Monitoring and Controlling System for Mango Logistics Based on Machine Learning

Buyung Achmad Hardiansyah¹, Heru Sukoco², Sony Hartono Wijaya³
¹²²²Department of Computer Science, IPB University, Bogor, Indonesia
¹buyung.hardiansyah@gmail.com, ²hsrkom@apps.ipb.ac.id, ³sony@apps.ipb.ac.id

Abstract

Fruits are highly perishable goods, which means they have a short shelf life and can pose significant challenges in trade. A long supply chain can trigger the process of fruit spoilage. The logistics environment, both internal and external, can also affect the decrease in quality of goods. One common issue facing producers is the variability in consumer demand for fruit quality. To address this problem, a machine learning-based logistics monitoring and recommendation system can be developed, utilizing the Long Short-Term Memory (LSTM) and Decision Tree algorithms. Using machine learning algorithms, the system can analyze data from devices equipped with the Internet of Things (IoT), such as temperature and humidity sensors, to identify potential issues in the supply chain and provide recommendations to optimize logistics operations. In this study, a machine learning-based monitoring system is developed to monitor the shelf life of perishable goods, with a specific focus on mango fruit. The system utilizes LSTM to predict mango ripeness and decision tree algorithms to recommend fruit ripeness. The objective is to provide producers with recommendations that optimize the logistics process for high-quality mangoes and meet the consumer demands for quality fruit. The implementation of a machine learning-based logistics monitoring and recommendation system can provide significant benefits to mango producers. Using advanced technologies, such as LSTM and Decision Tree algorithms, producers can optimize their logistics operations, improve fruit quality, reduce waste, and improve customer satisfaction.

Keywords: decision tree; LSTM; machine learning; mango logistics


1. Introduction

Fruits are highly perishable products with a relatively short shelf life and their cultivation is limited to specific areas with varying levels of durability. As a result, fruit distribution often leads to a decrease in quality and damage before reaching the customer [1]. A lengthy supply chain can trigger the fruit spoilage process, and poor management decision-making can result in the loss of value of the product. The logistics stages, which include storage and transportation, are crucial activities in the fruit supply chain and are significantly affected by temperature and time.

These two factors can determine changes in fruit quality, and therefore optimal storage temperatures must be maintained during the logistics process. By controlling the environmental conditions during transportation and storage, the quality of the fruit can be preserved and the shelf life can be extended. This can result in higher value for the product and better customer satisfaction [2].

The logistic environment, both internal and external, can affect the decline in the quality of goods [3]. Temperature and humidity are critical factors that can significantly affect the quality of perishable goods during logistics operations. Therefore, to maintain the quality of these goods, it is important to monitor the temperature, humidity, and other relevant storage parameters in real time [4].

In recent years, real-time monitoring systems for perishable goods have been widely developed. In past research, an automated infrastructure has been presented to monitor the supply chains of perishable goods and manage their functionality. The system utilizes wireless sensor network technologies for the monitoring of goods and parallel 'real-time' processing for the prediction of shelf life [4].
A real-time monitoring system has been developed to estimate quality losses for perishable products during storage and transportation using a wireless network and regression model [5]. Past research has focused on developing a real-time monitoring system that integrates IoT sensors, a big data platform, and data mining methods. The system uses smartphone-based sensors, MongoDB, and outlier detection methods to create an efficient and cost-effective platform to monitor environmentally sensitive agricultural food products [6].

This system is designed to provide recommendations to producers and allow real-time monitoring of the logistic environment, encompassing factors such as temperature and humidity. In addition, it aims to predict the ripeness of mango fruit throughout the logistics process, thus facilitating producers in optimizing transportation and storage procedures to ensure the delivery of high-quality mangoes that align with consumer demands for fruit quality. The novelty of this research lies in the ability of the system to effectively monitor and predict the level of ripeness of mango fruit during logistics transport by employing machine learning techniques that analyze the temperature and humidity of the logistics environment.

The mango logistic recommendation system was developed using a decision tree algorithm. A decision tree is a classification model similar to a tree-shaped decision structure consisting of a collection of points containing test conditions (internal nodes) and class labels (leaf nodes) [7]. Decision trees are used to build predictive models in machine learning, data mining, and statistics. The decision tree algorithm uses collected observations to predict the value of an object through a series of decisions [8].

