6]. In consequence, enterprises are implementing a variety of energy efficiency measures (EEMs), including process technology-centric improvements (e.g., optimization of process parameters) and investment-oriented improvements (e.g., selection of energy-efficient machinery) [6]. A substantial body of research on energy efficiency has been conducted in the literature. In these studies, a variety of methods have been employed, including simulation-based approaches [7], data analysis studies [8, 9], production planning and hybrid simulation [5], and lean management [6].

Data analysis examines, cleans, transforms, and models data to achieve a specific goal or solve a problem. In this process, the raw data is transformed into meaningful information, thereby providing a basis for future decisions [10]. The objective of data analysis is to identify patterns, trends, and relationships within data sets through the application of diverse techniques and tools. Data analysis is widely utilized across numerous fields, including finance, marketing, health, and engineering. Businesses, organizations, and individuals can enhance their decision-making processes by employing data analysis, making them more informed and strategic [11].

2 Literature Review

Wold et al. (1987) made significant contributions to the literature on principal component analysis (PCA), focusing on issues such as the geometric interpretation of PCA, variable selection, correspondence analysis, and modeling power. PCA serves as a fundamental technique in multivariate data analysis. The article emphasizes that PCA is a versatile tool that can facilitate understanding complex datasets. However, it also cautions that the method should be applied with caution. This paper comprehensively examines PCA's theoretical underpinnings and practical applications [12].

Ghiasi et al. (2002) presented a comprehensive theoretical framework for energy-aware clustering in sensor networks. Their work addressed the energy optimization and balanced k-clustering problem. The authors develop algorithms that optimize the load distribution among the master nodes to improve energy efficiency and demonstrate the effectiveness of these algorithms through preliminary experiments. Furthermore, they make significant contributions to the existing literature, indicating potential avenues for future research that can be adapted to the dynamic nature of sensor networks [13].

Abdi and Williams (2010) significantly contribute to the literature on PCA by comprehensively understanding the subject matter. They concentrate on issues such as calculating observation contributions in PCA, the function of eigenvalues and factor scores, the interpretation of components, and data visualization. Furthermore, it enhances PCA's theoretical and practical aspects by addressing practical applications in the social sciences and generalizations, such as including qualitative variables [14].

Bro and Smilde (2014) explored principal component analysis (PCA) and emphasized its applicability, particularly in chemometrics. By providing a comprehensive description of PCA, they offer researchers guidance on critical issues such as component selection, outlier detection, and model validation. By emphasizing PCA's versatility and pragmatic utility in discovery studies, the article offers valuable insights into the existing literature from theoretical and applied viewpoints [15].

Benedetti et al. (2016) significantly contributed to the existing literature on industrial energy efficiency. The article underscored the significant role of compressed air systems in industrial energy consumption and the deficiencies in the prevailing measurement methodologies. In particular, the study identified shortcomings in current practices, including the low proportion of facilities measuring energy consumption and the fact that current benchmarks do not reflect actual operating conditions. The paper establishes a foundation for future research to enhance industrial energy management. It illustrates the necessity for enhanced energy metering systems and the establishment of dependable benchmarks [9].

Zhou et al. (2017) made a notable contribution to the existing energy efficiency and sustainability literature by presenting innovative methodologies. The study presents findings that are supported by empirical data and have reached a wider audience through the use of open access. The study underscored the potential of sustainable energy solutions to inform and enhance energy policies and bolster energy efficiency. Furthermore, it offers policy recommendations to facilitate the adoption of energy technologies and highlights potential avenues for future research. The study makes a valuable contribution to the field of energy studies by promoting collaboration and knowledge sharing, which are essential for overcoming the challenges currently facing the energy sector [16].

Chen and Xiong (2017) contributed to the district heating system adaptation to climate change and building renovation scenarios. The study assesses the practicality of the function utilized for heat demand estimation and presents application and scenario-based analyses with real-world data. The findings elucidate the impacts of diverse climate and renovation scenarios on heat demand, underscoring the significance of parameter tuning to enhance the precision of predictions. This indicates the necessity for precise forecasting methodologies to administer energy resources effectively [17].

Sobottka et al. (2017) notably impacted the production planning and control (PPC) research stream by introducing a novel hybrid simulation-based planning tool, potentially enhancing energy efficiency in manufacturing processes. The tool optimizes energy use by integrating material flow and equipment behavior with multi-criteria optimization, supporting sustainable manufacturing practices. Furthermore, incorporating the genetic algorithm (GA) into the optimization process has facilitated the simultaneous attainment of multiple production objectives and the formulation of constraints that reduce the search space. The paper enhances these methods' applicability to various production environments while proposing new avenues for future research. These contributions provide a foundation for future developments in energy-efficient production planning and a crucial basis for the sustainability of modern manufacturing industries [7].

Porizka et al. (2018) offer significant insights into the application of principal component analysis (PCA) in the processing of laser-induced breakdown spectroscopy (LIBS) data. In light of the growing popularity of PCA in LIBS, the authors provide practical guidelines for its use in data visualization, dimensionality reduction, and modeling. In particular, the paper identifies the challenges associated with implementing data preprocessing and MVDA algorithms. It underscores the necessity to develop more comprehensive data libraries for the LIBS community. Furthermore, the paper asserts that PCA can address numerous analytical requirements and, as a result, should be considered as a preliminary step before transitioning to more intricate algorithms. In conclusion, the paper offers methodological contributions to the existing literature and provides detailed application guidelines by exploring the role of PCA in LIBS in depth [18].

