International Journal of Studies in Business Management, Economics and Strategies
Volume 5, Issue 1, January- 2026
ISSN (E): 2949-883X
Scholarsdigest.org

UTILIZATION MACHINE LEARNING TO BUILD A PREDICTIVE MODEL TO IMPROVE PACKAGING QUALITY IN PRODUCTION PROCESSES

Basma M. Hamad
College of Administration and Economics
Babylon University Babil, Iraq
Basma.mahdi@uobabylon.edu.iq

Abstract

The study is geared toward the development of a predictive model of packaging quality since it plays an important role in the food industry. Nutritional value losses and spoilage are directly associated with packaging defects. The study factors those variables that strongly influence the quality of packaging and attempts early prediction of defectiveness, hence a reduction in financial loss, which will drive continuous improvement efforts to make better quality management decisions. The fruits of data mining and the machine learning revolution have essentially made effective predictive methods possible. An applied analytical approach based on previous studies and production records from Noor Al-Kafeel Company's chicken and chicken parts factories was adopted by this research for constructing the proposed model. Methodology used machine learning techniques to analyze relationships between a set of operational, human, and environmental variables that may predict packaging quality. 300 real-world operational observations suitable for modeling comprised the sample. Data mining techniques and the Random Forest Algorithm were used. The predictive model produced an accuracy rate of 97.6%, hence proving robustness and applicability in industrial environments . Factors that influence packaging quality analyzed included defect rate, humidity, heat sealing, and machine speed; a revelation that these factors ranked high in terms of their importance. Due to automation, the human factor is less significant.

These findings make it possible for the proposed model to help in moving from post-inspection detection of defects to proactive, predictive methodology, which would improve production process efficiency, add value and sustain industrial organizations.

Keywords: Packaging quality, Predictive model, Machine learning, Random Forest Algorithm, Orange Data Mining.

Introduction

The food industry is among the most significant due to its direct relation with aspects of human health and the provision of daily sustenance. The chicken meat and meat product packaging plants are leading among these industries since chicken meat is nutritionally valuable while being a source of high-quality protein. However, this industry faces numerous challenges in its efforts toward maintaining quality in terms of minimizing packaging defects because sensitivity to microorganisms can result in spoilage as well as disease (Fontes et al., 2011). Factors that influence the quality of chicken meats include operational, environmental, and human factors. Packaged foods retain their sensory and nutritional attributes when they are microbiologically stable (Vytejckova et al., 2017). The literature confirmed that packaging, its monitoring, and the quality assurance of it are important because during the physical flow, products may be exposed to changes in their characteristics in an unacceptable way. From here, the importance of packaging quality has a great role in maintaining the delivery of food products up to their final destination and consumption by consumer.

The fast growth of artificial intelligence and machine learning has greatly improved the use of predictive analysis in finding defects in operations, making mitigation efforts better, and ensuring packaging quality. Earlier research has proven that such models allow for the early spotting of patterns and signals connected with possible quality problems by using machine learning algorithms. This early detection gives enough time to take corrective actions before defect rates go up; thus, it helps to lower defects in manufacturing and reduce waste during production (Zulqarnain et al., 2022).

Predictive modeling decision support building requires tools and techniques. Data Mining Orange is a dynamic electronic platform that merges Machine Learning schemes with Data Analysis methods plus Result Viewing and Exploration via Visualizations. It permits the comprehension of variables together with value (Altayeb, 2024 & Arabiat). This has also been applied by several scholars since it is a technology. To analyze data concerning food quality: Phoemchalard et al., (2024) conducted research aimed at creating a classifier on buffalo meat, beef, and goat; results compared several algorithms SVM, Naïve Bayes Neural Net not less than 100% accurate The Random Forest algorithm was best according to AUC and F1 indices over 97 percent Meanwhile, Şahin et al., (2024) studied deep learning-based detection models using "you only look once" algorithm achieved very high accuracies above 99% in packaging time meat classification Other industries were researched with the use of Data Mining Orange technology in view of the pharmaceutical industry and a global pharmaceutical company which intended to carry out an analysis of multidimensional data using the Patial Least Squares algorithm. This study has indicated how important data cleaning is plus the removal of outliers and their effect in making prediction models better. Previous studies conducted using Data Mining Orange indicate that it permits all steps beginning with data collection and processing up to building predictive models and assessing their performance across various industrial settings.

It opens with theory. It mostly method, shares every step of data collection, and processing, in addition to their predictive model development. The results share the output from the machine learning exercise, give an assessment on how significant the research variables are, and look at the performance indicators for the model. After that come findings. It is a study about machine learning and attempts to practically predict packaging quality within the food industry.

2. Theoretical framework

Operations management is the architecture, functioning, and control of highly efficient production systems thus making it a major pillar in the performance of industrial organizations. Production management is an important function of all industries that need to turn raw material into finished goods efficiently and effectively through activities ranging from purchasing and manufacturing to packaging and distributing. The end product should be high-quality goods at the lowest cost possible. Among many objectives toward production management, quality assurance and cost reduction will keep a firm competitive as well as sustainable in the highly globalized world markets. These goals are realized through strategic planning and resource optimization accompanied by continual improvement processes.