A decision tree is a type of supervised learning prediction model that aims to predict a predetermined target value. Decision trees can work with categorical and continuous input and output variables. They can be used for classification with a categorical target variable or for regression with a continuous target variable. In the context of decision trees, there are three types of nodes: root nodes, internal nodes, and leaf nodes. These nodes are illustrated in Figure 1 [9].

Previous research has used decision tree algorithms and achieved good accuracy in related areas. A study was conducted to compare the performance of logistic regression, decision tree, and neural network models for predicting traffic congestion. The results showed that the decision tree model outperformed the other two methods, achieving an accuracy of 97% [10]. The proposed work to detect network intrusions using a decision tree classifier demonstrated good performance, with higher accuracy, recall, and F1 score, as well as shorter execution times [11]. A decision tree algorithm is used to predict the transhipment schedule and the resulting model fits well, with a determination coefficient of 0.73 [12].

In this system, we also use the Long-Short-Term Memory (LSTM) algorithm to develop a system that predicts mango quality and monitors environmental conditions during mango fruit logistics. LSTM is a type of repetitive neural network that is capable of capturing long-term dependencies and efficiently learning from input sequences of varying length [13]. Long-Short-Term Memory (LSTM) is a type of neural network that was developed to address the problem of learning long-term dependencies. It was first introduced by Hochreiter and Schmidhuber in 1997 as a type of processing module for recurrent neural networks (RNNs).

An LSTM network consists of repeating LSTM units that interact with each other to model sequences and recognize patterns over long periods [14]. There are several unit architectures used in LSTMs, the most commonly used ones being forget gates, input gates, and output gates. The forget gate is responsible for deciding which information will be discarded from the cell state, while the input gate decides which new information will be stored in the cell state. The output gate determines which information will be output from the cell state. The cell state in an LSTM stores a value or state for a short or long period. An example of the LSTM architecture can be seen in Figure 2 [15].

![LSTM Architecture](image)

**Figure 1. LSTM Architecture**

LSTM has been used in several studies for prediction and produces good accuracy. Prediction of air quality with concentrations of CO, NO2, O3, PM10 and SO2 in the city of Madrid [14]. The LSTM RNN model is an adequate interpolation tool to predict the efficiency of
nutrient loss from the A-A-O MBR process in wastewater treatment systems [13]. The LSTM RNN model is used for online monitoring and prediction of sintered chemical composition, this method performs quite well with an R2 value of more than 0.92 and the mean square error (MSE) and mean absolute error (MAE) are close to zero [16].

This research aims to develop a machine learning-based system that can monitor and control mango transport from the point of maturity determination to the point of delivery to customers. The system will provide recommendations for the optimal mode of transport, monitor and predict the quality of the fruit during the logistics process, and suggest treatments to maintain the quality of the mangoes for optimal conditions upon arrival. Ultimately, the system aims to ensure that the quality of the fruit is maintained throughout the logistics process to meet customer demands.

2. Research Methods

Our research methodology adopts a systematic approach to develop machine learning models that are integrated within back-end systems, thereby delivering output through web and mobile applications.

The research involved several stages. The initial stage of planning was to determine the necessary software components and functionalities required for the research project. Subsequently, relevant data on mango fruit logistics, including temperature, humidity, and ripeness levels, were acquired and prepared for analysis. Through these data, two machine learning models were developed using the decision tree algorithm for recommendation and the LSTM algorithm for predictions.

Following the development of these models, the system design phase began, involving a use case diagram, class diagram, sequence diagram, and the overall architecture and integration of the developed models. The subsequent system implementation phase included integrating the prediction and recommendation models, as well as creating web and mobile applications to display the system's output. Finally, the implemented system was tested in system tests to ensure its functionality, accuracy, and reliability in predicting mango quality and monitoring environmental conditions during the logistics transportation process. A visual representation of the research stages can be found in Figure 3.

2.1 Planning

The planning stage aims to identify software requirements and outline the system to be built. The expected output from this stage is a clear understanding of the requirements of the system and the business processes involved.

The system is expected to have the ability to monitor and control mango logistics, starting from shipping planning, determining the quality of mango fruit, and managing transportation logistics. It also monitors environmental conditions during the logistics period and allows for the prediction of mango fruit maturity, ensuring the maintenance of fruit quality until it reaches the customer.

2.2 Data Acquisition

The monitoring and control system will be equipped with a prediction model that uses machine learning to provide information if there is an anomalous process during the logistics transport period.

The data used in the machine learning model will be obtained by observing the ripeness of mangoes using a device equipped with temperature and humidity sensors in the logistic transport of mangoes. The data collected consist of sensor device data on the logistics of mango fruit transport during its journey from Surabaya to Jakarta, as well as from Surabaya to the city of Balikpapan.