In their study, Bonfa et al. (2019) significantly contributed to advancing energy efficiency in district heating systems. The paper put forth a novel forecasting methodology based on the correlation between outdoor temperature and heat demand and demonstrated its application to a range of buildings in Lisbon. The findings indicated that retrofit scenarios impact the accuracy of the predictions and that alterations in the slope of the heat demand function are crucial for long-term planning. This study underscored the significance of more precise and responsive forecasts in energy planning and strategies to mitigate greenhouse gas emissions [8].

In their study, Cebeci (2020) introduces the R package 'fcvalid,' which significantly contributes to the field of fuzzy and probabilistic clustering analysis. The package offers a comprehensive tool for validating clustering results and provides a user-friendly experience with generalized internal validity indices. Furthermore, the package specializes in fuzzy clustering needs, allowing data points to belong to more than one cluster. The graphical review and performance evaluation functions of 'Fcvalid' represent a valuable tool for researchers interested in improving clustering methodologies, and the package makes an essential contribution to the existing literature on the subject [19].

The study by Clayman et al. (2020) represents a significant advancement in bioinformatics and gene expression analysis. The study enhanced the efficacy of clustering in highly variable datasets by integrating principal component analysis (PCA) and K-means clustering methods to examine 978 landmark genes within the L1000 dataset. The utilization of landmark genes resulted in the formation of more discrete clusters and enhanced clustering efficiency compared to randomly selected genes. Moreover, the paper offers valuable insights for future research on the potential clinical applications of these genes, underscoring the necessity for a more comprehensive investigation into the interrelationships between genetic and clinical variables [20].

In a recent study, Jansson et al. (2022) have made significant contributions to geochemical exploration techniques, focusing on Zn skarn deposits and industrial carbonates in the Sala region of Sweden. The integrated application of PCA and K-means clustering techniques enabled an objective classification of dolomite samples while facilitating the comprehension of the spatial distribution of mineralization through geochemical signatures. The findings demonstrate that data-driven approaches yield strong concordance with geological correlations, thereby enabling the identification of dolomite types with economically significant potential for mining. This study makes a valuable contribution to the existing literature on geological investigations by emphasizing the significance and practical implications of data-driven classification methods [21].

In a recent study, Ilu et al. (2022) proposed a novel methodology for forecasting the spread of the novel coronavirus (2019-nCoV), also known as SARS-CoV-2, which has caused a global pandemic. By integrating the LSTM algorithm with PCA and K-means clustering, the study demonstrated remarkable proficiency in nonlinear data processing and outlier detection. Compared to other machine learning algorithms, it exhibited the superior performance of LSTM and achieved high prediction accuracy on large datasets. The findings offer significant guidance for enhancing public health predictions and further investigating the applications of machine learning in epidemiological modeling [22].

Liao et al. (2024) proposed a two-layer optimization method for DPV and ESS based on IDEC-K clustering to reduce voltage violations and network losses in distribution networks. The results demonstrate that this method effectively mitigates overvoltage risk and reduces line losses by narrowing the voltage range and stabilizing fluctuations. This paper makes a significant contribution to the existing literature on energy management by emphasizing the crucial role of advanced optimization techniques in this field [23].

The study by Eid et al. (2024) on the determination of recharge and salinity origins in Siwa Oasis, Egypt, offers significant insights that contribute to the existing literature on water resources management. By integrating stable isotopes, mixing models, and K-mean cluster analysis, the study provides a detailed analysis of the types of water resources and the dynamics of increasing salinity in the region. This comprehensive approach elucidates how the salinity of the TCA aquifer has increased over time and the extent to which various factors have contributed to this phenomenon. It also furnishes recommendations for water management strategies. Moreover, the study offers a crucial reference point for future research endeavors to address analogous hydrological challenges in arid regions [24].

In a recent contribution to the field, Dugan et al. (2024) provide valuable insights into assessing foot deformities in children with cerebral palsy (CP). The study employs a multivariate functional principal component analysis (MFPCA) in conjunction with a k-means clustering method to facilitate a more detailed and dynamic classification of foot types. The findings underscore the necessity of developing personalized treatment plans and the importance of dynamic assessment methods. The efficacy of MFPCA in clinical trials signifies a substantial advancement in the management of foot deformities and offers a robust foundation for future research endeavors [25].

A literature review reveals that PCA and the K-Means algorithm have been employed in numerous studies on energy efficiency. However, there is a lack of research investigating the application of these methods to enhance energy efficiency in the wood furniture industry. In this context, the objective of this study is to utilize PCA and the K-Means algorithm to enhance the energy efficiency of machinery used in facilities that produce value-added wood products. High energy consumption presents a significant challenge to achieving sustainability goals, resulting in increased costs and a larger carbon footprint. Furthermore, inefficient energy management contributes to elevated operational costs and an increased environmental burden. This study aims to enhance energy efficiency and achieve cost savings by analyzing the energy consumption data of a wood products manufacturing company, identifying existing inefficiencies, and proposing effective optimization strategies.