Packaging is the most important factor when it comes to the safety and quality of food against environmental factors, spoilage, and contamination. It creates a protective barrier against physical factors (shocks, temperature variations), biological factors (microbial contaminants), and chemical factors (oxygen, moisture) that may attack the content of the package thereby ensuring sanitation, freshness, and quality storage period of food for the whole storage time (Agarwal et al., 2022). In addition, food packaging can protect products from damage caused by pathogenic organisms transmitted through food thereby conserving food safety as well as properties under environmental influences thus preserves the quality of a given product (Chen, 2023). Additionally, it increases the shelf life of a product hence reducing wastage. High barrier materials are mostly made of tinplate or low permeability plastics; antimicrobial agents and indicators of freshness incorporate active and intelligent packaging. This makes the delay in oxidation and development of harmful bacteria possible, making shelf life better (Vasile & Baican, 2021; Yan et al., 2022). These novelties ensure that the quality of the product is maintained within its shelf life by monitoring food safety (Chen, 2023). Also, packaging provides information regarding nutritional content or any factor on safety plus facilitating handling, storage, and transportation for consumers about the particular food product (Vasile & Baican 2021). Though packaging is very important in providing food safety as well as maintaining its quality, it has an adverse effect on environmental sustainability due to the material used in packaging. Recent consumer perception has led them to rate more interest in sustainable packaging; hence there should be a balance among safety, quality, and environmental friendliness.

Stages of the packaging process in the food industry include the choice of appropriate packaging materials, filling and sealing, labeling, inspection, storage, and transportation. All these emphasize product quality, safety, and consumer satisfaction. The nature of the food product determines what material will be used for its packaging, shelf life, and strength of protection required. The materials used are glass, metal, plastic paper, and others. Shelf lives provide an excellent moisture and oxygen barrier that glass and metal supply as discussed by Alamri et al. (2021) Plastic is known for alternatives for flexibility and possibilities in engineering while paper and cardboard are known for sustainable solutions. Flexible polymer films are generally used as a combination with the use of some other type of packaging material, e.g., plastic bags or rigid plastic containers that can be closed with transparent plastic lids, since fresh chicken meat has a high moisture content. This protects

against exposure to the air and must be ensured to be safe (non-poisonous) and non-reactive with the meat; otherwise, it would release contaminants or with undesirable flavors.
The packaging process starts by placing an appropriate quantity of chicken in the package. "Automatic" machines are used to weigh and divide portions so that there is uniformity as well as prevention against wastage. The area should be clean and sterile where packaging takes place thus reducing contamination Dudbridge,(2016) Accurate weighing for portion control, pricing, ensuring the right quantity to consumers has highlighted importance advanced weighing technologies enhance efficiency with minimal losses sealing labeling follows Sealing incorporates moisture oxygen controllers a type active extend shelf life maintain quality Bhardwaj et al.(2019) Seal integrity is important in preventing contamination and spoilage-both of which result in enormous losses besides causing a setback to consumer confidence, as explained by Dudbridge (2016). The labels also carry basic information on the product content details among other nutritional information expiry date etc. This added another dimension to what had earlier been articulated about advanced technologies and materials at the packaging stage-that biodegradable packaging material is equally sustainably oriented because it reduces environmental impact while sustaining functional integrity within systems of packaging. Green Packaging pursuit is a trend for the food industry towards sustainability accompanied by increased public health awareness (Han et al., 2018).
The type of packaging, production line speed, and operator skill are some very critical factors that will influence packaging quality. In return, all these factors play a vital role in implementing the standards of safety, efficiency, and satisfaction of customers. The choice of packaging material may directly relate to protection as well as preservation of food (Tkalec et al., 2018). Creative solutions for active and intelligent packaging technology introduce shelf life extension plus improvement on total safety conditions about food attainable modern requirements based on consumer demands (Han et al., 2018). High seal integrity in high-speed production lines might be compromised thereby causing spoilage at the end product; more waste (Dudbridge, 2016). Companies desire high-speed production that compromises quality due to defects sealing or final packaging, so too does Huynh (2019) argue. Worker skills also directly affect product quality. Skill is key to machine maintenance and the assurances of the packaging process (Dudbridge, 2016), hence training without end and following packaging standards can lead to a great change in the efficiency and effectiveness of the packaging process (Huynh, 2019).
Product quality is the main goal that all industries strive to achieve. In the food industry, it represents the degree of conformity of the final product to sensory and functional standards as well as regulatory requirements that satisfy the consumer and ensures safety. This quality has attributes including microbiological stability, nutritional values, plus the product’s appearance and taste. Process quality speaks to the effectiveness of the production system and its ability to achieve a consistent level of product quality. This is attained through sound operational practices and control measures instituted at different stages of the production, packaging, and transportation processes (Semercioz-Oduncuoglu& Luning, 2025). Recent studies in this field have revealed that it is digital transformation technologies under big data analytics and the Internet of Things integrated with quality management systems that

improve process efficiency by controlling variability reflected in improved product quality meeting international standards (Peres et al., 2025). Product quality and process quality are therefore two complementing pillars for food safety, supply sustainability, chain competitiveness increases for manufacturers.

Quantitative criteria are measured under the defective rate and waste rate. The defective rate comprises percentage units which are not up to mark with required specifications wherein the waste rate denotes material or units wasted during the production process expressed as a percentage. Quality improvement strategies comprise basic methodologies of process and product improvements such as PDCA, and Six Sigma among others. Plan, Do, Check, Act is a simple cycle continuously improved by planning, doing, monitoring, and revising it; thus making continuous improvement very easy to be understood by any person and implemented fostering continuous improvement as part of the culture in various industries. Meanwhile, Six Sigma applies its more structured methodology through data and the framework of DMAIC (Define, Measure Analyze Improve and Control). This relates to upgrading quality and cutting defects in complex settings. The method has seen use mostly in making and service fields bringing about big cost cuts and better efficiency (Saxena & Rao, 2019), (Ghelani, 2023).

Quality control, together with data analysis, has found its place as growing concerns in the packing line complemented by increasing advances of digital transformation technologies and Internet of Things systems. Real-time production rates and defect rates or characteristics of packaging will allow for early spotting deviations in KPIs such as incorrect filling percentage, spoilage rate, and frequency of line downtime. The quality control relies on production-based statistical analysis to input defects in packaging lines and where their causes lie. This, therefore, provides an avenue for companies to analyze systematically the defects that consequently improve the product quality and optimization of processes. Apart from that, when data mining is combined with Six Sigma methodology, historical data about defects is analyzed fully and patterns are disclosed for future improvements. Quality control measures that are based on data have been seen to lower defect rates hence better operational efficiency as well as increased customer satisfaction. But data analysis shows the needs for more training and gaps in the process that must be filled to hit top quality marks. So, one can say that packaging lines data analysis is a key tool for active quality control to keep the process going and hold consumer trust in food products.

