Journal of Theoretical and Applied Information Technology

30th November 2023. Vol.101. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

APPLICATION OF MACHINE LEARNING ALGORITHMS FOR FORECASTING SALES OF SHRIMP SEEDS

DEVI ALFITRIA ANGGRAENI[1], NILO LEGOWO[2]

[1,2]Information Systems Management Department, BINUS Graduate Program – Master of Information System Management, Bina Nusantara University, Jakarta, Indonesia 11480

E-mail: [1]devi.anggraeni@binus.ac.id, [2]nlegowo@binus.edu

ABSTRACT

Indonesia has a great potential for marine and fisheries, particularly in shrimp cultivation. There are many types of shrimp cultivated by Indonesian farmers, but Vannamei shrimp is one of the main types of shrimp in Indonesia, highly demanded both domestically and for export. With such a large demand for shrimp, shrimp farmers need a supplier of shrimp seeds so they can be cultivated into shrimp. PT Prima Akuakultur Lestari, a shrimp seed supplier, who has been in operation for more than 5 years and can produce 130 million shrimp seeds per year. However, this company faced challenges in achieving their sales targets. Sales targets are set based on the sales history of previous months. However, companies are still manually determining their sales targets so that targets are often not achieved. Machine learning is believed to be a method strategy to predict future shrimp sales. Before using Machine Learning, companies must first know which Machine Learning model is suitable for their use. The aim of this research is to test several machine learning models to find a model that is suitable for companies to use in forecasting shrimp seed sales. The research utilized the CRISP-DM method and analyzed shrimp seed sales data from March 2017 to August 2022. Three machine learning models were tested: K-Nearest Neighbor (KNN), Support Vector Regression (SVR), and Neural Network (NN). RapidMiner software was used for data analysis. The results obtained from this research show that the K-Nearest Neighbors model has the highest accuracy value among the three other models tested, with an RMSE value of 6326408.735 and R^2 0.215.

Keywords: *Machine Learning, Forecasting, Shimp Seeds, Sales, CRISP-DM*

1. INTRODUCTION

Indonesia is the largest archipelagic country in the world which includes large and small islands with a total of more than 17,000 islands with an area of 5,193,250 km² (1). Its territory is divided into 1.9 million km² of land and 3.2 million km² of sea (2). Indonesia has extraordinary marine and fishery potential due to its vast sea area (3).

One of the breakthrough programs of the Ministry of Marine Affairs and Fisheries (KKP) is the development of export-based aquaculture with shrimp as one of the national leading commodities (4). Shrimp is one of Indonesia's mainstay fishery commodities which has great export potential. As one of the national leading commodities, shrimp has always been the choice to be involved in efforts to increase national income and achieve the target of increasing production by up to 250% by 2024. Currently shrimp ponds are spread in almost all corners of Indonesia.

In Indonesia itself there are various types of shrimp seeds. However, based on the records of the Fish Quarantine and Quality Supervision Agency of the Ministry of Maritime Affairs and Fisheries (BKIPM-KKP) it shows that Vannamei Shrimp seeds are the most widely traded type of shrimp seeds in the country (between provinces). The number of vannamei shrimp seeds traded in 2017 was superior to other shrimp seeds, reaching 15.87 billion heads or around 72.81% of the total national fish seed traffic (5).

The Fish Quarantine and Quality Control Agency of the Ministry of Maritime Affairs and Fisheries (BKIPM-KKP) also noted that there were five provinces supplying national vannamei shrimp seeds, namely Lampung Province (37.84%), East Java Province (24.49%), Banten Province (17 .81%), Bali Province (12.76%) and South Sulawesi Province (5.92%) (5).

With the current era of digital transformation, every company or organization increasingly needs an information system to improve the quality of the company's/organization's internal information flow (6), manage daily operations, integrate data (7), and build cooperation with

external parties. Companies/organizations need an integrated and unified Management Information System (7).

Management information systems are different from ordinary information systems because SIM is used to analyze other information systems that are applied to company/organization operations (8).

A management information system is a unit that includes the use of human resources, documents, technology, and procedures in an integrated and integrated manner designed to collect, process, store and disseminate information related to automation or decision-making support by humans, for example Decision Support Systems in management and activities within an organization (9).

There are several companies supplying vannamei shrimp seeds in Lampung Province, one of which is PT Prima Akuakultur Lestari, it is a company engaged in the field of cultivation, especially the cultivation of vannamei shrimp seeds from 2017 until now. This company has transacted more than millions of shrimps seeds every month, so it can be said that the shrimp seed transactions carried out by this company are quite good.

However, the company experienced a problem, namely not being able to meet predetermined sales targets. Because the goods sold by the company are living things, the company cannot store the goods sold. Determining the amount of inventory that is too small will result in the loss of the company's opportunity to gain more profit by selling more shrimp seeds. However, determining the amount of excessive inventory will result in the company discarding existing shrimp seeds.

