

## Research Article

# DEVELOPMENT A FORECAST MODEL FOR PREDICTING SCHOOL UNIFORM ORDERS USING DATA MINING TECHNIQUES

\* Phatarapon Vorapracha

Department Division of Information and Digital Technology Management, Faculty of Industrial Technology, Phranakhon Rajabhat University, Bangkok, 10220, Thailand.

Received 27<sup>th</sup> November 2022; Accepted 28<sup>th</sup> December 2022; Published online 30<sup>th</sup> January 2023

### ABSTRACT

The development of a forecast model for predicting school uniform orders using data mining techniques aimed 1) to create and test a school uniform order forecast model. 2) to find the efficiency of the system and the satisfaction with the forecasting system and can predict student uniform orders using data mining techniques. The researcher used a method to categorize the results of the student uniform order data classification. The model developed with Cross Validation 10 Folds method using the Decision Tree (J48) algorithm, Naive Bayes algorithm, and the Neighbour search algorithm found that the Decision Tree (J48) algorithm had an accuracy of 86.60%, an accuracy of 87.10%, a recall of 84.50%, and a balance of 86.40%. The Naive Bayes athlete has an accuracy of 79.77%, an accuracy of 78.70%, a recall of 78.70% and a balance of 75.66%, the Near Neighbour Algorithm has an accuracy of 82.65%, an accuracy of 81.60%, a recall of 80.44%, and a balance of 81.88% The results showed that the classification of student uniform order data using the Decision Tree (J48) algorithm, the data classification was 84.84% accurate and there were errors. 15.16 percent, Naive Bayes algorithm, 74.11% data classification accuracy and 25.89 percent error, and Proximity algorithm. The data classification results were 81.60 percent accurate and there was an error of 18.40 percent, respectively. It can be seen that the Decision Tree (J48) algorithm has the highest student-order data classification value than any other technique students effectively.

**Keywords:** School Uniform, Data Mining, Model.

### INTRODUCTION

Currently, the student's school attendance requirement is to wear a school uniform as required by that school. The law is the School Uniforms Act B.E. 2551, Section 5 requires students to wear school uniforms. by the characteristics of the school uniforms, the method of dressing, the conditions for dressing and exemption from having to wear a school uniform to be in accordance with the regulations prescribed by the Ministry of Education Any student who does not wear a school uniform without exemption under paragraph two may be subject to disciplinary action in accordance with the regulations set by the Ministry of Education. Many parents have a need to buy school uniforms for their children at least, but parents have to buy 2-5 sets of school uniforms for each student on average. In each school, there are different types of school uniforms for each class period. which has different styles of school uniforms from such needs It was found that the school uniform store used a forecasting method for the stock of each type of school uniform. and each year in advance which from the inquiry found that have a residual rate and selling school uniforms out of stock varies. Some years have a lot of leftovers, which leads to outstanding costs. In addition, the remaining school uniforms have deteriorated quality when they are sold the following year. And some years, the school uniforms prepared are not enough to sell. Thus causing a lot of loss of income opportunities. As a result of such problems, the store was unable to predict the number of school uniforms in each category. Including the size of the school uniform from the manufacturer to serve with the number of customers who will order It was found that if parents go to buy school uniforms during the school's opening period May not be able to find a school uniform that is suitable for the shape of the student. From the problems that arise, the researcher sees the importance of therefore, a forecasting model

was developed to predict the order of school uniforms. The researchers used data mining techniques to help solve the problem. As a system to help ordering school uniforms into the store's stock and help support the decision to buy school uniforms to sell in the store to meet the needs of parents and do not allow a large number of outstanding products in stock