Artificial intelligence (AI) and machine learning technologies development have significantly influenced human life positively in various areas. AI may be defined as “the simulation of human minds to make computers think and behave like humans by performing such activities as learning, and problem-solving” (Zhang & Lu, 2021). The integration of Artificial Intelligence with data analytics revolutionizes the way firms extract nuggets from huge quantities of data therefore enhancing decision making (Kumar, 2023). In addition, it strengthens the capacity for quick and accurate assessment of large volumes of information thereby helping overcome challenges posed by increasing amounts of data resulting from rising trends in digitization (Kumar & Singh, 2024). AI enables firms to optimize operations besides enhancing strategic planning through unveiling concealed patterns and tendencies (Gandomi et al., 2023; Kumar & Singh, 2024).

Artificial Intelligence (AI) is a discipline that seeks to create systems that can undertake tasks with a level of competency at comprehension, reasoning, and decision-making normally associated with humans. It comprises Machine Learning and Deep Learning as major elements. The basic differentiation among them is in the approach usable for processing information and fields of application.

Machine learning is defined as the study of those algorithms and systems whereby a computer can learn from data and subsequently improve performance over time on some tasks without explicit programming for each task. Machine learning heavily relies on training large datasets to computational models so that these models acquire the competence to perform such tasks as prediction, classification, or general pattern detection (Janiesch et al., 2021; Baierle et al., 2024). The three variants of machine learning are supervised, unsupervised, and reinforcement learning. Data with known input and output (labels or labels) used by supervised learning train a model about their relationship; this type is found in classification (like defining quality) or regression. Data labels are not used in unsupervised learning but new structures or patterns inside the data are discovered. It is used for clustering or reducing dimensions. The other type is called reinforcement learning and this essentially means improving strategies of decision-making in a particular environment based on experiences and steps that are rewarded or penalized(Asongo et al., 2024).

The decision tree algorithm thus helps in understanding how the decisions are made based on the features of the data. It is a machine learning algorithm used for building a model through a tree structure to make decisions or classifications. The algorithm works by repeatedly splitting the data based on particular features, routing it down the branches of the tree, and finally arriving at a specific decision or classification at the leaves. This method, though simple and attractive for many nonlinear or somewhat complex data, has found its application in prediction, classification, risk assessment, and many more due to their easy accessibility (Asongo et al., 2021). Conversely, random forest algorithms are utilized for developing prediction as well as classification models. It is based on constructing an ensemble of individual decision trees and then merging their outputs into one final output that will be even more precise and robust. The benefits of this algorithm are its capacity for high-dimensional data, the mitigation of overfitting problems, and tremendous strength when it comes to issues of classification as well as prediction. It is also known to resist imbalances in the data. This consequently makes random forest an extremely powerful yet versatile algorithm and hence leads to high usability. While Naive Bayes works on a probability principle under the guidance of Bayes theorem by assuming that for every class the input variables are mutually independent thus reducing computational complexity, this model happens to be highly efficient with high-dimensional data and is also quick in terms of training and scalability. That shows how effective it is toward dealing with incomplete data and insignificant features that do not play a great role toward accuracy in a model, thus making the Naive Bayes algorithm simple as well as efficient.

It is key to understand that results and execution of prescient models, particularly inside the domain of machine learning and information science, are profoundly impacted by how cautiously highlights and target factors are characterized. Highlights essentially speak to engaging or input factors given to the predictive calculation, on which designs are recognized and induced. Alternately, the target factor too known as the result or subordinate variable speaks specifically to those comes about or categories that the demonstrate is endeavoring to anticipate, gauge, or classify (Kwon & Jeong, 2023) (Chen et al., 2023).

3. Research Methodology

3.1. Materials and Methods

This study uses a quantitative applied approach in line with the main aim of improving the quality of packaging processes while reducing defects for the concerned company under investigation. More particularly, this is done by forming a predictive model through machine learning based on historical data related to conditions of operation; performance characteristics of equipment; human factor characteristics and raw material properties from that production line. It also assessed this model using some statistical indicators: accuracy, recall, and the F1 index besides confusion matrix where recall and F1 indices were found more appropriate concerning balanced data for evaluation of this model. The recall and F1 indices were found more appropriate to evaluate this model concerning balanced data. Model reliability, together with generalizability to such industrial environments, will have improved when real company historical data is used. Also, going over and checking results again helped keep the model's performance steady, along with being able to look at how different models work. The model's value and real use got better because it matched up with the factory's known quality signs.

3.2. Research population and sample

The research community consisted of operational data related to the packaging lines at a chicken and poultry processing facility. Noor Al-Kafeel Industrial and Agricultural Investments Company operates as part of project efforts towards food industry development as well as achieving self-sufficiency in animal products within the local sector; this chicken and poultry processing plant forms part of its most integrated projects, including production processes from poultry rearing to slaughtering and cutting up to packaging and distribution. The capacity for operation in the plant is more than 25 tons. Three hundred valid-for-the-model actual operating notices were picked from that plant between June 1, 2024, through May 30, 2025. Days with different reasons for production stoppage are a typical representation of all possible types of conditions included in that period. Data revealed an imbalance in the distribution of packaging quality categories because sound cases represent the majority of data compared to cases of quality defects. This goes to say that the actual nature of any process has its real-world implications.