To overcome this, companies need to make optimal planning, both production planning and sales planning. Good communication between companies and consumers is needed to find out the demand from consumers, so that it can assist companies in determining the number of products to be produced.

The level of customer satisfaction is one of the main factors that must be considered by the company, with a feeling of satisfaction with a product/service, the customer will automatically return to buy or use the product (10) even customers will help to promote the product to his closest relative. The best company must be able to meet the needs of its customers, because the quality of service provided by the company is needed to provide customer satisfaction.

Several studies have shown that in a business strategy, retaining old customers is more profitable than attracting new customers (11). Loyal customers reflect customer satisfaction, where price is not the main factor for them to buy goods/use company services but is something that is relative (12).

It is more important to establish good relations with customers than to make customers the company's sales targets. If the company's relationship with the customer is good, whatever the company offers, the customer will prefer that company because they are sure of their satisfaction (13).

Therefore, it is necessary to implement a better strategy that can help this business become more effective in improving marketing and customer support (13). One way to build and maintain relationships with these customers is to implement a Customer Relationship Management system.

Customer Relationship Management is a corporate strategy that is of particular concern to companies because of the inseparable relationship between the Customer Relationship Management function and customer satisfaction and loyalty.

Customer Relationship Management is currently one of the strategies used by companies to build, manage, strengthen company and customer relationships, and get to know and understand their customers better. So that companies can provide the best service and build and establish better long-term relationships with their customers (14).

The company several times failed to achieve sales targets. To overcome this, companies need to make optimal planning. Good communication is needed between the company and the customer to find out about the requests from customers. Then it also requires good planning by the company in determining the number of sales targets based on customer requests. One way that can be used to overcome this problem is to predict shrimp seed sales to be able to respond to fluctuations in market demand and meet sales targets.

According to Niladri Syam and Arun Sharma (15) machine learning has proven successful in predicting sales. Forecasting processes can be carried out using Machine Learning, a branch of Artificial Intelligence (AI) and computer science that focuses on using data and algorithms to mimic the way humans learn, and gradually improve their accuracy (16).

Forecasting using machine learning to determine the number of shrimp sales in the following month is very important for companies. Because by knowing the number of sales that will occur, companies can be more precise in planning shrimp seed cultivation. So that the company does

not lose the opportunity to obtain maximum profits. Currently, the company still uses manual methods to determine the number of shrimp seeds that must be cultivated, and this results in the sale of shrimp seeds being ineffective.

In this research the author will carry out experimental testing on several models to determine which model is most appropriate for forecasting shrimp seed sales using machine learning. So that the company can maximize the profits it will obtain. After the author finds the right model, the company can use the model for use in forecasting.

This research will use K-Nearest Neighbor, Support Vector Regression, and Neural Network as research models.

2. LITERATURE REVIEW

2.1 Relevant Theories

2.1.1 Forecast

Forecasting is a process that is carried out to estimate or predict something that will happen in the future by processing historical data in the past or present (17). Forecasting has an important role in almost all fields, be it business intelligence, politics, economics, meteorology, finance, and telecommunications (18).

2.1.2 Data Mining

Data mining is a process carried out by companies to convert raw data into useful information (19). Data mining is the process of mining and processing data to gain knowledge or find certain patterns from many data sets and transform that information into an understandable structure for further use (20).

Data mining helps companies improve decision making by providing information that can be used according to predetermined data mining objectives, for example making predictions based on past trends and current conditions, developing more effective marketing strategies, increasing sales, and reducing costs (21).

This process can be done using software with the help of statistical calculations, mathematics, database technology, and the use of artificial intelligence technology (22). Data mining is also known by other terms such as Knowledge Discovery in Database (KDD), Data/Pattern Analysis, Knowledge Extraction, Data Archeology, and others (23).

2.1.3 CRISP-DM

In data mining, there are several models of data analysis techniques that can be used by data practitioners, one of which is the Cross-Industry Standard Process for Data Mining (CRISP-DM) (24). CRISP-DM is a data mining method co-developed by Daimler Chrysler (formerly Daimler-Benz), SPSS (formerly ISL), NCR Corporation, OHRA, and Teradata in 1996 (25). Then it was developed by hundreds of organizations and companies in Europe in various workshops between 1997-1999 [23] and finally the first version of CRISP-DM was published in 1999 (26).

CRISP-DM has been established as the de facto standard for developing data mining and knowledge discovery projects (26) and can be used across a wide range of industry sectors and a variety of tools (27). In addition to the application of CRISP-DM to the data mining process, the choice of algorithm also greatly influences the performance comparison of data mining methods (28).

CRISP-DM is not the only standard in data mining, but it is the most widely used methodology for data mining projects according to many polls (29). Based on the results of a poll conducted by datascience-pm, almost half of the respondents most often use CRISP-DM (30).

