K-nearest neighbor analysis to predict the accuracy of product delivery using administration of raw material model in the cosmetic industry (PT Cedefindo)

H B Novitasari¹, N Hadianto¹, Sfenrianto², A Rahmawati¹, R Prasetyo¹, J Miharja¹, W Gata¹

¹STMIK Nusa Mandiri, Jakarta, Indonesia
²Information Systems Management Department, Bina Nusantara University, Jakarta, Indonesia

E-mail: hafifah.bella@gmail.com

Abstract. The purpose of this study is to apply the K-Nearest Neighbor (KNN) algorithm to predict the accuracy of product delivery using the raw material administration model, which has 3 attributes, namely administration of material purchases, administration of material readiness and processing work orders compared to existing targets. From this study, it was found that the KNN Algorithm was effective in predicting the accuracy of product delivery using the administration of raw material model applied to the production history data from July 2018 to December 2018 with a small error ratio. Thus, data mining with the KNN algorithm can be used in decision making within the company to predict the accuracy of product delivery.

1. Introduction

Production planning is an important process in maximizing company performance and aims to serve different customers. Accuracy in planning the production of a product may require smart-resources [1]. Production companies have several production module planning that utilizes the allocation of employee activity resources, materials, and production capacity [2]. There are three important components in production planning, namely: process, scheduling, and material resource planning (MRP) [3].

With production planning, it is expected that a company can deliver products according to the time determined by the customer. But often planning does not match the reality in operation so that it occurs not following the lead time. Because of this, the company needs data analysis or data mining to predict in the future.

In this study continued the previous research entitled Analysis of Regression Algorithm to Predict Administration, Production, and Delivery to the Accuracy of Delivery of Products in the Cosmetic Industry. From that study, it can be seen that the attribute that most influences the accuracy of product delivery is the Administration which has a coefficient of 0.925, greater than the Production and Delivery.

This study aims to apply the K-Nearest Neighbor algorithm to predict the accuracy of product delivery using the Administration of Raw Material model.

Administration time is meant is an Office Administration, where humans are the most important factor in achieving satisfactory and efficiency office services [4]. The administration is that part of the activity of a body or group of men which is concerned with the management of the affairs of that body or group [5]. Based on PT Cedefindo's production history data, it can be seen that the Administration
K-Nearest Neighbor (KNN) is an optimization approach in various fields such as production optimization, pattern recognition, image processing, etc. The KNN approach is suitable for algorithms that have large training data [6]. The KNN algorithm has accurate optimization results and aims to classify new objects based on attributes and training samples [7]. Optimization using KNN is more suitable to build a classification method without making assumptions about the form of the "smooth" function $f$ [8]:

$$x = (x_1, x_2, ..., x_n)$$

$$y = f(x)$$

that relates to $y$ (the response variable) to $x$ (predictor variables). The function $f$ is non-parametric because it does not involve any estimation of parameters in any shape or form. In KNN, given a new point $p = (p_1, p_2, ..., p_n)$, we dynamically identify $k$ observations in the training data set that are similar to $p$ (the $k$ nearest neighbors) [8]. The Euclidean distance between the points $(x_1, x_2, ..., x_n)$ and $(p_1, p_2, ..., p_n)$ is defined as [8]:

$$\sqrt{(x_1 - p_1)^2 + (x_2 - p_2)^2 + \cdots + (x_n - p_n)^2}$$

For each queried n-dimensional object (i.e., an object that has n attributes), the Euclidean distances between the queried object and all the training data objects are calculated, and the queried object is assigned the class label the most of the $k$ closest training data has [8].

Several previous studies used the k-nearest neighbor algorithm. The first study, Forecasting for prediction of disturbances using past events based on neighbor k-Nearest. Data used from industrial process factories. This study also explains the prediction of measurable disturbances in an unsupervised manner, on changes in the characteristics of the disturbances that occur [9].