3 Material And Methods

EcoWoodProduct is a medium-sized enterprise specializing in producing value-added wood products, with operations spanning domestic and international markets. Additionally, EcoWoodProduct is located in Türkiye and is renowned for its superior quality and innovative products within the woodworking industry. The company employs modern production technologies and an environmentally conscious approach to increase energy efficiency and optimize production processes per its sustainability goals.

The study is comprised of four phases. The data collection and preprocessing phase was followed by exploratory data analysis, cluster analysis, and cluster analysis.

In the initial phase of the study, data about operational parameters, energy consumption, environmental conditions, and production output from five high-capacity wood processing machines utilized in the wood processing industry were gathered daily over the past three years. The data set was then divided into ten categories:

- Date: Date of the data record.
- Machine ID: Unique ID for each machine.
- Runtime Hours: Total runtime of the machine (in hours).
- Load Percentage: Load percentage according to the machine's maximum capacity.
- Electricity Usage: Electricity consumption (in kWh).
- Gas Usage: Natural gas consumption (in cubic meters).
- Temperature: Temperature in the plant (in degrees Celsius).
- Humidity: Humidity level in the plant (in percent).
- Production Output: Production output (in units).
- Energy Efficiency: Energy efficiency metric (production output per energy consumed).

Table 1 presents a summary of the statistical characteristics of the data set utilized in the study.

The data were loaded into a Pandas DataFrame, converted to date and time format, and subjected to quality control checks, including identifying missing values and outliers. Subsequently, the numerical features were standardized.

In the study's second phase, exploratory data analysis was conducted to identify the dataset's characteristics. In this phase, central tendencies and variability for each feature in the entire dataset were calculated with summary statistics (mean, standard deviation, minimum, maximum, etc.). Additionally, data consistency was evaluated by identifying any missing values. The key features were identified, and histogram plots were constructed to illustrate their distribution. Furthermore, correlation matrices were employed to elucidate the interrelationships between the numerical attributes within the data sets. The data were resampled monthly to identify trends and seasonal patterns in energy consumption and production output, and average values were calculated.

In the third stage of the study, a cluster analysis was conducted. In this stage, the data sets were organized by selecting six features for analysis: runtime hours, load percentage, electricity usage, gas usage, temperature, humidity, production output, and energy efficiency. The selected features were standardized using the Standard Scaler. To reduce the dimensionality of the data set and retain the most meaningful information, a principal component analysis (PCA) was conducted, reducing the data set to two principal components. PCA is a dimension reduction method that aims to express a large number of correlated variables in a data set with a smaller number of new independent variables, which are known as principal components [26]. This method employs the correlation between the original variables to generate new components that collectively encompass the maximum possible variance in the data set. Each new component is calculated as a linear combination of the original variables, and these components are uncorrelated [27]. PCA represents data in a more concise form by minimizing the loss of information, particularly in multivariate data sets, and is therefore widely used for data reduction, modeling, and identifying significant patterns among variables [26].

Table 1. Statistical Summary of Operational Data for Woodworking Machinery

Variable	**Machine ID**	**N**	**N***	**Mean**	**SE Mean**	**StDev**	**Min**	**Median**	**Max**
Runtime Hours (h)	M01	1096	0	16.829	0.123	4.060	10.013	16.927	23.995
	M02	1096	0	17.035	0.122	4.031	10.001	16.819	23.987
	M03	1096	0	17.017	0.121	4.012	10.016	16.766	23.992
	M04	1096	0	17.187	0.122	4.035	10.006	17.333	23.992
	M05	1096	0	17.095	0.118	3.921	10.030	16.915	23.999
Load (%)	M01	1096	0	75.081	0.429	14.194	50.080	75.317	99.998
	M02	1096	0	74.775	0.446	14.777	50.031	74.792	99.999
	M03	1096	0	74.314	0.431	14.279	50.002	73.819	99.881
	M04	1096	0	74.314	0.437	14.457	50.023	74.226	99.994
	M05	1096	0	74.755	0.433	14.338	50.008	74.500	99.903
Electricity Usage (kWh)	M01	1096	0	300.24	3.46	114.47	100.09	298.85	498.92
	M02	1096	0	294.27	3.51	116.31	101.01	290.74	499.37
	M03	1096	0	295.89	3.51	116.11	100.05	295.20	499.36
	M04	1096	0	304.17	3.48	115.06	101.03	304.53	499.61
	M05	1096	0	300.10	3.53	116.82	100.82	303.96	499.86
Gas Usage (m^3)	M01	1096	0	127.08	1.33	44.02	50.16	129.51	199.79
	M02	1096	0	126.19	1.29	42.69	50.03	127.69	199.96
	M03	1096	0	123.30	1.29	42.63	50.40	121.68	200.00
	M04	1096	0	120.96	1.28	42.44	50.27	117.72	199.87
	M05	1096	0	123.96	1.28	42.39	50.01	123.71	199.82
Temperature (^oC)	M01	1096	0	24.959	0.0869	2.875	20.007	25.007	29.980
	M02	1096	0	24.879	0.0887	2.936	20.005	24.746	29.992
	M03	1096	0	25.029	0.0873	2.891	20.000	25.118	29.990
	M04	1096	0	24.997	0.0892	2.953	20.003	24.977	29.995
	M05	1096	0	24.788	0.0875	2.897	20.004	24.758	29.994
Humidity (%)	M01	1096	0	49.892	0.349	11.569	30.013	49.884	69.983
	M02	1096	0	49.581	0.350	11.583	30.024	49.749	69.988
	M03	1096	0	50.254	0.353	11.680	30.037	50.048	69.987
	M04	1096	0	49.750	0.346	11.454	30.040	49.683	69.990
	M05	1096	0	49.638	0.351	11.616	30.025	49.772	69.926
Production Output (Units)	M01	1096	0	1250.7	13.3	438.7	501.4	1240.2	1996.9
	M02	1096	0	1248.5	13.1	434.6	501.2	1250.8	1999.0
	M03	1096	0	1267.3	13.3	441.2	500.1	1264.0	1998.6
	M04	1096	0	1250.1	13.0	431.2	503.0	1269.5	1998.9
	M05	1096	0	1226.1	13.1	432.8	500.3	1217.7	1999.6
Energy Efficiency (Production Output Per kWh)	M01	1096	0	3.2114	0.0470	1.5571	0.7965	2.9332	9.8057
	M02	1096	0	3.2758	0.0501	1.6596	0.8647	2.9291	10.1497
	M03	1096	0	3.3728	0.0530	1.7536	0.8234	3.0222	11.3280
	M04	1096	0	3.2491	0.0491	1.6249	0.7948	2.9205	11.1761
	M05	1096	0	3.2067	0.0490	1.6238	0.7773	2.9295	11.8756