This methodology was developed to guide the most common steps in a data mining project (26). This methodology defines a non-rigid sequence of six life cycle phases (31) with arrows indicating the most important and frequent dependencies between phases (32). In fact, most projects move between phases as needed because CRISP-DM is flexible and adaptable (32). For data that can be processed with CRISP-DM there are no specific requirements or characteristics because the data will be processed again at the stages in it (33).

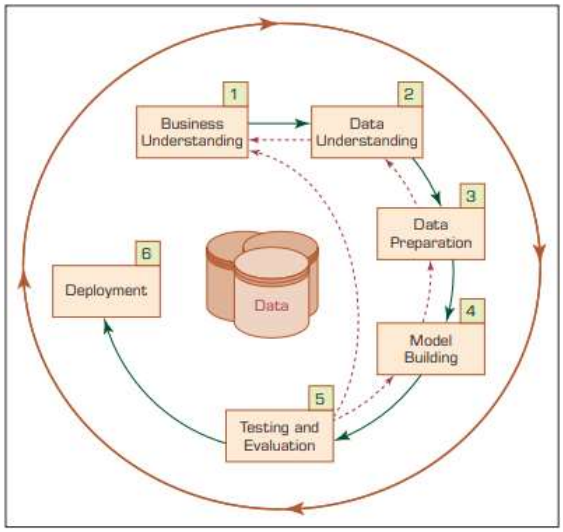

Figure 1: Six Step CRISP-DM Data Mining Process

Figure 1 depicts the six phases of CRISP-DM and their interactions (34). The following is a brief explanation of each phase (33):

1. Business Understanding

Understand the goals and needs of the project from a business point of view, then assess the current conditions to have a limitation of existing problems, then determine plans and strategies to achieve data mining goals.

2. Data Understanding

Initial data collection, if the data comes from more than one database, then the process of integration, data description, data exploration, and data quality evaluation is carried out. Data exploration is also carried out to detect interesting parts of the data that can be used to form hypotheses for hidden information.

3. Data Preparation

This stage includes all activities to build the final dataset (data to be processed at the modeling stage) from raw data. This stage can be repeated several times. Some of the things that are done at this stage include cleaning data, selecting tables, records, and attribute data (data selection), as well as transforming data and creating new attributes to be used as input at a later stage.

4. Modelling

At this stage it directly involves Machine Learning to determine data mining techniques, data mining tools and data mining algorithms. Several modeling techniques were selected and applied, and some parameters were adjusted to obtain optimal values. There are several different techniques that can be applied to the same data mining problem. Often someone realizes there is a data problem when doing modeling or getting an idea to build new data, so that, if necessary, data preparation can be done again so that the data can be in accordance with the required data mining technique.

5. Evaluation

This stage is carried out by looking at the performance level of the pattern produced by the algorithm. At this stage, one or more models are formed which seem to have high quality from the point of view of data analysis. Evaluate the model obtained more thoroughly before use and determine whether the model can achieve the goals set at the initial stage (business understanding). Determine if there are important business issues that are not being handled properly. At the end of this phase, a decision must be made about the use of the data mining results.

6. Deployment

This stage is carried out by creating and presenting simple reports so that users can use them using the resulting model. Moreover, the resulting model is implemented in a real-time environment.

If the model obtained is not good enough to be used to support business, then a new iteration is determined for CRISP-DM (31).

2.1.4 Machine Learning

Machine Learning is a machine capability that focuses on using data and algorithms developed to be able to learn and imitate human behavior without being programmed explicitly (35) and gradually increase its accuracy through experience (36). Machine Learning is a branch of Artificial Intelligence which is based on the idea that machines can collect knowledge from data, find hidden patterns, and make decisions with little human involvement (37).

Machine Learning draws on ideas from various disciplines such as Artificial Intelligence, probability and statistics, computer science, information theory, psychology, control theory, and philosophy (38). The adoption of machine learning methods has been used and applied in almost various aspects of life such as health, economics, education, marketing, and others (39).

The process starts with inputting good quality data and then training the machine (computer) by building a Machine Learning model using the data and different algorithms to solve the problem. The type of algorithm used depends on the type of problem to be solved, the number of variables, and so on (35).

Machine Learning and Data Mining are intertwined. Data Mining uses Machine Learning algorithms to analyze databases and find hidden information in data, while many machine learning algorithms use data mining methods to process data before learning the necessary tasks (38).