3.3. Identifying the variables affecting packaging quality

A set of variables was identified based on previous studies that dealt with packaging quality and which can have an impact on the final product. The independent variables were (packaging machine speed, machine temperature, machine temperature, sealing pressure, humidity ratio, and operator experience). On the other hand, the quality of product packaging was classified based on the quality standards adopted in the factory (high, medium, low) as a dependent variable, as listed in Table No. (1), which shows a description of the variable with the unit/or allowed values, and the limits or allowed range to ensure packaging quality.

Table (1) Variables Affecting Packaging Quality

No.	Variable	Unit / Allowed Values	Range Notes	Source
1.	Packaging speed (PS)	Packages/ minute	To avoid errors and minimize defects Optimal range: 160–200 packs/minute	Gurunathan,et al., 2022
2.	Machine temperature (MaT)	°C	To ensure a tight seal without damaging the film The optimal temperature range: 72–78°C	Mathew, et al., 2016
3.	Sealing Pressure (SP)	Bar	The optimal sealing pressure values during closure are 2.2–2.8 Bar to avoid leakage or damage to the .casing	Mathew, et al., 20216
4.	Humidity Level (HL)	%	To reduce condensation and maintain packaging .quality, the optimal range is 40–65%	Gurunathan,et al., 2022
5.	Film Thickness (FT)	Micron	depending on the type of film used 60–40	Vytejckova, et al., 2017
6.	Material Type (MT)	A / B / C	Depending on the type of supplier or the quality of the material	Vytejckova, et al., 2017
7.	Product Temp (PT)	°C	A temperature between 2–6°C is preferred, according .to Codex Alimentarius recommendations	Gurunathan,et al., 2022
8.	Label Accuracy (LA)	%	Acceptable value ≥ 93% to avoid labeling errors	Mathew, et al., 2016
9.	Defect Rate (DR)	%	Optimal value < 3% to guarantee quality	Gurunathan,et al., 2022
10.	Seal Strength (SS)	N/m	Optimal sealing force > 13 N/m to prevent leakage during transport	Vytejckova, et al., 2017
11.	Operator Experience (OE)	years	.It affects the quality of the operational process	Gurunathan,et al., 2022
12.	Packaging Quality	1 = Good / 0 = Not Good	target variable for prediction	Combined from all three studies

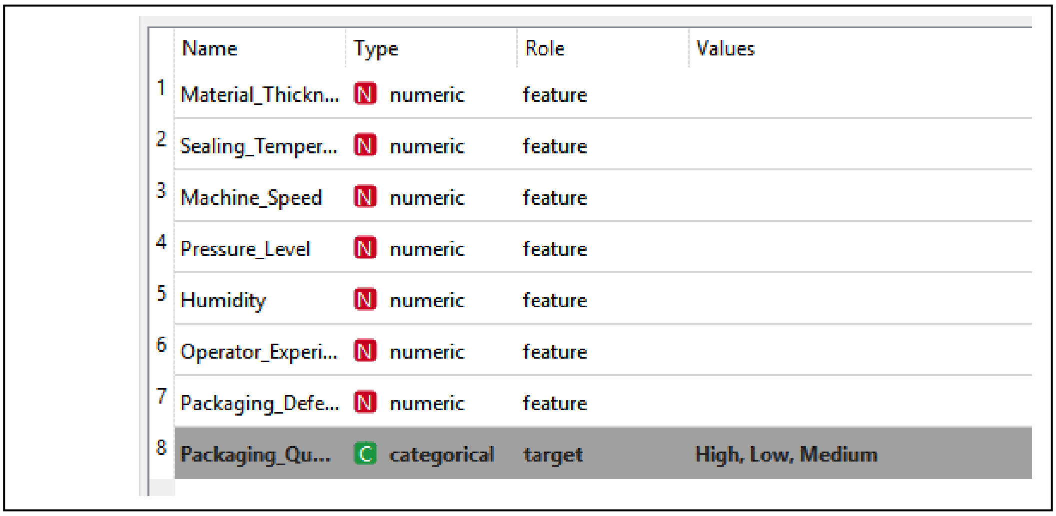
From the table above, seven main variables were adopted as independent variables for the predictive model based on a set of operational considerations and what the literature indicates in the field of industrial packaging processes. This adoption was after conducting a statistical significance analysis of the variables as in Figure (1). The results show that these variables represent factors that strongly influence packaging quality. Also, limiting the influential variables and working to select the most influential ones achieves a systematic balance between the accuracy and simplicity of operating the model. This contributes to reducing manifestations of overadaptation. In addition, they represent measurable variables that can be monitored within the factory sample of the study, which enhances the efficiency of the model and its applicability.

		#	Info. gain	Gini
1	N Packaging_Defect_Rate		0.528	0.1
2	N Humidity		0.022	0.0
3	N Machine_Speed		0.019	0.0
4	N Operator_Experience_Years		0.014	0.0
5	N Pressure_Level		0.014	0.0
6	N Sealing_Temperature		0.010	0.0
7	N Material_Thickness		0.009	0.0

Figure 1: Results of the statistical significance analysis of the independent variables.