### LITERATURE REVIEW

The researchers conducted development of a forecast model for predicting school uniform orders using data mining techniques the following related research studies. [1] The study is to implement data mining techniques for reducing the resignation rate of employees as well as decreasing expenses caused by the recruitment processes in an insurance company. The analysed data set is the data of resigned employees and employees who are still working from the year 2013-2017. The data of 1,000 items with 11 attributes is analysed using association rule mining to find the patterns and relations of employee resignation. Moreover, five predictive models for the resignation of employees using decision tree, support vector machine, neural network, naive Bayesian and nearest neighbour techniques have been conducted. The efficiency comparison among those five models are performed based on 5-fold cross-validation technique. It is found that the factors that are related to resign decision are salary, working age, career advancement, performance evaluation and relationship with supervisors. The results of the model evaluation show that the decision tree predictive model yields the highest accuracy at 91.03%, while support vector machine technique's predictive model fields accuracy of 90.93%. Neural Network technique predictive model fields accuracy at 90.75%. NaiveBayes and K-Nearest Neighbours models yield 89.60% and 82.10% of accuracy, respectively. These results can be applied to compensation strategies for employee retention. Also, it can be a guideline for human resource information systems in the future.

\*Corresponding Author: Phatarapon Vorapracha,

Department Division of Information and Digital Technology Management, Faculty of Industrial Technology, Phranakhon Rajabhat University, Bangkok, 10220, Thailand.

[2] The objective of this research were 1) to development of model to predict the yield of cassava using data mining techniques, 2) to develop a forecast yield of cassava. With queries system to evaluate satisfaction of users of the system through the website. The process on data from the Kamphaeng Provincial Agriculture Extension Office. A predictive model was based on techniques Classification with a decision tree, using an algorithm of five, including the J48, Random Tree, Simple Cart, NaïveBayes, and LAD Tree. Then tested the predictive models provide a technical Cross-validation The result found that the accuracy was only 70.96%. Which was not good enough. Therefore, testing methods had improved by splitting the series into two parts Including training set and 5 test set and then create predictive models. Using the following: Algorithm J48, Simple Cart, LAD Tree, Random Tree and Naïve Bayes again. The model obtained from Training set method were more accurate than from Cross-validation Test ins in all algorithms used, which were J48, Random Tree, Simple Cart, NaïveBayes, and LAD Tree Thus the selection of predictive models that provided the highest accuracy. In terms of Test Set, Algorithm J48 for maximum accuracy at 75.64%, Simple Cart with 80.12%, and LAD Tree with 89.55% while algorithms Random Tree and Naïve Bayes the accuracy lower than 70% were not taken into consideration. Then only the first three models were used to implement the development of model to predict the yield of cassava. The model was integrated with the web-based information system for information search on cassava, which was designed and developed in this research to provide basic information such as update, add, and reports. The decision support respondents were rated their satisfaction over the site of such information. Officials of kamphaeng provincial agriculture extension, users, and administrator a total of 30 people. There was satisfaction in the use of such systems by averaging the responses to a satisfaction rating of over 91%, which was considered very good. [3] Supply chain management is a collection of activities of flow and processing of goods from raw materials through various processes until reaching the final consumer. The same is true for the flow of information that is associated with increasing supply chain relationships. In order to maintain success for a competitive advantage, it starts with setting the same objectives and continually exchanging information in order to reduce quantitative uncertainty for management performance requirements and decision making process.[4] Web application development for water resources, this web application provides 3 web services: 1) management, presentation and geographic storage 2) support for modelling water resources 3) use of water resources. appropriate This web application is developed with programming languages such as PHP Ajax, JavaScript Java based libraries such as Open Layers jQuery and Open Software such as Geo Server PostgreSQL Post GIS. This web application can be utilized at any time and can be accessed everywhere able to create instant collaboration patterns with multiple users Flexibility for additional components and services.

**MATERIALS AND METHODS**

This paper is a data mining technique used in analysis. By comparing the three techniques as follows. [5] [6] Classification is data mining to classify the data to be analysed as being in the category. Based on the model or model that was created classification will use a certain amount of existing data or datasets to model. After the model has been designed, it must be tested with another group of data. To check the accuracy or the reliability of that model data mining by categorization used for discrete values such as. The data provided is the result of deciding yes or no, risky or not risky. Classified data mining tools include algorithms for generating decision trees, Bayesian theory, NaïfBays, neural networks. Bayesian belief, rule-based analysis in IF\_THEN rule model, neural network or

neural network SVM correlation law, finding similar properties or similar by k-nearest neighbour method, reasoning and effect to answer hypothesis by case base reasoning method and genetic algorithm. Predictive data mining tools include regression analysis, chi-square analysis, and validation. The function of classification starts from importing data by using classification techniques to create a model by applying the model to identify the type of data that is the target value as shown in Figure 1.