2.2 State of Art

The first research was written by Nasim Sadat Mosavia, Francisco Freitasa, Rogério Piresb, César Rodriguesb, Isabel Silvab, Manuel Santosa, Paulo Novaisa with the title “Intelligent energy management using data mining techniques at Bosch Car Multimedia Portugal facilities” (40). The aim of this research is to predict future demand for electrical energy consumption at Bosch Car Multimedia Portugal facilities. In addition, the influence of indicators such as humidity and temperature on electrical energy consumption is also investigated. The algorithms used are Random Forest, Decision Tree, Support Vector Machine, Linear Regression, and K-Nearest Neighbor. The results of this research are that the Random Forest algorithm as a regression approach has potential in making predictions, but FB

Prophet is the technique with the most potential in estimating future electrical energy needs related to E333 and E260 which are HVAC sensors. FB Prophet has the lowest RMSE value with a value of 8.07. FB Prophet is selected as a potential approach to forecast future electrical energy consumption where upper and lower bounds are adopted to determine an intelligent alarm system.

The second study entitled "Sugarcane Yield Prediction Through Data Mining and Crop Simulation Models" was written by Ralph G. Hammer, Paulo C. Sentelhas, and Jean C. Q. Mariano (41). The purpose of this study was to identify and coordinate the main variables that condition sugarcane yield, according to their relative importance, and to develop a mathematical model to predict sugarcane yield using data mining (DM) techniques. The algorithms used are Random Forest, Gradient Boosting Machine, and SVM. The results of this research are that the Random Forest algorithm as a regression approach has potential in making predictions, but FB Prophet is the technique with the most potential in estimating future electrical energy needs related to E333 and E260 which are HVAC sensors. The algorithm that has the lowest RMSE value is the FB Prophet with a value of 8,07 and the algorithm that has the lowest RMSE value is Random Forest with a value of 19,702 and the algorithm that has the fastest processing value is the Gradient Boosting Machine with a time of 0.034 hours.

The third study entitled "Prediction of Local Government Revenue using Data Mining Method" was written by Muhammad Zuhri Infusi, Gede Putra Kusuma, Dewi Annizah Arham (42). The goal of this research is to find the most accurate forecasting model to predict Regional Original Income in the coming year. The algorithms used are Linear Regression, Artificial Neural Network, dan Deep Learning. The results of this research are that Linear Regression is the best model for predicting Regional Original Income because it has the lowest RMSE value and the highest R^2 value. Linear Regression has an RMSE value of 97,318,950,671,576 and an R^2 value of 0.949. The performance of the Multiple Linear Regression method is better than the ANN and Deep Learning methods. The advantages of the Multiple Linear Regression method can be seen in the performance measurements and predicted values produced.

3. METHODOLOGY

The research framework outlines the logical flow of the research and looks at the relationships between variables to obtain results in accordance with the research objectives.

This research uses an experimental design, namely the author will use several variables in testing carried out systematically. The data used is shrimp seed sales data from 2017 to 2022. With a focus on forecasting shrimp seed sales at the company, this research will help companies in planning shrimp seed cultivation.

Seeing the current reality that companies are still planning cultivation manually and the results obtained by the company are not yet optimal, the application of appropriate machine learning models can help companies plan cultivation better. So that the company can obtain maximum profits, where the company achieves sales targets and can maximize cultivation.

The research framework can be described as follows:

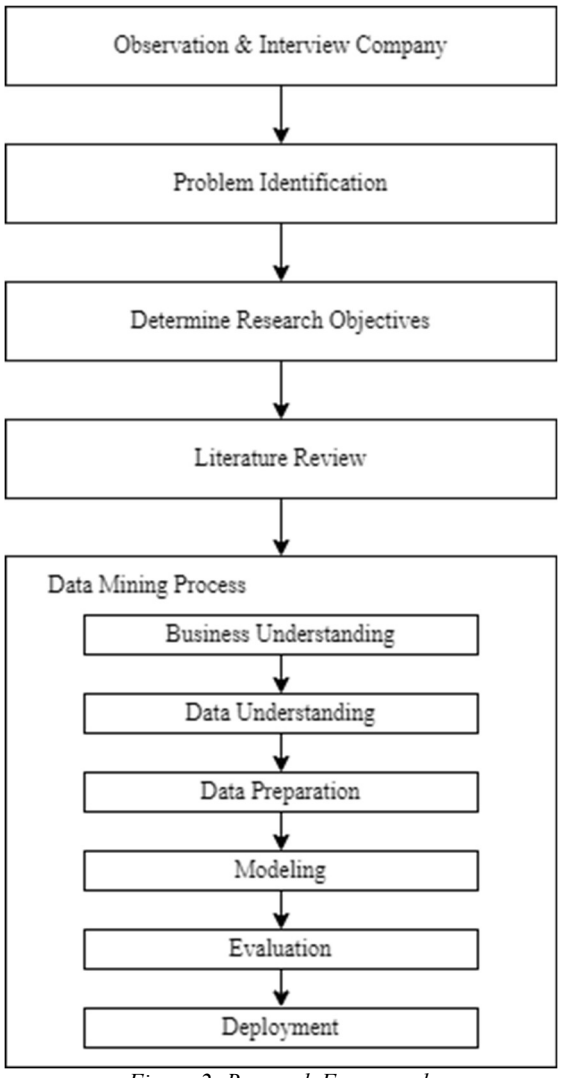

Figure 2: Research Framework

Based on the framework in Figure 2, the following are the steps the author took in conducting this research:

1. The first step the author took was to conduct a field study by direct observation at company hatchery to see the development process of shrimp seeds. Apart from that, the author also interviewed one of the production staff and one of the company's financial staff to obtain the information and data needed to conduct this research.