2.3 Data Preprocessing

Data pre-processing includes tasks such as cleaning, removing unnecessary features, and transforming the data as needed. Once the data have been pre-processed, they will be divided into training and test datasets to build and evaluate the machine learning model.

2.4 Training Process

After the pre-processing step, the data are divided into training and test datasets. The training dataset is used as input data to generate the recommendation model using the decision tree algorithm and the prediction model using LSTM.

2.5 Model Evaluation

The decision tree model is evaluated using a performance evaluation matrix that includes precision, recall, accuracy, and the F1 score. A higher value of the evaluation matrix indicates better performance of the decision tree model [17]. The confusion matrix is used to visualize the distribution of the number of actual classes compared to the predicted classes.
To assess the performance of the LSTM model, we use the Root Mean Square Error (RMSE), which is an alternative method for evaluating prediction techniques and measuring the accuracy of a model’s forecast results. The RMSE value represents the square root of the average of the squared errors between the predicted values and the actual values in the regression model. A lower RMSE value indicates that the predicted values are closer to the actual values and therefore the model has better accuracy. This technique is easy to implement and has been widely used in various studies related to forecasting and prediction [18].

2.6 Design System

The system comprises a web application and a mobile application developed using the Flutter programming language. This system has three main features, including a prediction and monitoring system, a mango ripeness recommendation system, and transportation logistics recommendations. The use case diagram for monitoring and controlling mango quality can be found in Figure 4.

Class diagrams are graphical representations that illustrate the relationships and structure between classes or objects in a software system, including the attributes and operations associated with each class [19]. Figure 5 shows the class diagram of the mango transportation logistics back-end system, which includes four classes with distinct functions. One of the classes, 'Locations,' is responsible for retrieving location data, such as city names, latitude, and longitude attributes.

The 'Distance' class includes the necessary operations to calculate the distance between two geolocations based on their latitude and longitude values. The 'Monitoring' class is responsible for obtaining results from the prediction model generated using the LSTM algorithm. Finally, the 'Recommendation' class includes an operation to recommend the optimal maturity level of mangoes and suggest the most suitable logistics for transporting the mangoes.

2.7 Implementation System

The system was built and implemented using both hardware and software components. Specifically, the
machine learning models were developed using the Python programming language and the TensorFlow library, with the Keras API and sci-kit-learn. A PostgreSQL database was used for data storage, while the front-end system was developed using Flutter, and the back-end was developed using the Python programming language.

2.8 System Testing

The testing process utilizes the black-box testing method to observe the results of the system execution and ensure that the system runs as intended. A black-box testing technique is used to assess the functionality of the software by observing the input and output results of the system. This testing method is used to test the interface of a component with other components and systems [20].

3. Results and Discussions

The developed system comprises a web application and a mobile application created using the Flutter programming language. This system encompasses three key features, including a prediction and monitoring system, a mango ripeness recommendation system, and transportation logistics recommendations integrated with machine learning models.

The intended purpose of the developed system is to enable the monitoring and management of mango logistics, from the planning of shipping, the assessment of the quality of mango fruit and the management of transportation logistics. In addition, it monitors environmental conditions throughout the logistics process and facilitates the prediction of mango ripeness, guaranteeing the preservation of fruit quality until it reaches the customer.

3.1 Recommendation Model

The mango logistics recommendation system is developed using machine learning techniques, specifically employing the decision tree algorithm. This system offers recommendations for the initial maturity of mango fruits and logistics transportation. By considering these recommendations, mango fruit producers can ensure that the fruit reaches customers with the desired level of ripeness. The system enables producers to specify the quality of the initial mango fruit and to make informed decisions about logistics transport based on the recommendations provided by the system.

The recommendation model uses logistic data on the transport of mango fruit equipped with temperature and humidity sensors from the city of Surabaya to the cities of Jakarta and Bali. The recommendation model uses logistic data on mango fruit transport, incorporating temperature and humidity sensors. The collected data consists of various attributes, including temperature, humidity, initial mango ripeness, mango ripeness during the trip, travel time, and mode of transportation.

The next stages are the data preprocessing stage and the data analysis stage. The pre-processing stage involves converting the data type of the transportation mode attribute from object to numeric and selecting the relevant features to be used in the model. The data analysis stage focuses on understanding the underlying patterns and characteristics of the dataset.