A K-means clustering algorithm with three clusters was then applied to the reduced data set using PCA, and the resulting cluster labels were added to the original data frame. The clusters were then delineated in the feature space that PCA had reduced to facilitate an examination of the cluster separations. The K-means algorithm is designed to partition a given data set into k clusters, with each sample assigned to the cluster with which it is most closely associated. Even without an a priori determination of the optimal number of clusters, the optimal value of k can be identified through cluster adjustment. The optimization of clusters is achieved by minimizing the intra-cluster variance or squared error function [28]. The operation of the algorithm is illustrated in Figure 1.

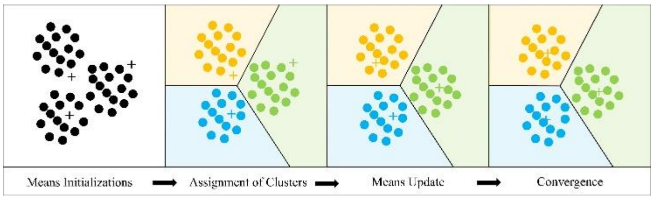

Fig. 1. K-Means Algorithms [29].

The algorithm is comprised of the following steps.

1. The data set is partitioned into k groups.
2. The center of each cluster is initially selected at random.
3. Subsequently, the distance of each object to the selected centers is calculated, and all objects are allocated to the cluster with the closest center among the k clusters.
4. The mean or centroid of all instances within each cluster is calculated.
5. These steps are repeated until each sample has been assigned to a cluster.

In the final stage of the analysis, the clusters were examined. During these analyses, the characteristics of each cluster were identified, graphical representations were constructed to illustrate the distribution of salient features within each cluster, and areas for operational improvement were identified.

4 Results

A three-year data set comprising ten different data types was collected daily to evaluate the energy efficiency of five high-capacity machines used in the wood processing industry. The dataset was initially loaded into a Pandas DataFrame, and the "Date" column was converted to the date and time format. The dataset was then cleaned, whereby any missing or abnormal data points were identified and removed. Subsequently, the numerical features were standardized to ensure they were expressed on a consistent scale.

Statistical summaries, including the mean, standard deviation, minimum, and maximum values, were calculated to ascertain each feature's central tendencies and variability within the dataset. These computations identified and rectified any missing values and inconsistencies in the data.

A correlation matrix was constructed to analyze the dataset further and understand the data points' relationships. The resulting correlation matrix is presented in Figure 2.

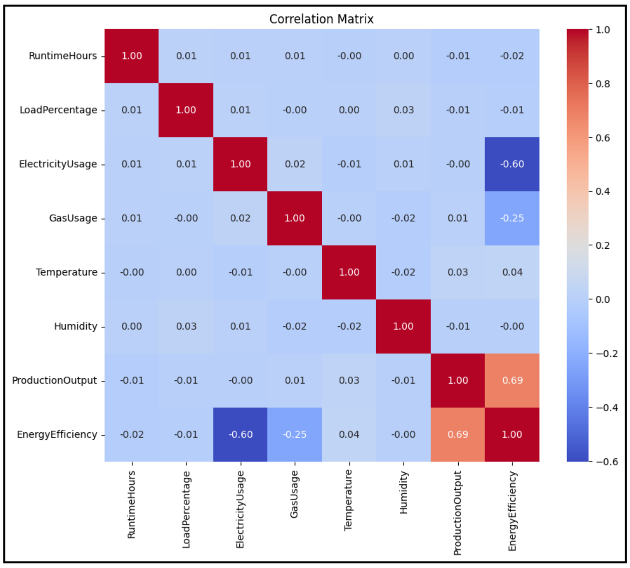

Fig. 2. Correlation Matrix

As illustrated in Figure 2, there are notable correlations between the "Energy Efficiency" data and the "Electricity Usage," "Gas Usage," and "Production Output" data. There is a robust positive correlation between energy efficiency and production output. These two variables exhibit a strong correlation, demonstrating a tendency for simultaneous increases and decreases.