**Fig.1 Basic functions of classification.**

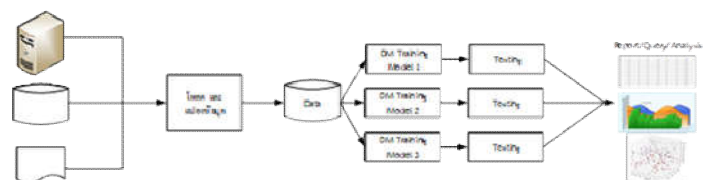
[7][8] Techniques used to create models or models of classifications. The decision tree is a technique that relies on a tree-like structure. It consists of a branched root node representing data attributes and leaves representing an assortment of data. [9] Naive Bayes algorithm, a data classification model that uses probability principles. Based on the Bayes' Theorem and the assumption that the occurrence of events is independent, if P(h) is given, the probability of occurrence of event h and P(h|D) is the probability that An event h occurs when an event D occurs from a given variable. We can detect events considering events as follows:

$$P(h | D) = P(D | h) * P(h) / P(D)$$

from the above equation The probabilities of Bayesian classification can be calculated as follows:

$$P(d | h) = P(a_1, \dots, a_n | h) = P(a_1 | h)$$

[10] K-Nearest Neighbors Classification refers to the method used to classify classes. By this technique, it is decided Which class will represent a new condition or case by checking a certain number of "K" in a case's nearest neighbor algorithm or the same conditions or as close as possible. It will find the sum (Count Up) of the number of conditions or cases for each class and define new conditions for classes that are the same as those that are closest to each other. The technique of nearest neighbor algorithm is used to find the distance between each variable (Attribute) in the data and then calculate the value. This method is suitable for numerical data with discrete variables. It can be done, but requires more specialized management. [11] Developing a Forecast Model for Predicting School Uniform Orders Using Data Mining Techniques. [12] It is a system that will facilitate the ordering of school uniforms into the stock as shown in Figure 2.



**Fig. 2 Data import prototyping concept.**

Prototyping by taking school uniform sales data from the store and various factors, both internal and external factors affecting school uniform sales for that year, load and convert the data into a format that can be used. The analyzes were stored in the database, and three prototypes were tested using the training data to learn the prototypes. When learning the prototype, then use the learned data to make a testing set from the actual data that has been shared from the learning data. The results obtained from the comparative analysis test

selected the prototype suitable for the student uniform order forecasting system.

### RESULTS AND DISCUSSION

The results of studying various methods to be applied in research. which can describe the research results into 4 parts as follows:

- a) Study student dress in this research, the researcher studied the student's dress style and was able to choose the student's dress style in 45 different ways. Study the process of data mining. The researcher has established a standardized process for implementation which uses a standard known as CRISP-DM (CRISP DM – Cross Industry Standard Process for Data Mining) to help work more efficiently consists of 6 steps: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment. Classification studies by researchers study techniques used to forecast student uniform orders. Decision Tree Technique, Naïve Bayes Technique and K-Nearest Neighbors Techniques. The researcher collected data from 3 branches of Sereephan stores using the data of the school uniform purchases each year and the school uniform sales data from 2018 - 2021, a total of 587 product items each year.
- b) The researcher studied the data. There are 3 branches of Sereephan stores using data on ordering and selling school uniforms from 2018 - 2021, totaling 587 items, which consists of the following information: in Figure 3.
  - Product type
  - Size
  - Number of school uniforms ordered
  - Number of school uniforms sold
  - Number of inventories

Fig. 3 Order information example and sales of school uniforms, 2018 – 2021.

Data cleaning and transformation. The researcher collected data from 3 branches of Sereephan stores using the order information. and sales of school uniforms. From 2018 - 2021, the researcher selected only complete data, meaning information on product types, sizes, number of school uniforms ordered and the number of school uniforms sold. The researcher chose the data from 2018 - 2020 because it was found that in 2021 the sales of school uniforms have dropped considerably. Due to the lack of a semester for students due to the 2019 COVID-19 situation, it has been announced for students to study online (students do not have to go to school), causing data in the year 2021 to have sales declines as shown in Table 1.