2. The author identifies the problems currently being experienced by the company based on the results of interviews conducted previously. The problem experienced by the company was that the company was unable to meet the sales target that had been set based on the estimated need for shrimp seeds in the following month.

3. After finding the problems experienced by the company, the author determined the research objectives based on the existing problems. The aim of this research is to find a machine learning algorithm that can help companies determine the right sales targets so that companies can maximize sales of shrimp seeds.

4. The author conducted a literature study to obtain information related to the topic raised in this research and to find out the results of previous research by other researchers with the same topic of discussion, namely forecasting.

5. Data Mining Process, this research uses the Cross-Industry Standard Process for Data Mining (CRISP-DM) method. The following is an explanation of the steps taken at each stage of CRISP-DM:

a. Business Understanding. In accordance with what has been explained in the background, the problem faced by the company is that the company cannot meet its sales targets. This sales target is set based on the number of requests for shrimp seeds in the previous month. This sales target will be used as a reference by the production department for cultivating broodstock shrimp for fertilization. The company has the risk of losing customers if the company cannot meet market demand and the company has the risk of throwing away shrimp seeds if the supply of shrimp seeds exceeds market demand. New ways to predict future sales are needed.

Based on the results of literature studies that have been carried out and discussed previously, it is believed that machine learning can be used as a good approach to apply in predicting shrimp seed sales. After knowing the problem and seeing the company's needs, the aim of this research is to obtain the best machine learning model that can be used to predict the number of shrimp seed sales in the coming month.

b. Data Understanding. A dataset is needed to make predictions, namely historical data on sales of shrimp seeds in previous months. The data taken is raw data. The dataset used is historical data on shrimp seed sales transactions from March 2017 to August 2022. This data was obtained directly from the company's financial staff.

In this research, an analysis of historical data on sales of PT Prima Akuakultur Lestari shrimp seeds will be carried out and experiments will be carried out by creating several machine learning models using the regression method. Data processing was carried out using Microsoft Excel and RapidMiner. The dataset has information from 828 sales transactions and has 7 attributes. Dataset attributes are presented in table 1.

Table 1: Attributes of PT Prima Akuakultur Lestari Shrimp Seed Sales Dataset

Attribute Name	Type	Description
Date	Categorical	Shrimp seed sale transaction date
Invoice No.	Categorical	Shrimp seed sales invoice number. This invoice number is unique, will not be repeated in other transactions
Customer	Categorical	The name of the buyer of shrimp seeds, can be an individual or a company
Price	Numerical	Unit price of shrimp seeds
Quantity	Numerical	Number of shrimp seeds sold
Quantity	Numerical	Number of shrimp seeds sold
Quantity Added	Numerical	Number of additional shrimp seeds in one invoice number
Total Bill	Numerical	Total bill for payment of shrimp seeds

Table 2: Examples of Attribute Values and the Number of Missing Value

Attribute Name	Value	Missing Value
Date	25/03/2017	0
Invoice No.	01/INV-PL/03/2017	1
Customer	Mr. Riyan	0
Price	53	0
Quantity	2.103.068	1
Quantity Added	138.525	812
Total Bill	105.153.400	8

The target variable for this research is "Quantity" which will be used to see how many sales will occur in the following month.

c. Data Preparation. The data that has been collected is identified first and checked for data quality. Data was cleaned to remove noise and inconsistent data. After author checking the data, it was found that several attributes had empty values, namely attribute No. Invoice, Quantity, Quantity Added, and Total Bill. However, in this research the author only needs the Quantity and Quantity Added attributes. The Quantity Added attribute is not an attribute that must have a value, so many values are empty. The Quantity Added attribute will be combined with the Quantity attribute based on No. Invoices.

In table 1 the Quantity attribute has 1 empty value. For this record the author will fill in the value based on the Total Bill attribute, because the Quantity attribute is an attribute that contains the number of shrimp seed sales. So, the Quantity attribute can be filled in based on the Total Bill attribute which contains the number of bills for selling shrimp seeds. The author will divide the Total Bill value by the Price attribute to get the Quantity value.

This research will use monthly sales data. Because the available dataset contains sales figures based on Invoice No., the author will create a new attribute, namely Sales Amount, which contains the number of shrimp seed sales per month which contains the Quantity and Quantity Added attributes based on month of sale. Examples of data transformation can be seen in table 3.