Conversely, a robust inverse correlation is observed between "Energy Efficiency" and "Electricity Usage." This strong negative relationship indicates that as one of these two variables increases, the other decreases and this relationship becomes robust. The relationship between "Energy Efficiency" and "Gas Usage" is relatively weak and negative. The weak negative relationship indicates that while one of these variables increases, the other decreases, but this relationship is relatively weak.

Subsequently, histogram plots were constructed to illustrate the distribution of the principal attributes identified by the correlation matrix (i.e., "Energy Efficiency" versus "Electricity Usage," "Gas Usage," and "Production Output"). The graphs mentioned above are presented in Figure 3.

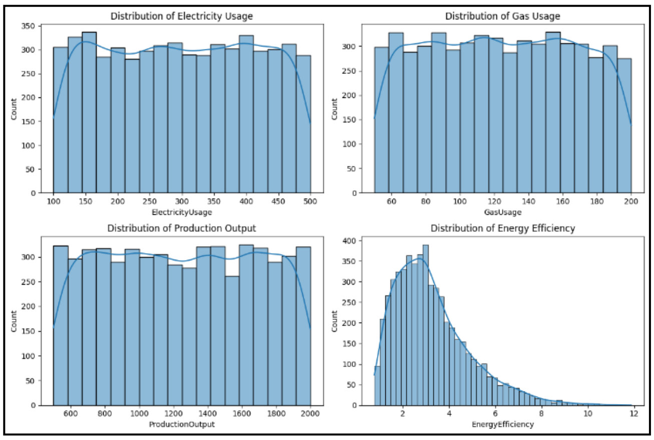

Figure 3. Energy Use and Efficiency Breakdown Graphs

An examination of the histogram plots presented in Figure 3 reveals that the data about "Electricity Usage," "Gas Usage," and "Production Output" are distributed similarly to one another. These key attributes are distributed relatively uniformly across a broad range, exhibiting a relatively flat distribution with no discernible concentration.

In contrast to the attributes above, the "Energy Efficiency" distribution exhibited a rightward skew. This suggests that energy efficiency is predominantly concentrated at lower values, lacking high-efficiency observations. This indicates that, in general, energy efficiency is at relatively low levels, although it is possible to achieve high efficiency in certain instances.

Following the interpretation of the histogram plots of the key attributes, a cluster analysis was conducted on the data set. Before the clustering analysis, all data were standardized using the StandardScaler function. Subsequently, principal component analysis (PCA) was employed to reduce the dimensionality of the dataset while retaining the most salient information. As a consequence of this analysis, the dataset was reduced to two principal components, thus facilitating enhanced visualization and clustering.

The K-Means algorithm was then applied to the data set, which had been reduced in size through PCA, to obtain three clusters. The labels corresponding to the abovementioned clusters were then appended to the data frame. The PCA dimensionality reduction is presented in Figure 4 to facilitate a visual examination of the cluster distinctions.

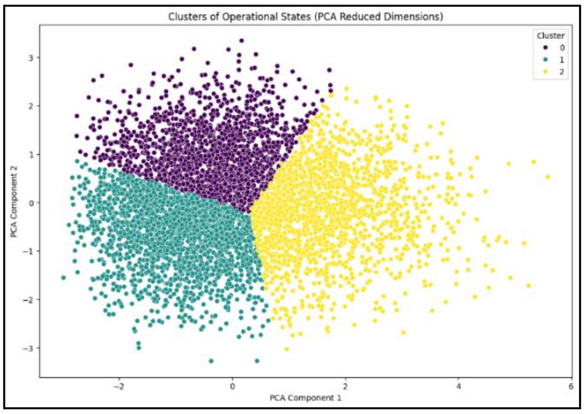

Figure 4. Clusters in Reduced Feature Space with PCA

As illustrated in Figure 4, the principal component analysis (PCA) technique has formed three distinct clusters in the reduced-dimensional feature space. The characteristics and averages of these clusters are as follows:

Cluster 0:

- Electricity Usage: High (average: 337.17 kWh)
- Gas Utilization: High (average: 144.34 m^3)
- Production Output: Medium (average: 1291.58 units)
- Energy Efficiency: Medium (average: 2.74 units/kWh)
- Load Percentage: Medium (average: 68.86%)

Cluster 1:

- Electricity Use: Very High (average:336.99 kWh)
- Gas Utilization: High (average:120.64 m^3)
- Production Output: Low (average:943.63 units)

- Energy Efficiency: Low (average: 2.14 units/kWh)
- Load Percentage: High (average: 78.82%)

Cluster 2:

- Electricity Use: Low (average: 209.79 kWh)
- Gas Utilization: Low (average: 107.90 m³)
- Production Output: High (average: 1595.03 units)
- Energy Efficiency: High (average: 5.26 units/kWh)
- Load Percentage: Medium (average: 75.38%)

The three clusters identified by PCA in the reduced feature space exhibit disparate energy consumption and efficiency profiles. Cluster 0 indicates a group with elevated energy consumption and moderate efficiency, whereas Cluster 1 is distinguished by markedly elevated energy consumption and low efficiency. Cluster 2 represents a group with low power consumption and high efficiency.