Table 1 Show the number of product sales for the year 2018 - 2021

Year	2018	2019	2020	2021
Sales amount	5,046	5,938	6,231	1,725

From Table 1, sales of products sold in 2018 were 5,046 items, in 2019 there were 5,938 items, and in 2020, there were sales at 6231 items and the year 2021 had sales of 1,725 items. It was found that the items of product sales in the year 2021 had decreased significantly. The researcher therefore chose to use the data to create a prototype only in the years 2018-2020. The researcher collected the data in the year 2018-2020 with the total number of product lists of 45 items as shown in Table 2

Table 2 Product List

No.	Product	Number of Entries
1	Free male student shirt	13
2	Free naval officer's shirt	10
3	Student Polo Shirt Free	7
4	Student shirt lotus neck cut all the time free	9
5	Girls shirt high school free	8
6	Toray Free Scout Shirt	11
7	Seri Girlfriend Shirt	8
8	Free Red Cross youth shirt	9
9	Kindergarten Shirt Bua Seree	5
....	....	.....
45	Student Skirt 6 Pleated Durian Blue Look	26
Summation		587

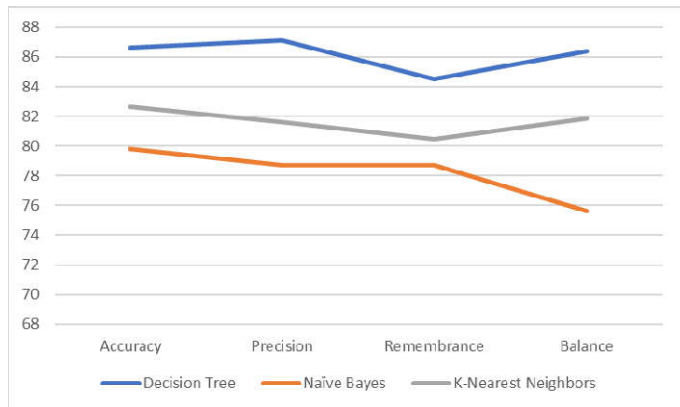
From Table 2, it is found that there are 45 product items and total number of items in each category is 587 items. Then the researcher selects complete data, meaning that there is a complete list of products for sale and purchase into the Seripan store. All 3 years, a total of 587 items, then the information obtained And convert the data by replacing all 45 item types with X01, X02, X03, ....., X45 respectively. Convert all data item size data to a CSV file format for preparation. The model was tested using WEKA program. An example is shown in Figure 4.

Fig. 4 Sales/order data representation and stored in a .CSV file format.

- c) The model was created and tested using WEKA with the Decision Tree Algorithm (J48), Naïve Bayes Algorithm, and the Near Friend Algorithm (K- Nearest Neighbors), which the resulting model is in the form of data classification rules to classify purchase orders into warehouses by product type and size of that item.

**Table 3 Performance values from Cross Validation 10 Folds model testing**

Model Test	Accuracy	Precision	Remembrance	Balance
Decision Tree	86.60	87.10	84.50	86.40
Naïve Bayes	79.77	78.70	78.70	75.66
K-Nearest Neighbors	82.65	81.60	80.44	81.88



**Fig. 5 Performance graphs from model testing.**

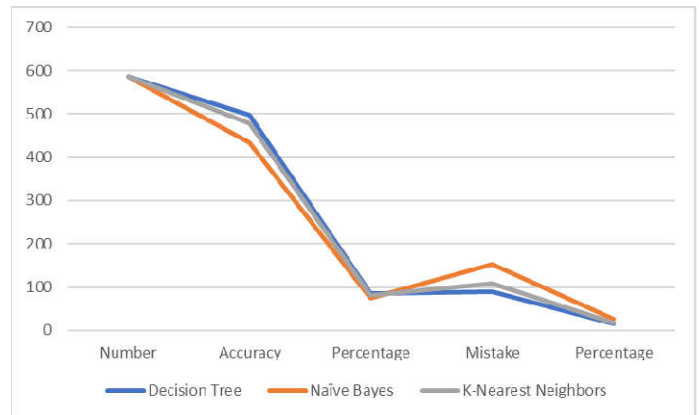
From Table 3 and Figure 5, the results of classification of student uniform order data The model was developed using the Cross Validation 10 Folds method using the Decision Tree (J48) algorithm, Naïve Bayes algorithm, and neighbor search algorithm. It was found that the Decision Tree algorithm (J48) had an accuracy of 86.60%, an accuracy of 87.10%, a recall of 84.50%, and a balance of 86.40%. Naïve Bayes has an accuracy of 79.77%, an accuracy of 78.70%, a recall of 78.70%, and a balance of 75.66%. It has an accuracy of 82.65%, an accuracy of 81.60%, a recall of 80.44%, and a balance of 81.88%.