Table 3: Example of Data Transformation

Before			After	
Invoice No.	Quantity	Quantity Added	Month	Sales Amount
15/INV/ 05/2019	890.000	741.165	May 2019	2.799.943
16/INV/ 05/2019	360.000	305.577		
17/INV/ 05/2019	503.200			
01/INV/ 07/2019	100.520	875.000	July 2021	2.343.536
02/INV/ 07/2019	877.107			
03/INV/ 07/2019	327.273	163.636		

To improve the performance of the forecasting model, the Month attribute value data type will be changed to a numerical type. Based on the limitations of this research problem, the data used is sales data from March 2017 to August 2022. The total sales records for shrimp seeds used are 66 records. Statistical data from the data sources used in this research can be seen in Figure 3.

Figure 3: Graph of Shrimp Seed Sales of PT Prima Akuakultur Lestari

d. Modelling. The approach used in this research is regression. This research aims to predict the number of shrimp seed sales in the following month. The algorithms that will be used are K-Nearest Neighbors (KNN), Support Vector Regression (SVR), and Neural Network (NN). The tool used to perform modelling is RapidMiner version 9.10.

e. Evaluation. The dataset has sixty-six records and has one target variable. The target variable in this research is the Quantity variable which will predict the number of shrimp seed sales in the following month. Performance measurements for all models evaluated were assessed using Root Mean Square Error (RMSE) and R Squared (R^2). In total there were three models evaluated.

f. Deployment. The results obtained at this stage are in the form of sales data analysis reports using the best model for forecasting shrimp seed sales in the following month. Based on the test results of the three algorithms used, the K-Nearest Neighbors model has the highest accuracy value. With the highest accuracy value, companies can apply machine learning to predict shrimp seed sales using the K-Nearest Neighbors model.

4. RESULT AND DISCUSSION

4.1 Result

Table 4 shows a comparison of the performance results of the three models used in forecasting shrimp seed sales. Of the three models evaluated, K-Nearest Neighbors had the best performance. Performance measures for all evaluated models were assessed using Root Mean Square Error (RMSE) and R Squared (R^2). A total of three models were evaluated.

Table 4: Comparison of the performance results of the 3 models tested

No.	Algorithms	RMSE	R^2
1.	K-Nearest Neighbor	6326408.735	0.215
2.	Support Vector Regression	7008731.263	0.177
3.	Neural Network	6913028.1045	0.182

The RMSE value of each algorithm evaluated shows quite high numbers. Forecasting results will be more accurate if the RMSE value is close to 0 (zero). Conversely, the higher the R-Squared value, the better the model.

The author also tries to predict sales of shrimp seeds from September 2022 to February 2023 using the best model, namely K-Nearest Neighbors model.

Table 5: Results of Sales Forecasting for PT Prima Akuakultur Lestari for Six Months Using the K-Nearest Neighbors Model

No.	Month	Prediction
1.	September	8,283,444
2.	October	9,696,680
3.	November	9,855,207
4.	December	8,686,780
5.	January	7,926,580
6.	February	9,379,257

4.2 Discussion

In recent years, there has been increasing interest in using machine learning techniques for sales forecasting. Numerous researchers have explored various algorithms, such as regression models, time series analysis, and more advanced techniques such as neural networks and support vector machines. These methods have shown promising results in accurately forecasting sales trends and patterns. Forecasting sales will have a positive impact on optimization through determining the right amount of shrimp seed production/cultivation.

This research proposes a machine learning model that will be used to predict the number of shrimp seeds that should be cultivated to maximize company revenue. Just like the three previous studies, the researchers utilized data mining to extract data sets. The methodology used to carry out data processing is CRISP-DM.

The results found in this research are different from the three previous studies because the data and forecasting models used in the previous research are different from this research. In the research conducted by (40) the algorithms used were Random Forest, Decision Tree, Support Vector Machine, Linear Regression, and K-Nearest Neighbor where the best model was Random Forest. In the research conducted by (41) the algorithms used were Random Forest, Gradient Boosting Machine, and SVM with the best model being Random Forest but FB Prophet is the technique with the most potential in estimating future electrical energy needs. In subsequent research carried out by (42) the algorithms used were Linear Regression, Artificial Neural Network, and Deep Learning with the best model being Linear Regression.

Even though the goals and the data and models used are different, there are similarities between the other three studies and the current research, namely optimizing resource allocation, improving decision making, and making accurate estimates.

By forecasting shrimp seed sales, companies can plan future policies to optimize aquaculture production according to needs. In this research, the author faced limitations such as data quality (only having a few supporting variables). The accuracy of the model may change according to the use of new variables used. Future research should further examine what factors can influence the accuracy of the model used.

5. CONCLUSION

The dataset used is historical data on the sale of shrimp seeds by the company, totaling 828 data. Of the three Machine Learning models built, after testing it was found that the K-Nearest Neighbor model is the model with the best accuracy in predicting the number of seed sales, better than the Support Vector Regression and Neural Network models. The sampling method used is Split Validation sampling with 80% of the data used for training and 20% of the data used for testing. The RMSE and R^2 values of the K-Nearest Neighbor model are 6326408.735 and 0.215. Meanwhile, the Support Vector Regression model is 7008731.263 and 0.177. And the Neural Network model is 6913028.1045 and 0.182.