A bar chart is provided for each attribute to facilitate a more effective visual comparison of the average values for the three clusters. The bar chart is presented in Figure 5 for reference. This visualization provides a more detailed representation of the main differences between the clusters and their impact on energy efficiency.

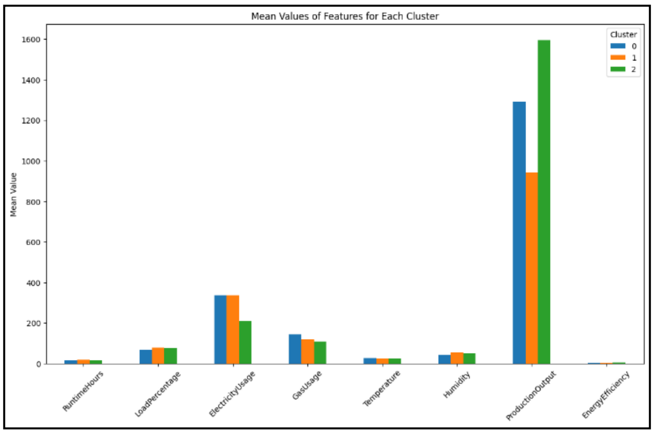

Figure 5. Mean Values of Attributes by Clusters

As illustrated in Figure 5, the most notable distinction between the clusters is the production output. Cluster 2 exhibits a markedly elevated production output compared to the other clusters. This indicates cluster 2 has more efficient production processes or a higher production capacity. The discrepancy in production output between clusters 0 and 1 is less pronounced, with cluster 0 exhibiting a marginal increase in the mean production output.

Concerning electricity consumption, cluster 0 exhibits the highest average, while cluster 2 displays the lowest average. This indicates that Cluster 0 is associated with the utilization of a highly energy-intensive process or machinery, whereas Cluster 2 is distinguished by its enhanced energy efficiency. The electricity consumption of Cluster 1 is nearly identical to that of Cluster 0, exhibiting only a slight decrease.

Cluster 0 has the highest average gas consumption, followed by Clusters 1 and 2. This indicates that there may be a correlation between gas consumption and production output. It can be postulated that clusters that consume more energy may also have higher gas consumption values.

No significant discrepancies are evident between the clusters regarding load percentage, although Cluster 1 and Cluster 0 exhibit slightly elevated averages.

No notable discrepancy is observed among the clusters regarding energy efficiency, which indicates that the clusters may exhibit comparable characteristics.

Other parameters, including working hours, temperature, and humidity, were not significantly different between the clusters.

In conclusion, Cluster 2 is distinguished by high production output and low energy consumption, which can be regarded as an advantageous scenario in terms of energy efficiency. Despite its high electricity and gas consumption, Cluster 0 does not have the highest production output, suggesting that the machines or processes within this cluster may be energy-intensive but inefficient. Cluster 1 occupies a mid-range position concerning both energy consumption and production output.

The analysis yielded the development of optimization strategies for each cluster to enhance energy efficiency within the enterprise:

For Cluster 0, it is recommended that;

- A detailed energy audit could be conducted to identify the underlying causes of the elevated level of energy consumption.
- Machine settings and production schedules could be optimized to reduce energy consumption.
- Preventive maintenance practices could be implemented to enhance the efficiency of the machines in question.

For Cluster 1, it is recommended that;

- Operational processes could be redesigned to increase production output while reducing energy consumption.
- Advanced energy management systems could be implemented to monitor and control real-time energy use.
- Employees receive training on energy efficiency and operational strategies.

For Cluster 2:

- The cluster's most effective energy efficiency practices could be documented and standardized.
- Investment in technology improvements should be made to increase energy efficiency further.
- Continuous improvement efforts could be undertaken to maintain and enhance existing efficiency.

It is anticipated that the implementation of the aforementioned insights about energy efficiency within the company will result in an estimated reduction of 4.5% to 10% in energy consumption [6-8] and a concomitant decrease in energy costs by 20% to 30% [5].

5 Discussion

The study by Benedetti et al. (2016) on energy efficiency for compressed air systems in large and energy-intensive industrial companies is significantly aligned with the optimization strategies proposed in our study [9]. In particular, the approach of "identifying and standardizing best practices" proposed for Cluster 2 in our study is consistent with the "benchmarking indicators" put forth by Benedetti et al. Furthermore, the strategies of "implementing energy metering systems," "adjusting operational parameters," and "transferring knowledge and identifying best practices across companies and sectors" align with the optimization strategies identified for Cluster 0, Cluster 1, and Cluster 2, respectively.