Another study, analysis of the nearest neighbors to predict cesarean section. This study proposes Cephalopelvic disproportion (CPD) and the closest neighbor analysis (NN) called the CPD-NN algorithm. There are three phases of the CPD-NN algorithm, namely: the initial phase, the distance measurement phase, and the prediction phase. The results showed that efficiency and accuracy occurred using the prediction of CPD-NN [10].

The KNN approach is also used for stock price predictions using. The sample data used are six large companies on the Jordanian stock exchange. The results of this study can be used to help provide correct investment information for investors [11].

The following study of data mining systems and recommendations for automatic web use using KNN. Data used is based on user behavior based on user click activity on the Really Simple Syndication (RSS) site. The results show that KNN classifiers are easy to understand and have good accuracy than most other machine learning techniques [12]. Also, the KNN approach for short-term wind speed predictions. This study explains the K-NN classification model implemented by using object-oriented programming techniques [13].

2. Results And Discussion
This study uses Knowledge Discovery Database (KDD) method, there are 5 steps that must be done in this method, namely: Data Selection, Pre Processing, Transformation, Data Mining, and Interpretation Evaluation (see Figure 1).
1. **Data Selection**
   In this step, data were selected from a series of operational data. This paper uses the production history data of PT Cedefindo for the delivery date in 2018, starting from July 2018 to December 2018 with a total history of 644 data.

2. **Data Pre Processing**
   At this stage, data cleaning is done including removing data duplication, checking data inconsistencies, and data errors. The production history data selected was 644 data, then after preprocessing it was 610 data.

3. **Data Transformation**
   After doing data selection and pre-processing data, the next step is to do data transformation. Data that has been through pre-processing as many as 610 data consists of 8 attributes, namely Date Order, Order Number, Product Code, Delivery Target, Raw Material (RM) Date, Material Readiness Date, Processing Work Order (PWO) Date and Send Date are transformed into 4 attributes, namely Administration of Material Purchases (AMP), Administration of Material Readiness (AMR), PWO and Target.

4. **Data Mining**
   This research uses data mining algorithms to process data for analysis, namely the K-Nearest Neighbor algorithm. The transformed data then processed using the K-Nearest Neighbor algorithm.

5. **Interpretation or Evaluation**
   Data generated from algorithms become information that is expected to help improve the performance of production planning in achieving product delivery targets.

In this study use software tools to process data for analysis with the K-Nearest Neighbor algorithm. The software tools used are RapidMiner version 9.0, in order to be able to use software tools, we need to prepare production history data that has gone through KDD stages, namely the pre-processing and data transformation stages. The following is the example of production history data in table 1.
Based on KDD method, the first step which is data selection as explained in table 1. Data is selected from product history data delivery dates from July 1, 2018, to December 31, 2018, as many as 644 data. From these data pre-processing is done so that it becomes 610 data, then transformation data is carried out which produces 4 attributes namely: (1) AMP = Total days from Order Date to RM Date; (2) AMR = Total days from RM Date to Material Readiness Date; (3) PWO = Total days from Material Readiness Date to PWO Date; and (4) Target = Total days from Order Date to Delivery Target. The following is data that has been transformed and then processed in Rapidminer in table 2.