We analyze logistics transportation data for mangos, including the following attributes: transportation period, initial mango ripeness, mango ripeness upon arrival at the destination, type of transportation, and distance between the starting point and destination. To continue our analysis, we explore the characteristics and distribution of the data. This data exploration stage is critical to determining the necessary preprocessing steps. Descriptive statistics, such as standard deviation, variance, minimum, maximum, mean, and median values, are used to provide information on each attribute [21].

Table 1 shows that the time in hours attribute has the highest variance among the attributes, ranging from a minimum value of 1 to a maximum value of 115, with a standard deviation of 32.71. On the other hand, the initial maturity attribute has the smallest.

Table 2 shows the entropy and information gain values. Entropy is a measure of the amount of uncertainty associated with a random variable. The entropy increases as the level of uncertainty or randomness increases and decreases as the uncertainty decreases. Information gain is the difference between the original level of information required to classify the data correctly and the new level of information required after splitting the data using a decision tree [21]. The time-in-hours attribute has the highest information gain value, indicating that it will serve as the root of the decision tree model. The preprocessing stage involves cleaning the data by removing empty values. The data are then divided into 80% training data and 20% test data, with the training data used as input for the decision tree model.
To generate the recommendation model, we used the decision tree algorithm in the Python programming language and the sklearn library. The model parameters, including maximum depth, minimum sample split, and criterion, were specified during the model generation process.

To determine the optimal model, we executed a hyperparameter tuning process using k-fold cross-validation and grid search methods to determine the best value for the maximum depth parameter. The hyperparameter tuning process involved evaluating the mean square error value for maximum depth values ranging from 1 to 10, as shown in Figure 7. The x-axis in Figure 7 represents the maximum depth values, while the y-axis represents the mean square error values. The graph indicates that the maximum optimum depth value is 4, which was selected to develop the decision tree model.

The next step is to determine the optimal values for the min sample split and the criterion parameters using the grid search cross-validation method. The grid search process involves evaluating different combinations of the maximum depth parameter in the range of 1 to 10, and the minimum sample split parameter in the range of 10 to 60. This process will help us identify the optimal combination of parameters for the decision tree model.

The test data were applied to the decision tree model and performance was evaluated using precision, recall, accuracy, and F1 score metrics. The model achieved an accuracy score of 0.92.

3.2 Prediction Model

The mango quality prediction system uses the Long-Short-Term Memory (LSTM) algorithm to develop a system that can predict the quality of mango fruit and monitor environmental conditions during the mango logistics process. The system can predict the quality of mango fruit during the transport period based on factors of temperature and humidity during mango transport. Fruit producers are expected to be able to effectively monitor the logistics environment, including real-time temperature and humidity conditions.

Furthermore, the system enables the prediction of mango quality by utilizing temperature and humidity parameters, along with travel time, for an efficient mango distribution. The prediction model uses monitoring data for mango logistics transportation,
The attributes used for the modeling are those that have a significant impact on the respiration process of mangoes during transport. These attributes include external factors such as temperature and humidity in the environment. By incorporating these attributes into the model, we can better predict the maturity of the mango upon arrival at its destination [22].

The next step is data preprocessing, which involves converting unstructured data into structured data. The data preprocessing stage includes selecting relevant features, converting date and time values to datetime format, and combining date and time values to create datetime columns. The resulting dataset is visualised in Figure 9. Finally, the data are divided into 80% training data and 20% test data for model training and evaluation.

To generate the prediction system model, we utilized the long-short-term memory (LSTM) algorithm, implemented using the Keras API in the Python programming language. The LSTM architecture was designed with 200 LSTM layers, 1 dense layer, an Adam optimizer, 500 epochs and a batch size of 72.

The performance of the LSTM model was evaluated using the root mean square error (RMSE) metric. The RMSE value obtained was 0.06116661, indicating good performance of the model. Furthermore, the model demonstrated a strong correlation value.

Figure 10 shows the training loss and the validation loss values during model training. At epoch 20, the validation loss values began to decrease, and the optimal RMSE value of 0.06116661 was achieved. Figure 11 provides a comparison between the actual and predicted values.

3.3 Implementation System

The system is composed of a front-end that can be accessed via a web or mobile application and a back-end with a REST API architecture. The back-end uses the PostgreSQL database and incorporates embedded machine learning models, as illustrated in Figure 12.