The study by Bonfa et al. (2018) on energy efficiency for compressed air systems in pharmaceutical manufacturing plants identified similar optimization strategies [8]. Among the strategies identified, two are particularly noteworthy: "specific maintenance interventions related to compressors detected through monitoring" and "energy monitoring and control charts for compressors." However, it is notable that strategies such as "determining the optimal activation sequence" and "developing a centralized control system to manage the system compressors" were not proposed in this study. Conversely, using data analysis techniques to achieve energy efficiency is a common approach in our research, as evidenced by the two studies in the literature.

The findings of this study have several implications for academic, practical, and managerial contexts. From an academic standpoint, the combination of data analysis approaches has been shown to yield noteworthy results in studies. Applying PCA and K-clustering methods for energy efficiency in the wood processing industry has substantially contributed to the existing literature on the subject. From a practical standpoint, cluster-based analyses present a viable avenue for enhancing energy efficiency by focusing on specific processes or machinery within an enterprise. Furthermore, the efficacy of data-driven decision-making in operational processes is underscored. From a managerial perspective, this study offers insights into the optimal allocation and prioritization of resources within enterprises, facilitates the integration of energy efficiency as a core organizational value, and provides a framework for developing strategic plans.

As is the case with any study, this one has limitations. The limitations of the data set and the extensive data requirements are particularly noteworthy. Moreover, the study's generalizability is constrained at the sectoral or enterprise level.

Applying the methodology used in this study to other industries may yield different results in future studies. Furthermore, integrating machine learning and artificial intelligence can facilitate more productive and sophisticated outcomes.

6 Conclusions

The wood processing industry places a significant emphasis on energy efficiency. The machinery utilized in the wood processing industry is characterized by high capacity, resulting in considerable energy consumption in proportion to the output produced. Accordingly, this study undertook a comprehensive investigation into optimizing energy efficiency in wood processing facilities. Clustering operational data can identify specific operational situations and characteristics, and targeted optimization strategies can be proposed to improve energy efficiency. This approach can be further refined by integrating additional data and advanced analytical techniques, thereby facilitating continuous monitoring and improvement of the performance of wood processing machines.

The findings of this study can be summarized as follows:

- The dataset utilized in the study revealed that energy efficiency, electricity usage, gas usage, and production output were the most salient features.

- A total of three clusters (Cluster 0, Cluster 1, and Cluster 2) were formed in the reduced feature space using PCA.

- Cluster 2 is distinguished by its highly efficient operations, low energy consumption, and high production output. Enhancing energy efficiency requires employing advanced technologies and prioritizing continuous improvement initiatives.

- Cluster 0 indicates inefficient operations with high energy consumption and moderate production output. It is recommended that machine settings and production schedules be optimized and preventive maintenance plans be established to improve energy efficiency.

Cluster 1 encompasses inefficient operations characterized by high energy consumption and low production output. To improve its energy efficiency, advanced energy management systems capable of monitoring and controlling real-time energy use would be beneficial. Furthermore, training employees on energy efficiency and operational strategies would be advantageous.

7 Declarations

7.1 Study Limitations

None

7.2 Funding source

None.

7.3 Competing Interests

There is no conflict of interest in this study.

7.4 Authors' Contributions

Author contributed equally to this work.

References

[1] Statista. “Estimated value of the furniture market worldwide from 2022 to 2030”. https://www.statista.com/statistics/977793/furniture-market-value-worldwide/ (10.08.2024).

[2] Fortune Business Insights. “Furniture Market Size, Industry Share & COVID-19 Impact Analysis, By Raw Material (Wood, Metal, Plastic, and Others), Category (Indoor and Outdoor), End-User (Residential, Office, Hotel, and Others), and Regional Forecast, 2023-2030”. https://www.fortunebusinessinsights.com/furniture-market-106357 (5.08.2024).

[3] Global Market Insights. “Wooden Furniture Market - By Product Type (Indoor Furniture, Outdoor Furniture), By Wood Type (Hardwood, Softwood), By Price, By Application, By Distribution Channel Forecast 2024 – 2032”. https://www.gminsights.com/industry-analysis/wooden-furniture-market (5.08.2024)..

[4] Imarc Transforming Ideas into Impact. “Wood Furniture Market by Wood Type (Hardwood, Softwood), Distribution Channel (Retail, Online), End User (Residential, Commercial), and Region 2024-2032”. https://www.imarcgroup.com/wood-furniture-market (07.08.2024).

[5] Sihn W, Sobottka T, Heinzl B, Kamhuber F. “Interdisciplinary Multi-Criteria Optimization Using Hybrid Simulation to Pursue Energy Efficiency Through Production Planning”. CIRP Annals, 67(1), 447-450, 2018.

[6] Wen X, Cao H, Hon B, Chen E, Li H. “Energy Value Mapping: A Novel Lean Method to İntegrate Energy Efficiency into Production Management”. Energy, 217, 119353, 2021.

[7] Sobottka T, Kamhuber F, Sihn W. “Increasing Energy Efficiency in Production Environments Through an Optimized, Hybrid Simulation-Based Planning of Production and its Periphery”. The 24th CIRP Conference on Life Cycle Engineering, Kamakura, Japan, 8-10 March 2017.

[8] Bonfa F, Benedetti M, Ubertini S, Introna V, Satolamazza a. “New Efficiency Opportunities Arising from Intelligent Real Time Control Tools Applications: The Case of Compressed Air Systems’ Energy Efficiency in Production and Use”. 10th International Conference on Applied Energy (ICAE2018), Hong Kong, China, 22-25 August 2018.