3.4. Data Collection and Processing

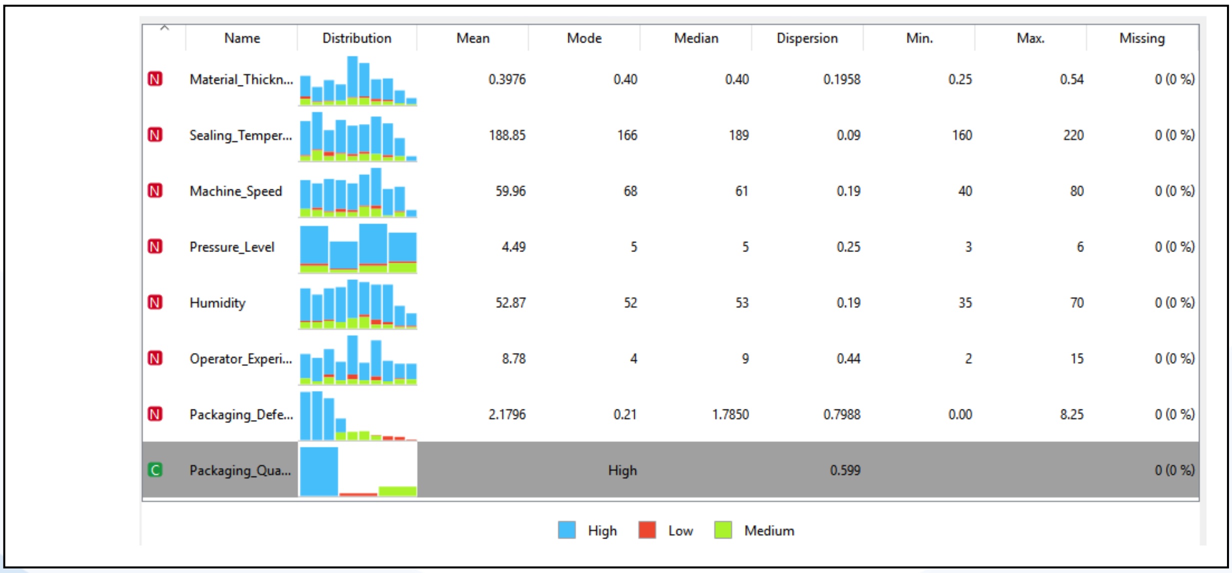
The data, as presented in Table (1), was collected from the factory's production records and quality control and monitoring systems. It underwent several processing stages before being used in the analysis phase. These processes included verifying data completeness, addressing missing values, identifying anomalies, and finally, standardizing the variable formulas. Since the data used was real, there was no need to employ the SMOTE (Synthetic Minority Amplification) technique, which is used to address imbalances in groups by increasing the sample size of the minority group to ensure data balance. The reason for not using SMOTE was to maintain the realistic distribution of operational quality problems and ensure the model's applicability in an industrial environment. To guarantee the accuracy of the predictive model, the data was divided into two groups: the first, comprising 70%, was allocated for training, and the second, comprising 30%, was allocated for testing. Orange Data Mining (ODM) was used to analyze the data and build the predictive model using machine learning algorithms that are appropriate to the nature of industrial data. Analyzing the relationships between operational variables and packaging quality was the focus of the modeling process, exploring the variables shown in Table (1) that have a high impact and cause packaging defects. Figure (2) shows the identification of these variables as attributes and the identification of packaging quality as the target for building the model.



	Name	Type	Role	Values
1	Material_Thickn...	N numeric	feature	
2	Sealing_Temper...	N numeric	feature	
3	Machine_Speed	N numeric	feature	
4	Pressure_Level	N numeric	feature	
5	Humidity	N numeric	feature	
6	Operator_Experi...	N numeric	feature	
7	Packaging_Defe...	N numeric	feature	
8	Packaging_Qu...	C categorical	target	High, Low, Medium

Figure 2: Description of variables.

Descriptive statistics were performed on the independent and dependent study variables, and Figure (3) shows the readiness of the data to build the predictive model, as the data showed a relative balance in distribution, in addition to the absence of missing values. We also note that the statistical values for the thickness of packaging materials, the average speed of the machine, and the average closing temperature are values that fall within the operational ranges, while the variables of worker experience, humidity, and pressure showed moderate variation. This is due to the operating conditions that differ between production batches. The figure also shows a relative dominance of the high quality category with respect to packaging quality, which reflects the natural industrial reality, and this is a justification for adopting a classification algorithm that can deal with this variation without the need to generate artificial data.



	Name	Distribution	Mean	Mode	Median	Dispersion	Min.	Max.	Missing
N	Material_Thickn...		0.3976	0.40	0.40	0.1958	0.25	0.54	0 (0 %)
N	Sealing_Temper...		188.85	166	189	0.09	160	220	0 (0 %)
N	Machine_Speed		59.96	68	61	0.19	40	80	0 (0 %)
N	Pressure_Level		4.49	5	5	0.25	3	6	0 (0 %)
N	Humidity		52.87	52	53	0.19	35	70	0 (0 %)
N	Operator_Experi...		8.78	4	9	0.44	2	15	0 (0 %)
N	Packaging_Defe...		2.1796	0.21	1.7850	0.7988	0.00	8.25	0 (0 %)
C	Packaging_Qua...			High		0.599			0 (0 %)

Figure 3: Descriptive Statistics

Figure (3) shows that the operating environment is heterogeneous, and to analyze this complexity, the need arises to use intelligent predictive models.

3.5. Building the predictive model

After identifying the independent and dependent variables and processing the data, the random forest algorithm was chosen to build a predictive model capable of accurately classifying packaging quality due to its ability to handle high-dimensional data, reduce the problem of overadaptation, and achieve strong performance in classification and prediction. This algorithm is also characterized by its resistance to data imbalance.

Figure (4) illustrates the construction of the predictive model for packaging quality based on historical data from the research sample, following a systematic sequence using Orange Data Mining. This includes stages starting with data loading, identifying dependent and independent variables, data preparation and processing, and conducting exploratory analysis to extract descriptive statistics. The data was divided into two ratios (70% and 30%) for training and final testing, and the classification was selected according to the target variable. This maintains the proportions of the categories within training and testing to ensure balanced representation of the dependent variable (packaging quality) and to achieve an objective evaluation of the model. The algorithm builds and trains the predictive model using (100) trees. To enable repetition, the iterable training feature was activated. The minimum node size was set to (5) to reduce overlearning, which allows for unrestricted tree depth. This contributes to the model's ability to represent complex patterns in the research sample factory data.