| No | Order Date | Order Number | Product Code | Delivery Target | RM Date | Material Readiness Date | PWO Date | Send Date |
|----|------------|--------------|--------------|----------------|---------|--------------------------|----------|-----------|
| 1  | 2017-05-30 | OPE-17-05-00120 | 6773AMPLEN10 | 2018-03-30 | 2017-11-28 | 2017-11-28 | 2018-06-11 | 2018-07-17 |
| 2  | 2017-06-22 | OPE-17-06-00133 | 5803ANLMPB10 | 2018-09-07 | 2018-08-07 | 2018-09-11 | 2018-08-31 |
| 3  | 2017-07-10 | OPE-17-07-00028 | 6433SDCCR10  | 2018-08-30 | 2018-06-22 | 2018-07-18 | 2018-08-28 |
| 4  | 2017-07-10 | OPE-17-07-00029 | 6433SDCCR10  | 2018-11-30 | 2018-06-22 | 2018-08-02 | 2018-09-28 |
| 5  | 2017-07-10 | OPE-17-07-00030 | 6433SDCCR10  | 2018-10-31 | 2018-10-01 | 2018-10-03 | 2018-10-29 |
| 6  | 2017-06-29 | OPE-17-06-00127 | 5803ANLMPB10 | 2018-08-10 | 2018-09-07 | 2018-09-23 | 2018-09-25 |
| 7  | 2017-08-16 | OPE-17-08-00146 | 6313SJFSPM10 | 2018-08-13 | 2017-11-30 | 2018-08-03 | 2018-08-09 |
| 8  | 2017-10-10 | OPE-17-10-00079 | 6103C2CPC10  | 2018-06-04 | 2018-06-22 | 2018-07-04 | 2018-07-17 |
| 9  | 2017-10-10 | OPE-17-10-00080 | 6103C2CPC10  | 2018-06-10 | 2018-06-22 | 2018-07-13 | 2018-07-23 |
| 10 | 2017-10-20 | OPE-17-10-00094 | 5803ANNVPB10 | 2018-07-06 | 2017-11-30 | 2018-07-12 | 2018-07-18 |

**Table 2. Examples of transformation data.**

| AMP | AMR | PWO | TARGET |
|-----|-----|-----|--------|
| 182 | 0   | 195 | 304    |
| 411 | 0   | 42  | 442    |
| 347 | 0   | 41  | 508    |
| 448 | 0   | 2   | 478    |
| 404 | 0   | 16  | 407    |
| 106 | 0   | 246 | 362    |
| 245 | 0   | 12  | 227    |
| 245 | 0   | 21  | 227    |
| 41  | 0   | 224 | 259    |
| 242 | 0   | 10  | 281    |
| 242 | 0   | 10  | 281    |
| 237 | 0   | 20  | 266    |
| 209 | 0   | 0   | 234    |

In study use 2 as the maximum number of K. By experimenting with different K values on the training dataset, starting from K = 1 to K = 9. From the experiment, it was found that predictions can work better with K = 2 with a minimum error rate. Then, table 3 presents the results of predictions with the K-Nearest Neighbor using RapidMiner. This research uses Cross Validation operator on Rapid
Miner, where Cross Validation has two subprocesses, namely subprocess training, and subprocess testing. The dataset is automatically divided into multiple sets based on 10 number of folds.

Table 3. The Results after applying KNN using RapidMiner.

| TARGET | PREDICTION | AMP | AMR | PWO |
|--------|------------|-----|-----|-----|
| 246    | 199        | 194 | 0   | 9   |
| 253    | 224        | 211 | 0   | 25  |
| 218    | 155        | 168 | 0   | 0   |
| 203    | 73         | 153 | 0   | 0   |
| 78     | 92         | 74  | 0   | 56  |
| 78     | 78         | 74  | 0   | 31  |
| 92     | 121        | 72  | 0   | 91  |
| 86     | 114        | 105 | 0   | 0   |
| 119    | 119        | 139 | 0   | 30  |
| 93     | 78         | 86  | 0   | 0   |

Delivery prediction after applying the KNN Algorithm to 611 data from the period July 2018 to December 2018 is the only 10 data are shown in table 3. RapidMiner shows the total RMS error is 0.847 and Square error is 0.718. From the results of these predictions it can be seen that between the target and the prediction have residues, the following is the residue of the prediction results, in table 4.