[9] Benedetti M, Bertini I, Bonfà F, Ferrari S, Introna V, Santino D, Ubertini S. “Assessing and Improving Compressed Air Systems’ Energy Efficiency in Production and Use: Findings from an Explorative Study in Large and Energy-Intensive Industrial Firms”. The 8th International Conference on Applied Energy (ICAE2016), Beijing, China, 81 October 2016.

[10] Emre İ.E, Selcukcan Erol C. “Statistics or Data Mining for Data Analysis”. Journal of Information Technologies, 10(2), 161-167, 2017.

[11] Bennert T, Hanson D, Maher A. “Influence of pavement surface type on tire/pavement generated noise”. Journal of Testing & Evaluation, 33(2), 94-100, 2005.

[12] Wold S, Esbensen K, Geladi P. “Principal Component Analysis”. Chemometrics and Intelligent Laboratory Systems, 2(1-3), 37-52, 1987.

[13] Ghiasi S, Srivastava A, Yang X, Sarrafzadeh M. “Optimal Energy Aware Clustering in Sensor Networks”. Sensors, 2, 258-269, 2002.

[14] Abdi H, Williams L.J. “Principal Component Analysis.” WIREs Computational Statistics, 2(4), 433-459, 2010.

[15] Bro R, Smilde A.K. “Principal Component Analysis”. Analytical Methods, 6(9), 2812-2831, 2014.

[16] Zhou W, Xu K, Yang Y, Lu J. “Driving Cycle Development for Electric Vehicle Application Using Principal Component Analysis and K-Means Cluster: With The Case of Shenyang, China”. 8th International Conference on Applied Energy (ICAE2016), Beijing, China, 8-11 October 2016.

[17] Chen Z, Xiong R. “Driving Cycle Development for Electric Vehicle Application Using Principal Component Analysis and K-Means Cluster: With The Case of Shenyang, China”. 9th International Conference on Applied Energy (ICAE2017), Cardiff, UK, 21-24 August 2017.

[18] Pořízka P, Klus J, Képeš E, Prochazka D, Hahn D.W. “On The Utilization of Principal Component Analysis in Laser-Induced Breakdown Spectroscopy Data Analysis, A Review”. Spectrochimica Acta Part B, 148, 65-82, 2018.

[19] Cebeci Z. “fcvalid: An R Package for Internal Validation of Probabilistic and Possibilistic Clustering”. Sakarya Unıversıty Journal of Computer and Informatıon Scıences, 3(1), 11-27, 2020.

[20] Clayman C.L, Srinivasan S.M, Sangwan R.S. “K-means Clustering and Principal Components Analysis of Microarray Data of L1000 Landmark Genes”. Procedia Computer Science, 168, 97-104, 2020.

[21] Jannson N.F, Allen R.L, Skogsmo G, Tavakoli S. “Principal Component Analysis and K-Means Clustering as Tools During Exploration for Zn Skarn Deposits and Industrial Carbonates, Sala Area, Sweden”. Journal of Geochemical Exploration, 233, 106909,2022.

[22] Ilu S.Y, Rajesh P, Mohammed H. “Prediction of COVID-19 Using Long Short-Term Memory By Integrating Principal Component Analysis And Clustering Techniques”. Informatics in Medicine Unlocked, 31, 100990, 2022.

[23] Liao J, Lin J, Wu G. “Two-Layer Optimization Configuration Method for Distributed Photovoltaic and Energy Storage Systems Based on IDEC-K Clustering”. Energy Reports, 11, 5172-5188, 2024.

[24] Eid M.H, Eissa M, Mohamed E.A, Ramadan H.S, Czuppon G, Kovacs A, Szűcs P. “Application of Stable Isotopes, Mixing Models, and K-Means Cluster Analysis to Detect Recharge and Salinity Origins In Siwa Oasis, Egypt”. Groundwater for Sustainable Development, 25, 101124, 2024.

[25] Dugan E.L, Barbuto A.E, Masterson C.M, Shilt J, “Multivariate Functional Principal Component Analysis and K-Means Clustering to Identify Kinematic Foot Types During Gait in Children with Cerebral Palsy”. Gait & Posture, 113, 40-45, 2024.

[26] Gürcan E. K., Soysal M. İ., Genç S. “The Determination of Relationships Between Live Weight with Various Body Measurement in Japanese Quail Using Principal Component Analysis”. Journal of Poultry Research. 9(1), 27-33, 2010.

[27] Durgun B., Uygur V., Durgun B., Sukuşu E. “Assessment of Relations Between Micro Element Availability and Soil Properties in Isparta-Atabey Plain Using Principal Component Analysis”. Anadolu Journal of Agricultural Sciences. 32(2), 258-268, 2017.

[28] Çolak B., Durdağ Z., Erdoğmuş P. “Automatic Clustering with K-Means”. El-Cezerî Journal of Science and Engineering. 3(2), 315-323, 2016.

[29] Arya N. (2023, November 10). K-Means Clustering for Unsupervised Machine Learning. EJable. https://www.ejable.com/tech-corner/ai-machine-learning-and-deep-learning/k-means-clustering.