The front-end was developed using the Flutter programming language, while the back-end was developed using the Python programming language with the Django framework. The front-end communicates with the back-end via web services using a REST API architecture. In general, an API defines a set of data and functions that facilitate the interaction between computer programs and allow them to exchange information [23].
Figure 12. Main Page

Figure 13 shows the main page of the mobile application, which includes three menu buttons: the transportation recommendation system, the mango ripeness recommendation system, and the monitoring system.

Figure 14 shows a form page to recommend the initial maturity level of the mango fruit. To use this form, the user first inputs the starting and destination points using a text search input. The user then enters the final maturity level of the mangoes, the travel time in hours and the mode of transportation using the numerical and dropdown input fields. Once the user has entered all the required information, the system validates the inputs and displays the results of the initial maturity level recommendation for the mango fruit.

The front end communicates with the back end through an API endpoint, which incorporates a mango recommendation model that was trained using the decision tree algorithm described previously. The front-end then displays the results of the initial ripeness recommendation to the user.

Figure 15 shows a logistic transportation monitoring device for mangoes, which includes sensors to measure temperature and humidity, as well as predictions of mango ripeness. This feature presents real-time monitoring of the logistics environment, including temperature and humidity, through a graph display. Data are obtained by receiving information from sensors that are integrated into the logistics transport system. Furthermore, the system can predict the ripeness of the mango fruit during the transportation period using temperature and humidity parameters.

The front-end display periodically retrieves data through an API from the back-end system, which incorporates machine learning techniques using the LSTM algorithm. The graph in Figure 15 is generated from the API response, which incorporates an LSTM machine learning model embedded in the system.
3.4 System Testing

To ensure that the system runs as expected, we performed system testing using the black-box testing method. We tested the system with seven different scenarios and the results demonstrated that each test scenario produced the expected results and ran smoothly.

By implementing this machine learning-based monitoring and control system, significant benefits can be achieved for producers of perishable goods, particularly mangoes. Using advanced technologies such as LSTM and Decision Tree algorithms, along with mobile and web applications, producers can optimize their logistics operations, improve fruit quality, minimize waste, and improve customer satisfaction. This system empowers producers to make data-driven decisions and efficiently manage their supply chain, resulting in improved overall performance and profitability.

4. Conclusions

Fruit is a perishable item that requires special treatment during distribution so that the quality of mango fruit can be maintained when it reaches customers. The variety of customer requests regarding the ripeness of mango fruit, when they arrive at their destination, can be accommodated by determining the initial ripeness of mango fruit and determining the appropriate transportation to save costs by using a decision tree algorithm. This can help producers determine the ripeness of the early mangos so that the quality of the fruit is maintained and it reaches customers according to demand. The logistics recommendation model, which uses the decision tree algorithm, can provide accurate recommendations for the initial maturity level of mangos with a 92% accuracy rate.

The logistics environment, both internal and external, can significantly affect the quality of goods. Therefore, it is crucial to monitor temperature, humidity, and other relevant storage parameters in real time. This system can monitor the logistics environment, including temperature and humidity, by receiving data from sensors integrated into logistics transportation in real time. Furthermore, it can predict mango ripeness during the transportation period using temperature and humidity parameters. The mango ripeness prediction model, which uses the long-short-term memory (LSTM) algorithm, can accurately predict the level of mango ripeness during transportation. The evaluation results using the root mean square error (RMSE) metric showed a high degree of precision, with a value of 0.06116661.

The recommendation and prediction systems can be accessed through the Web and mobile interfaces. The REST API architecture facilitates information communication between the front-end, back-end, and IoT sensor devices. Using this system, producers can optimize their logistics operations, improve fruit quality, minimize waste, and improve customer satisfaction. It empowers producers to make data-driven decisions and efficiently manage their supply chain, leading to better overall performance and profitability.

This system may still have weaknesses in terms of the internet communication network, where the REST API architecture still relies on a stable internet connection. In certain cases, some areas lack internet network coverage, which can lead to potential failures in the transmission of data from logistics transport to the server. As a result, data sent from logistics transportation to the server may encounter interruptions or failures at certain points during mango distribution.

For future research, machine learning prediction and recommendation models could be further optimized by comparing them with other algorithms to achieve even better results. Furthermore, the monitoring and control system for mango logistics could be improved by adjusting the web-based display to match different resolutions on the monitor screen, implementing a user access control list (ACL) to improve security, and grouping IoT devices in logistics transportation to allow monitoring of multiple transportation logistics.

Acknowledgments

The data used in this research come from the RISPRO LPDP (Research Proposal of the Indonesia Endowment Fund for Education).

References