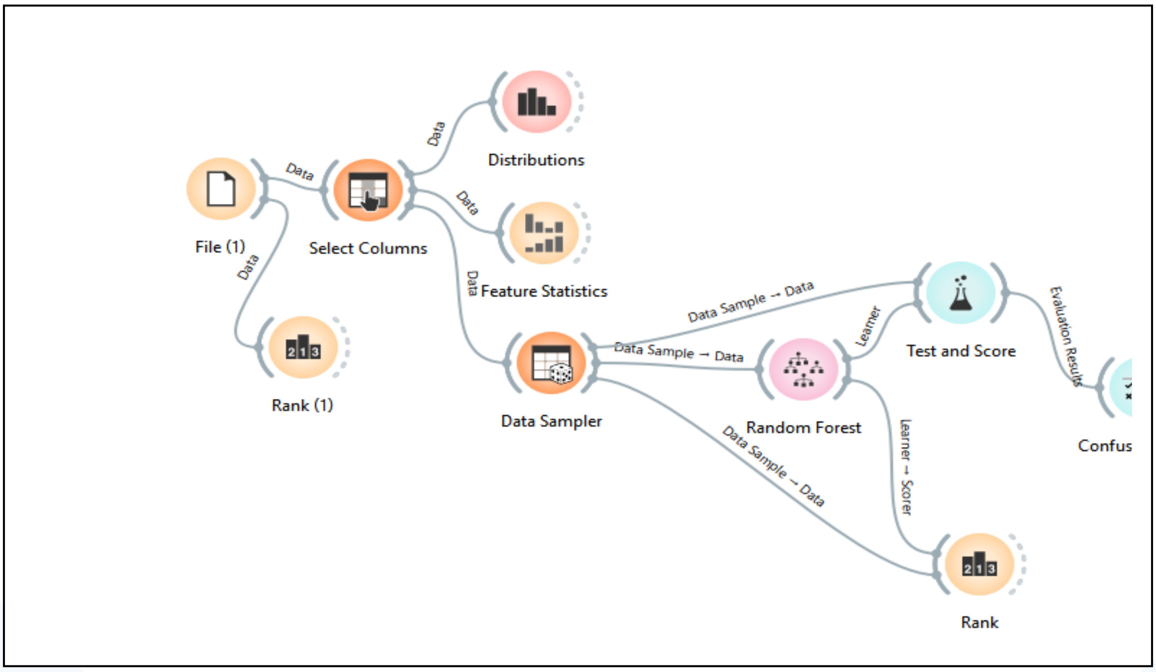


Figure 4: Flowchart of the development of the predictive model for packaging quality

The predictive model for packaging quality was assessed based on data from the research sample factory using several statistical indicators where overall accuracy, positive accuracy, recall, F1 index, Roc curve, and Matthews correlation coefficient (MCC) were major ones. These indicators provided an ability to comprehensively measure efficiency of the model under a situation where there is an imbalance in the dependent variable which is packaging quality. An ambiguity matrix was used to assist in analyzing and understanding the errors of the model as well as its accuracy in distinguishing between various levels of packaging quality. It was adjusted by displaying the number of actual cases for each category together with representing actual values on the rows and predicted values on columns.

An analysis of the significance of the variables most influential on packaging quality was performed, which was calculated based on the geni-decrease across multiple trees in the random forest.

4. Results

The results of the evaluation attest to the high predictive efficiency of the packaging quality model. The Area Under Curve (AUC) statistic posted a value of 0.996, reflecting a very high potential for differentiation between quality categories. In addition, the overall classification accuracy index reached a value of 0.976—very close to one, denoting an extremely high share of correct predictions out of the total number of cases.

High scores on other indices prove that the model can balance its ability to minimize positive errors with addressing cases of low quality. This is seen through the F1 coefficient at 0.974, a Precision index of 0.978, and a Recall index of 0.976. Furthermore, the Matthews Correlation Coefficient (MCC) (0.927) indicates good classification quality despite the multiplicity and relative imbalance of the target variable.

The results of the packaging quality classification are shown by the ambiguity matrix as in Figure (5). If it is classified into three categories (high, medium, low), it appears clear that the model correctly classified the high quality cases it collected, amounting to (136) cases, while (31) cases were classified as medium quality, and (5) cases were correctly classified as low quality. We note that the ambiguity matrix indicates a high level of correct prediction.

Actual \ Predicted	High	Low	Medium	Σ
High	169	0	0	169
Low	0	5	4	9
Medium	1	0	31	32
Σ	170	5	35	210

Figure 5: Correlation Matrix Results

When analyzing the importance of variables from highest to lowest impact in the predictive model using the Random Forest algorithm, Figure (6) shows us that the rate of packaging defects ranked first in importance in terms of impact on the prediction, while the impact of the rest of the variables varied, and this variation reflects the nature of the interactions between operational, environmental and human factors within the research sample environment.

		#	Info. gain	Gini	Rand...
1	N Packaging_Defect_Rate		0.528	0.192	
2	N Humidity		0.054	0.009	
3	N Sealing_Temperature		0.040	0.007	
4	N Machine_Speed		0.030	0.006	
5	N Pressure_Level		0.028	0.007	
6	N Material_Thickness		0.009	0.002	
7	N Operator_Experience_Years		0.006	0.001	

Figure 6: Analysis of the importance of variables affecting the prediction model

5. Discussion

The primary reason for using data mining techniques and packaging quality prediction models is to reduce defects and resource waste. Statistical results confirm the effectiveness of using machine learning techniques to achieve this, and they also support data imbalance correction, ensuring the detection of critical defects. These results align with recent literature indicating that the most important factor in improving production processes is controlling and managing operational and environmental variables.