The above data shows that models derived from the Decision Tree Algorithm (J48) and the K-Nearest Neighbors Algorithm can be used in the development of a model. Classification of student uniform ordering data is very accurate and accurate in classifying data as it is more than 80 percent accurate.

The results of the model test using the results of classification of student order data types using the Decision Tree (J48) algorithm, Naïve Bayes algorithm, and the Naïve Bayes algorithm. The closest approximation to test the accuracy and validity of the model by the confusion matrix. The results of classification of the learning data from the learning data files and the validity and error test data are shown in the table 4.

**Table 4 Data classification accuracy and error**

Classification Algorithm	Number	Accuracy	Percentage	Mistake	Percentage
Decision Tree	587	498	84.84	89	15.16
Naïve Bayes	587	435	74.11	152	25.89
K-Nearest Neighbors	587	479	81.60	108	18.40



**Fig. 6 The graph shows the accuracy and error of the data classification.**

From Table 4 and Figure 6, the results showed that the student uniform order data classification using the Decision Tree (J48) algorithm showed that the data classification was 84.84% accurate and the value error 15.16 percent, Naïve Bayes algorithm 74.11 percent accuracy of data classification and error 25.89 percent. The most similar results of data classification were 81.60percent accurate and 18.40% error, respectively. From the research results, it can be seen that the Decision Tree (J48) has the highest school uniform order data classification value than other techniques, can be used to effectively classify the learning data of the school uniform order.

d) The researchers measured the performance of the algorithm's data classification model to predict the 2022 orders and 2022 actual sales for performance testing. Using the data of 2022, which consisted of 587 product items, the quantity ordered, the researcher chose to classify the student uniform order data using the Decision Tree algorithm. (J48) as shown in Figure 7.

ลำดับ	ประเภทสินค้า	ขนาด	พ.ศ. 2565 จำนวนชุดนักเรียน ที่สั่งซื้อ
1	เสื้อเจ้ต้นักเรียนชายเสวี	เบอร์ 30	60
2	เสื้อเจ้ต้นักเรียนชายเสวี	เบอร์ 32	84
3	เสื้อเจ้ต้นักเรียนชายเสวี	เบอร์ 34	48
4	เสื้อเจ้ต้นักเรียนชายเสวี	เบอร์ 36	60
5	เสื้อเจ้ต้นักเรียนชายเสวี	เบอร์ 38	60
6	เสื้อเจ้ต้นักเรียนชายเสวี	เบอร์ 40	120
7	เสื้อเจ้ต้นักเรียนชายเสวี	เบอร์ 42	108
8	เสื้อเจ้ต้นักเรียนชายเสวี	เบอร์ 44	108
9	เสื้อเจ้ต้นักเรียนชายเสวี	เบอร์ 46	60
10	เสื้อเจ้ต้นักเรียนชายเสวี	เบอร์ 48	36
11	เสื้อเจ้ต้นักเรียนชายเสวี	เบอร์ 50	12
12	เสื้อเจ้ต้นักเรียนชายเสวี	เบอร์ 52	12

**Fig. 7 Example of a forecast of 587 school uniform orders in 2022.**

From Figure 7, it shows a list of 587 purchase order forecasts when the forecast amount to be ordered. Therefore, the researcher must take the decision amount of each order to be deducted from the number of inventories in stock in 2021 by subtracting from the number of products that will be ordered in the number of dozens (12 pieces) per item. Because the order forecast will come out in the number of pieces.