Table 4. Residual between Target and Prediction

| TARGET | PREDICTION | RESIDUAL | AMP | AMR | PWO |
|--------|------------|----------|-----|-----|-----|
| 246    | 199        | 47       | 194 | 0   | 9   |
| 253    | 224        | 29       | 211 | 0   | 25  |
| 218    | 155        | 63       | 168 | 0   | 0   |
| 203    | 73         | 130      | 153 | 0   | 0   |
| 78     | 92         | -14      | 74  | 0   | 56  |
| 78     | 78         | 0        | 74  | 0   | 31  |
| 92     | 121        | -29      | 72  | 0   | 91  |
| 86     | 114        | -28      | 105 | 0   | 0   |
| 119    | 119        | 0        | 139 | 0   | 30  |
| 93     | 78         | 15       | 86  | 0   | 0   |

Based on the residual value it can be seen how far the difference between the predicted value and the actual value of the target, the negative residual value indicates that the predicted value is greater than the actual value of the target. From all data in this study, it can be calculated that the total Target value is 49585 and total Prediction value is 48230, there is a difference between the Target value and the Prediction value which is 1355.

The difference between the Target value and the Prediction value also shows that the value of error is small and the prediction results using KNN Algorithm have an accuracy of 97.27%. From the previous study that entitled Analysis of Regression Algorithm to Predict Administration, Production, and Delivery to the Accuracy of Delivery of Products in Cosmetic Industry have an accuracy of 91.67% which is this research also focuses on product delivery but use Linear Regression algorithm. And also we have tried using Naive Bayes to predict the accuracy of product delivery using the administration of raw material model, but the results of RMS errors show a number of 0.944 which is greater than using the KNN algorithm. So, this study shows that the use of KNN Algorithm to predict the accuracy of product delivery using the administration of raw material model has high accuracy.
3. Conclusion
According to the results, the KNN Algorithm is effective in predicting the accuracy of product delivery using the administration of raw material model applied to the production history data from July 2018 to December 2018. This paper also shows that the KNN Algorithm is stable with a small error ratio, so the results are rationally approaching the actual value of the target. So, data mining with the KNN algorithm can be used in decision making for the company to predict the accuracy of product delivery using the administration of raw material model.

References
[1] Richard A, et al., 1996, United States Patent (19) no. 19.
[2] Fargher H E and Smith R A, 1992, Method and System for Production Planning 19.
[3] Zhou, Zude, ed., 2010, Manufacturing Intelligence for Industrial Engineering: Methods for System Self-Organization, Learning, and Adaptation.
[4] Eyre E C, 2015, Management in Office Administration.
[5] Pruthi R K, 2005, Theory of Public Administration.
[6] Su C, et al. 2008, Some Progress of Supervised Learning (Advanced Intelligent Computing Theories and Applications With Aspects of Artificial Intelligent) pp 661-666.
[7] Laaksonen, Jorma, and Erkki Oja, 1996. "Classification with learning k-nearest neighbors." In Proceedings of International Conference on Neural Networks (ICNN’96), vol. 3, pp. 1480-1483. IEEE.
[8] Parsian M., 2015, Data Algorithms.
[9] Borghesan, Francesco, Moncef Chioua, and Nina F., 2019, Thornhill. Forecasting of process disturbances using k-nearest neighbours, with an application in process control (Computer and Chemical Engineering vol 128) pp 188-200.
[10] Sodsee S., 2014, Predicting caesarean section by applying nearest neighbor analysis (Procedia Computer Science vol 31) pp 5-14.
[11] Alkhatib, Khalid, Hassan Najadat, Ismail Hmeidi, and Mohammed K. Ali Shatnawi, 2013., Stock price prediction using using K-Nearest Neighbor (KNN) algorithm (International Journal of Business, Humanities and Technology vol 3) pp 32-44.
[12] Adeniyi, David Adedayo, Zhaoqiang Wei, and Y. Yongquan., 2016, Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method (Applied Computing and Informatics vol 12) pp 90-108.
[13] Yesilbudak, Mehmet, Serer Sagiroglu, and Ilhami Colak., 2013, A new approach to very short term wind speed prediction using k-nearest neighbor classification (Energy Conversion and Management vol 69) pp 77-86.