High accuracy of the packaging quality level is delivered in the ambiguity matrix to underscore the efficiency of the model. The highquality category is achieved with great accuracy from the model, and no errors inside this category are recorded regarding its classification, therefore indicating stability of the model under certain fulfilled conditions. Limited errors recorded for results in medium and low-quality categories indicate that some degree of overlapping operational characteristics may be shared between these two categories. This can happen in actual industrial scenarios due to varying operating conditions. Generally, the model can function as a supporting decision-making instrument for upgrading packaging processes and reducing production faults grounded on the outcomes of the ambiguity matrix which affirms that the model has capability and efficacy to discriminate among various classes of packaging quality with minimal mistake levels particularly in those classes that are termed as crucial.

The quality defect rate is a variable that reflects the final outcome of the interaction of a set of operational factors, and this variable has received the highest level of importance. This finding implies that the cumulative level of defects in the production process has a direct impact on packaging quality, which is consistent with the industrial sector's drive for zero defects as an objective of quality management systems. Other determinants were environmental conditions wherein humidity was ranked second in importance. Defects are more likely to occur at higher humidity levels because adhesion is reduced. The sensitivity of temperature variations in the packaging process concerning seal failures and spoilage about food products explains why temperature control would be so emphasized. It obtained from human element as automated production processes lessen human variability due to standardized operating procedures. This does not stop the role of human intervention, but rather its impact has relatively reduced compared to technical factors due to technological advancement and adoption of automation. Lastly, continuous improvement programs can be guided by taking advantage of the order of importance of variables as presented by the model. This helps in transformation towards proactive quality management if departments take a lead and make use of forecasting models.

In conclusion, it must be noted that the results show us that the quality of packaging in production processes is the result of three sets of operational, human, and environmental factors. To achieve the goal of quality management in reducing defects, it is necessary to focus on the factors with the highest impact without neglecting the rest of the factors, which contributes to supporting improvement decisions.

6. Conclusions

The theoretical framework articulates that the food industry should enhance the quality of production processes and control defects. High defects spoil packaged food products wasting resources used which eventually turns into financial losses. This led to a conclusion for the study that indeed machine learning techniques, particularly Random Forest can be used as an effective predictive model in identifying factors that cause defects through certain operational, environmental, and human variables. In this industry due to its peculiar nature dealing with daily consumption needs and their dangers of spoilage on consumer health, it works in packaging processes improvement within the food industry. As the research results confirmed, adopting a defect prediction model contributes to supporting decisions aimed at improving product quality and the production process through early prediction of non-conformities, thus achieving proactive quality management a shift from reactive quality management. It is worth noting that these models contribute to the transition towards smart, data-driven manufacturing systems and the use of machine learning in managing the quality of production processes.

REFERENCES

1. Alamri MS, Qasem AAA, Mohamed AA, Hussain S, Ibraheem MA, Shamlan G, Alqah HA, Qasha AS .(2021). Food packaging's materials: A food safety perspective. Saudi J Biol Sci. Aug;28(8):4490-4499. doi: 10.1016/j.sjbs.2021.04.047.
2. Alzate Montoya, P. M., HURTADO NIETO, B. D., & Gómez Jiménez, M. L. (2022). Gestión de la producción: evolución y tendencias de investigación. Revista Ingeniería, Matemáticas y Ciencias de La Información, 9(18), 29–46. https://doi.org/10.21017/rimci.2022.v9.n18.a118.
3. Asongo, A.I, Barma, M, Muazu, H.G. (2021) Machine Learning Techniques, methods and Algorithms: Conceptual and Practical Insights. International Journal of Engineering Research and Applications. Vol. 11, Issue 8, pp. 55-64. DOI: 10.9790/9622-1108025564.

4. Baierle, I.C.; Haupt, L.; Furtado, J.C.; Pinheiro, E.T.; Sellitto, M.A. (2024). Forecasting Raw Material Yield in the Tanning Industry: A Machine Learning Approach. Forecasting 2024, 6, 1078–1097. https://doi.org/ 10.3390/forecast6040054.
5. Bhardwaj, A., Alam, T., & Talwar, N. (2019). Recent advances in active packaging of agri-food products: a review. JOURNAL OF POSTHARVEST TECHNOLOGY. 7(1), 33–62.
6. Chen, W., Zhou, T., Niu, P., Peng, B., Wang, W., Liu, H., Ma, Z., Wen, Q., & Sun, L. (2023). eForecaster: Unifying Electricity Forecasting with Robust, Flexible, and Explainable Machine Learning Algorithms. Proceedings of the ... AAAI Conference on Artificial Intelligence, 37(13), 15630–15638. https://doi.org/10.1609/aaai.v37i13.26853.
7. Chen, Y. (2023). Food Preservation Packaging. IntechOpen eBooks. https://doi.org/10.5772/intechopen.110043.
8. Dudbridge, M. (2016). Handbook of seal integrity for the food industry. https://eprints.lincoln.ac.uk/id/eprint/22771/.
9. Fontes, L.C.B, K.K. Ramos, T.C.Sivi, F.P.C.Queiroz, 2011, Biodegradable edible films from renewable sources- potential for their application in fried foods, Am. J. Food Technology, 6:555-567.
10. Ghelani, H. (2023). Six Sigma and Continuous Improvement Strategies: A Comparative Analysis in Global Manufacturing Industries. International Journal of Scientific Research and Management. https://doi.org/10.18535/ijsrm/v11i08.ec05.
11. Han, J.-W., Han, J.-W., Ruiz-Garcia, L., Qian, J., & Yang, X. (2018). Food Packaging: A Comprehensive Review and Future Trends. Comprehensive Reviews in Food Science and Food Safety, 17(4), 860–877. https://doi.org/10.1111/1541-4337.12343
12. He, Y., Huang, H., Li, D., Shi, C., and Wu, S.J. (2018) Quality and Operations Management in Food Supply Chains: A Literature Review. Journal of Food Quality Volume 2018, https://doi.org/10.1155/2018/7279491.
13. Huynh, T. (2019). Packaging Standardization in Lean Manufacturing: An application of Part Standardization. https://www.theseus.fi/handle/10024/265370.
14. Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. Electronic Markets. https://doi.org/10.1007/S12525-021-00475-2.
15. Kamble, S. S., Gunasekaran, A., Gawankar, S. A. (2020) Achieving sustainable performance in a data-driven agriculture supply chain: A review for research and applications. International Journal of Production Economics. Vol. 219, Pp.179-194. https://doi.org/10.1016/j.ijpe.2019.05.022.
16. Kumar, D. (2023). The Use of Artificial Intelligence in Data Analysis and Business Intelligence (pp. 311–320). Routledge. https://doi.org/10.4018/979-8-3693-0418-1.ch020.
17. Kumar, P., & Singh, L. (2024). Analysis of the impact of artificial intelligence on big data analysis. International Journal of Engineering Science & Humanities. https://doi.org/10.62904/ber12n15
18. Kwon, Y., & Jeong, J. (2023). Production Planning Forecasting System Based on M5P Algorithms and Master Data in Manufacturing Processes. Applied Sciences, 13(13), 7829. https://doi.org/10.3390/app13137829.