Forecasting of shrimp seed sales is carried out using the model with the best accuracy that has been tested, namely K-Nearest Neighbor, with the number of forecasts each month being 8,283,444 shrimp seeds in September 2022, 9,696,680 shrimp seeds in October 2022, 9,855,207 shrimp seeds shrimp seeds in November 2022, 8,686,780 shrimp seeds in December 2022, 7,926,580 shrimp seeds in January 2023, and 9,379,257 shrimp seeds in February 2023.

In this study there were only a few variables, so the author only used a few variables in carrying out the test. The author suggests that in future research the number of sample data and several other variables that are not in this study can be added. It is hoped that increasing the amount of sample data and variables can be a means of supporting the accuracy of Machine Learning models so that they become even better and can produce higher accuracy.

REFERENCES:

[1] DPR. Letak Geografis Indonesia. 2017 Jun.
[2] Fadhilah I. kompas.com. 2022. Berapa Luas Negara Indonesia?
[3] Pratama O. Direktorat Jenderal Pengelolaan Ruang Laut. 2020 [cited 2022 May 18]. Konservasi Perairan Sebagai Upaya menjaga Potensi Kelautan dan Perikanan Indonesia. Available from: https://kkp.go.id/djprl/artikel/21045-konservasi-perairan-sebagai-upaya-menjaga-potensi-kelautan-dan-perikanan-indonesia
[4] Setyawan EY. Badan Riset dan SDM Kelautan dan Perikanan. 2022 [cited 2022 May 18]. Produksi Budi Daya Udang di Indonesia. Available from: https://kkp.go.id/brsdm/sosek/artikel/39265-produksi-budi-daya-udang-di-indonesia
[5] BKIPM. Badan Karantina Ikan, Pengendalian Mutu Dan Keamanan Hasil Perikanan. 2018 [cited 2022 May 13]. Peta Lalulintas Benih Ikan dan Benur Udang Nasional 2018. Available from: https://kkp.go.id/bkipm/artikel/5880-peta-lalulintas-benih-ikan-dan-benur-udang-nasional-2018
[6] Ajie MD. Sistem Informasi. Academia; 1996.
[7] Sihotang HT. Sistem Informasi Pengagendaan Surat Berbasis Web Pada Pengadilan Tinggi Medan. Journal Of Informatic Pelita Nusantara. 2018;3(1).
[8] Hariyanto S. Sistem Informasi Manajemen. Publiciana. 2016;9(1):80–5.
[9] Sidh R. Peranan Brainware dalam Sistem Informasi Manajemen. Jurnal Computech & Bisnis (e-Journal). 2013;7(1):19–29.
[10] Rahman AA, Supaidi A, Aslamiah I, Ibrahim A. Implementasi Customer Relationship Management (CRM) Pelayanan Pelanggan (Corporate) Divisi BGES Pada PT Telkom Witel Sumsel. JRMSI-Jurnal Riset Manajemen Sains Indonesia. 2018;9(1):72–8.
[11] Kurniawan D. Penerapan Aplikasi CRM (Customer Relationship Management) Berbasis Web Dalam Bidang Jasa. Universitas Bina Nusantara Jakarta. 2009;
[12] Dyantina O, Afrina M, Ibrahim A. Penerapan Customer Relationship Management (CRM) Berbasis Web (Studi Kasus Pada Sistem Informasi Pemasaran di Toko YEN-YEN). JSI: Jurnal Sistem Informasi (E-Journal). 2012;4(2).
[13] Syabania R, Rosmawarni N. Perancangan Aplikasi Customer Relationship Management (CRM) Pada Penjualan Barang Pre-Order Berbasis Website. Jurnal Rekayasa Informasi. 2021;10(1):44–9.
[14] Wildyaksanjani JP, Sugiana D. Strategi Customer Relationship Management (CRM) PT Angkasa Pura II (Persero). Jurnal Kajian Komunikasi. 2018;6(1):10–23.
[15] Syam N, Sharma A. Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. Industrial marketing management. 2018;69:135–46.
[16] IBM. IBM. What is machine learning?
[17] Fardani DP, Wuryanto E, Werdiningsih I. Sistem Pendukung Keputusan Peramalan Jumlah Kunjungan Pasien Menggunakan Metode Extreme Learning Machine (Studi Kasus : Poli Gigi Rsu Dr. Wahidin Sudiro Husodo Mojokerto). Journal of Information Systems Engineering and Business Intelligence. 2015 Jun 25;1(1):33.
[18] Bontempi G, Ben Taieb S, Le Borgne YA. Machine learning strategies for time series forecasting. Business Intelligence: Second European Summer School, eBISS 2012, Brussels, Belgium, July 15-21, 2012, Tutorial Lectures 2. 2013;62–77.
[19] Budiyantara A, Wijaya AK, Gunawan A, Rolland M. Analisis Data Mining Hotel Booking Menggunakan Model ID3. JBASE-Journal of Business and Audit Information Systems. 2021;4(1).
[20] Gupta GK. Introducing to Data Mining with Case Studies. Third Edition. Delhi: PHI Learning Private Limited; 2014. 3 p.
[21] Damuri A, Riyanto U, Rusdianto H, Aminudin M. Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako. JURIKOM (Jurnal Riset Komputer). 2021;8(6):219–25.
[22] Setiawan R. dicoding. 2021 [cited 2023 Jan 28]. Apa itu Data Mining dan Bagaimana Metodenya? Available from: https://www.dicoding.com/blog/apa-itu-data-mining/