Table 5 Example of comparing the ordered quantity and the actual sold quantity in 2022

No.	Product Type	Size	2022				Accuracy (percent)
			2021 Remaining	Sold	Order	Remaining	
1	free male student shirt	30	0	49	60	11	100
2	free male student shirt	32	6	85	84	5	100
3	free male student shirt	34	2	47	48	3	100
4	free male student shirt	36	4	55	60	9	100
5	free male student shirt	38	3	59	60	4	100

From Table 5 shows the comparison of the ordered quantity and the actual sold quantity in 2022, showing the product type, the size of the product type, the number of inventories in the year 2021, the number of sets. Students sold in 2022, number of school uniforms ordered in 2022 (with dozens of orders and converted into pieces), number of inventories in year 2022 and the percentage of accuracy which from the information in the above table was found to be correct and inventories enough to meet the stock.

## CONCLUSIONS

From the results of the research, the results can be summarized as follows.

1. System analysis and design prototyping and the development of a student order forecast forecasting model found that analyzing the data by classification data was consistent with the development of a student uniform order forecasting model.
2. Development a forecasting model to predict school uniform orders and implement the system to help forecast school uniform orders. By using it in ordering in the year 2022, it was found that it was 100 % accurate that the ordered products and remaining sufficient to sell inventories in stock.

## Acknowledgements

This study was supported by Division of Information and Digital Technology Management, Faculty of Industrial Technology, Phranakhon Rajabhat University, Thailand.

## REFERENCES

1. WatanyutaNeelaphatrakun and ChutimaBeokaimook. (2019). The Study of Associated Factors of the Resignation Decision and the Efficiency Comparison of Various Predictive Models of EmployeeResignation: A Case Study of an Insurance Company. Graduate Student, College of Digital Innovation and Information Technology, Rangsit University.
2. PreechaLimtrakul, Vipajaroenpuntaruk and WittayaPornpatcharapong. (2016). Development of a Model to Predict Cassava Yield Using Data Mining. Vol. 3 No. 3 (2016): Science and Technology.
3. Ballou, R.H. (2004). Business Logistics/ Supply Chain Management. Prentice Hall. Supply Chain Management, 5, pp 234 –245.
4. David, A. and Shijia, G. (2019). Behavioral economics for decision support systems researchers, Decision Support Systems, 122, pp 1-12.
5. Vipajaroenpuntaruk, WaranyaPunnawat, ChailertPichitpornchai, SomlakLaungsri, Suwannee Asavakulchai, Chalerm SakLertwongsathien, ThanitPhusiri, RatchaneekulPinyoPanuwat, Kochakorn Na Nakhon Phanom and ChittinanSuanmalai. (2012). Data warehousing, data mining and business intelligence. Bangkok: Sukhothai Thammathirat Open University Press.
6. Jiawei, H., Micheline, K. & Jian, P. (2012). Data Mining Concept and Technique. 3rd ed. San Francisco, CA: Morgan Kaufmann Publishers is an imprint of Elsevier.
7. Dogan, S. Z., Arditi D. &Gunaydin, H.M. (2008). Using Decision Trees for Determining Attribute Weights in a Case-Based Model of Early Cost Prediction. Journal of Construction Engineering and Management, 134(2), 146-152.
8. Seema Sharma, Jitendra Agrawal, Sanjeev Sharma. Classification Through Machine Learning Technique: C4.5 Algorithm based on Various Entropies. International Journal of Computer Applications 82(16)(2013) 20-27.
9. Pouria Kaviani, Sunita Dhotre. Short Survey on Naive Bayes Algorithm. International Journal of Advance Engineering and Research Development 4 (11)(2017) 607-611.
10. Ballard, C., Farrell D. M., Gupta, A., Mazuela C. &Vohnik S. (2006). Dimensions Modeling In a Business Intelligent Environment. 1nd ed. IBM Corporation, International Technical Support Organization.
11. DursunDelen. (2009). The Three Most Common Data Mining Software Tools. <https://www.sciencedirect.com/topics/mathematics/>.
12. C. Moemeng , V. Gorodetsky , Z. Zuo , Y. Yang , C. Zhang , Agent-based distributed data mining: a survey, in: Data Mining and Multi-Agent Integration, Springer (2009) 47–58.