19. Li, R., Qiao, Y., Li, H., Duan, L., Zhang, J., & Jing, Y. (2021). Research on Integrated Quality Control in Product Manufacturing Stage. DEStech Transactions on Materials Science and Engineering. https://doi.org/10.12783/DTMSE/AMEME2020/35535.
20. Matboli, M., Khaled, A., Ahmed, M. F., Manar Yehia Ahmed, M.Y., Khaled, R., Elmakromy, G.M., Ghani, A.M.A., El-Shafei, M.M.,. Abdelhalim, M.R.M. and Abd-El Gwad, A.M. (2025) Machine learning-based stratification of prediabetes and type 2 diabetes progression. Diabetology & Metabolic Syndrome https://doi.org/10.1186/s13098-025-01786-6.
21. Peres, F.A.P.; Bondarczuk, B.A.; Gomes, L.d.C.; Jardim, L.d.C.; Corrêa, R.G.d.F.; Baierle, I.C.(2025) Advances in Food Quality Management Driven by Industry 4.0: A Systematic Review-Based Framework. Foods **2025**, 14, 2429. https://doi.org/10.3390/foods14142429.
22. Phoemchalard, C., Senarath, N., Malila, P., Tathong, T., & Khamhan, S. (2024). Using orange data mining for meat classification: The preliminary application of machine learning. International Journal of Agricultural Technology, 20(6), 2497–2512. retrieved from https://li04.tci-thaijo.org/index.php/IJAT/article/view/5614.
23. Ramadhani, R. A., Fitriana, R., Habyba, A. N., & Liang, Y.-C. (2023). Enhancing Quality Control of Packaging Product: A Six Sigma and Data Mining Approach. Jurnal Optimasi Sistem Industri. https://doi.org/10.25077/josi.v22.n2.p197-214.2023.
24. Şahin, D., Torkul, O., Şişci, M., Diren, D. D., Yılmaz, R., & Kibar, A. (2025). Real-Time Classification of Chicken Parts in the Packaging Process Using Object Detection Models Based on Deep Learning. Processes, 13(4), 1005. https://doi.org/10.3390/pr13041005.
25. Saxena, M. M., & Rao, K. S. (2019). Quality Management, Total Quality Management And Six Sigma. International Journal of Scientific & Technology Research, 8(12), 394–399. https://www.ijstr.org/paper-references.php?ref=IJSTR-1019-23899.
26. Semercioz-Oduncuoglu, A. S., Luning, P. A. (2025) Industry 4.0 technologies in quality and safety control systems in food manufacturing: A systematic techno-managerial analysis on benefits and barriers. Trends in Food Science & Technology 163. https://doi.org/10.1016/j.tifs.2025.105144.
27. Smętkowska, M., Mrugalska, B. (2018). Using Six Sigma DMAIC to improve the quality of the production process: a case study. Procedia - Social and Behavioral Sciences 238 (2018) 590 – 596. doi: 10.1016/j.sbspro.2018.04.039.
28. Vasile, C., & Baican, M. (2021). Progresses in Food Packaging, Food Quality, and Safety-Controlled-Release Antioxidant and/or Antimicrobial Packaging. Molecules, 26(5), 1263. https://doi.org/10.3390/MOLECULES26051263.
29. Vasile, C., & Baican, M. (2021). Progresses in Food Packaging, Food Quality, and Safety-Controlled-Release Antioxidant and/or Antimicrobial Packaging. Molecules, 26(5), 1263. https://doi.org/10.3390/MOLECULES26051263.
30. Whardana, G. P., Irwati, D., & Miharja, M. (2024). Analysis of Carton Packaging Quality Control Using Statistical Quality Control Methods at PT XYZ. International Journal of Innovative Science and Research Technology, 212–221. https://doi.org/10.38124/ijisrt/ijisrt24jul580.

31. Wu, X., Wu, K., Hao, S., Yu, E., Zhao, J., & Li, Y. (2025) Machine learning ensemble technique for exploring soil type evolution. Scientific Reports. 15:24332. https://doi.org/10.1038/s41598-025-10608-8.
32. Yagoubi, A., Sedaghat, N., & Ghanbarzadeh, T. (2013). Assessing Safety of Food Packaging Materials. https://profdoc.um.ac.ir/articles/a/1038241.
33. Yan, M. R., Hsieh, S., & Ricacho, N. (2022). Innovative Food Packaging, Food Quality and Safety, and Consumer Perspectives. Processes, 10(4), 747. https://doi.org/10.3390/pr10040747.
34. Yan, M. R., Hsieh, S., & Ricacho, N. (2022). Innovative Food Packaging, Food Quality and Safety, and Consumer Perspectives. Processes, 10(4), 747. https://doi.org/10.3390/pr10040747.
35. Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23(100224).https://doi.org/10.1016/j.jii.2021.100224.