[23] Chen MS, Han J, Yu PS. Data Mining: An Overview from a Database Perspective. IEEE Trans Knowl Data Eng. 1996;8(6):866–83.
[24] Maulid R. DQLab. 2022 [cited 2023 Feb 7]. Teknik Analisis Data CRISP-DM dalam Data Mining. Available from: https://www.dqlab.id/teknik-analisis-data-crisp-dm-dalam-data-mining
[25] Mauritsius T, Binsar F. BINUS University. 2020 [cited 2023 Feb 7]. Cross-Industry Standard Process for Data Mining (CRISP-DM). Available from: https://mmsi.binus.ac.id/2020/09/18/cross-industry-standard-process-for-data-mining-crisp-dm/
[26] Martínez-Plumed F, Contreras-Ochando L, Ferri C, Hernández-Orallo J, Kull M, Lachiche N, et al. CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. IEEE Trans Knowl Data Eng. 2019;33(8):3048–61.
[27] Yudha MA. Medium. 2021 [cited 2023 Feb 7]. CRISP-DM, Pendekatan Proses dalam Data Mining. Available from: https://andiyudha.medium.com/crisp-dm-pendekatan-proses-dalam-data-mining-68bf8c2dc908
[28] Hasanah MA, Soim S, Handayani AS. Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. Journal of Applied Informatics and Computing. 2021;5(2):103–8.
[29] Schäfer F, Zeiselmair C, Becker J, Otten H. Synthesizing CRISP-DM and Quality Management: A Data Mining Approach for Production Processes. In: 2018 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD). IEEE; 2018. p. 190–5.
[30] Saltz J. Data Science Process Alliance. 2022 [cited 2023 Feb 7]. CRISP-DM is Still the Most Popular Framework for Executing Data Science Projects. Available from: https://www.datascience-pm.com/crisp-dm-still-most-popular/
[31] Moro S, Laureano R, Cortez P. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. 2011;
[32] IBM. IBM. 2021 [cited 2023 Feb 7]. CRISP-DM Help Overview. Available from: https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview
[33] Purnama I, Saputra R, Wibowo A. Implementasi Data Mining menggunakan CRISP-DM pada Sistem Informasi Eksekutif Dinas Kelautan dan Perikanan Provinsi Jawa Tengah. In: Prosiding Seminar Nasional Ilmu Komputer UNDIP. 2012.
[34] Sharda R, Delen D, Turban E. Business Intelligence, Analytics, and Data Science: A Managerial Perspective. Fourth Edition. Harlow: Pearson Education; 2018.
[35] Mahesh B. Machine Learning Algorithms: A Review. International Journal of Science and Research (IJSR)[Internet]. 2020;9:381–6.
[36] Langley P, Simon HA. Applications of Machine Learning and Rule Induction. Commun ACM. 1995;38(11):54–64.
[37] Al-Sahaf H, Bi Y, Chen Q, Lensen A, Mei Y, Sun Y, et al. A Survey on Evolutionary Machine Learning. J R Soc N Z. 2019;49(2):205–28.
[38] El Naqa I, Murphy MJ. What is Machine Learning? Springer; 2015.
[39] Jordan MI, Mitchell TM. Machine Learning: Trends, Perspectives, and Prospects. Science (1979). 2015;349(6245):255–60.
[40] Mosavi NS, Freitas F, Pires R, Rodrigues C, Silva I, Santos M, et al. Intelligent energy management using data mining techniques at Bosch Car Multimedia Portugal facilities. Procedia Comput Sci. 2022;201:503–10.
[41] Hammer RG, Sentelhas PC, Mariano JCQ. Sugarcane yield prediction through data mining and crop simulation models. Sugar Tech. 2020;22(2):216–25.
[42] Infusi MZ, Kusuma GP, Arham DA. Prediction of Local Government Revenue using Data Mining Method. International Journal of Emerging Technology and Advanced Engineering. 2022 Jan 16;12(1):63–74.