Anti – Counterfeiting Detection for Near Field Communication (NFC) – Enabled Logistics

Manmeet Mahinderjit Singh†, Tan Han Yee††, and Rohail Hassan†††

†School of Computer Science, Universiti Sains Malaysia, 11700, Penang, Malaysia.
†† School of Computer Science, Universiti Sains Malaysia, 11700, Penang, Malaysia.
††† Department of Management and Humanities, Universiti of Teknologi PETRONAS (UTP), Malaysia. Bandar Seri Iskandar, 32610 Tronoh, Perak, Malaysia

Summary
Logistics is the process which involved in planning, implementing and controlling. Each process must be performed so it can provide maximum effectiveness and efficiency towards the customer. Preliminary study result shows that security factor is the most concern among logistics operation staff because they worried data information that stored in NFC tags been leaked and caused counterfeit items to occur. Hence, it brings lack of trust in the adoption of NFC in logistics. Due to the problem of Near Field Communication, proposed an enhanced approach to detect counterfeiting issue in NFC–enabled logistics to be conducted in this research. The researcher chose using data mining technique and enhances it to specifically detect on counterfeit items in logistics industry by creating detection rules to detect counterfeit items. Based, on the results, the J48 technique is the most accuracy on detect counterfeit items. Hence, research used J48 technique and enhanced with specific rules created to detect on counterfeit items. The result of enhanced J48 technique improved to 99.2% accuracy provides better detection than other data mining technique.

Keywords:
Near Field Communication (NFC), Tags, Counterfeiting, J48 technique, Logistics, Security attack.

1. Introduction

Logistics is a process which involves planning, implementing, and controlling. Each part of the process must be performed so that it can provide maximum effectiveness and efficiency to the customers. The current trend of logistics involves improvisation from the traditionally-based logistics to sensor-based logistics method. This is because sensor-based logistics method can help organisations to track product locations and prevent counterfeit issues. Two current trends in sensor-based logistics method chosen by most organisations are Radio Frequency Identifier (RFID) and Near Field Communication (NFC). Even though sensor-based logistics method is implemented, counterfeit cases still happen in the logistics industry. Counterfeit is creating items that look similar to original items and later are sold in the market for personal benefits. They are products that are manufactured without the original manufacturers' permission. The aim of counterfeit is to create items with the lowest price and the lowest quality. Counterfeit becomes more serious at the global level due to customers' demand for purchasing products at the lowest price [1]. This has led to third party person who starts manufacturing counterfeit items which are later sold to customers who request for the lowest price. As counterfeit is widely grown, it impacts not only manufacturers but also consumers. Once the counterfeit items are sold on the global market, profit of the manufacturers will lessen as some customers prefer to purchase counterfeit items that are cheaper compared to the originals. Moreover, customers will eventually lose trust in the manufacturers as they will not be able to differentiate which items are manufactured by original manufacturers. They will switch to other manufacturers' products that do not have counterfeit issues. In consumerism, safety and health of the consumers play an important part [1]. Counterfeit items manufactured at low quality have risks on ones' lives. As an illustration, there was a case of a pharmaceutical chain which failed to control properly and verify against counterfeit. As a result, the counterfeit items that had been sold with a different ingredient from the original medicine caused 76 people died in the year 2006 [2]. Counterfeit happens in logistics industry as NFC tag is easy to be duplicated. Attackers just need to use NFC Reader which can get all the information that has been stored in the NFC tag. Then, the attackers will duplicate the information and locate it into counterfeit items. This causes organisations to lose profit because the attackers can create counterfeit products and sold them at the global market. Aris Report mentioned that a total of 1.2 billion of USD dollars of counterfeit products have been flowing into the United States market [1]. Once the amount of counterfeit products is large, it may cause the public to lose trust in the organisations. In handling counterfeiting, Intrusion Detection System is widely employed in the logistics industry. One major problem with this system is that it is not able to prevent duplicate tags from happening in the logistics industry. It only monitors activities that are to identify any suspicious event within an organization. However, it is not able to detect duplicate tags [3]. For

Manuscript received February 5, 2018
Manuscript revised February 20, 2018

example, when counterfeit items are decreased with wanted items with NFC tags, Intrusion Detection System will not be able to detect because the system focuses on network detection like intercepting data information through the network and malicious data happening in the network. Besides, a preliminary study was done by the researcher to understand the factors that influence sensor-based method in logistics. Based on the results obtained from the survey conducted with the staffs of a logistics company, it shows that 70 percent of respondents agreed that performance expectancy, effort expectancy, social influence, facilitating condition, and security value are the factors that influence them for using the sensor-based method in the logistics industry. The hypothesis was also tested through Pearson's Correlation, and Linear Regression which shows that facilitating condition and security value are significant factors which influence sensor-based method in the logistics industry. In addition, with the usage of MCDM, two main factors that have impacts on NFC usage are the security and facilitating elements. Overall, security is a major concern raised by the employees for adopting NFC in logistics. Therefore, improvements of NFC technology are needed in order to influence operation staff to implement the sensor-based method in the logistics industry. In this research, the researcher focused on detection of counterfeiting attacks that happen in logistics industry by using data mining approach. Data mining has been implemented to detect duplicate data information to help logistics to intercept counterfeit items before they are flown into the global market.

2. Preliminary Study: Influential Factors of Sensor-Based Tagging in Logistics

Near Field Communication provides convenience to logistics industry as it can track product locations that ease operation staff to locate product locations within the logistics warehouses. Even though Near Field Communication benefits logistics industry, some logistics industries are still willing to use the traditional method to continue their daily businesses. Hence, a preliminary study was conducted to understand the factors that influence sensor-based method in the logistics industry. In order to seek such understanding, a survey was administered by collecting feedback from logistics industry. Once the feedback was collected, Pearson's Correlation and Linear Regression were used to analyze the survey result to prove the hypothesis of the researcher. As proven by the hypothesis, Multi-Criteria Decision Making was used to prioritize which factor was the most influential of the sensor-based method in the logistics industry. Based on the result generated from the Multi-Criteria Decision Making, security value makes the top factor of the sensor-based method in logistics. Therefore, improvement of NFC was needed to influence the operation staff in using NFC in the logistics industry. There were a few steps taken by the researcher to understand the main influential factors of the sensor-based method in the logistics industry. These included:

1. *Data collection via survey instrument*
2. *Data analysis via Pearson's Correlation Result*
3. *Data analysis via Linear Regression*
4. *Data result via Multi-Criteria Decision Making*

Further explanation of the survey, Pearson's Correlation, Linear Regression and Multi-Criteria Decision Making is documented in the following sections.

2.1 Data Collection via Survey Instrument

The survey is the first part of the preliminary study. In order to create a quality survey, pre-test questionnaires were administered to make sure all the questions that had been created were easy to understand. Samples of pre-test questionnaire questions are documented in the appendix section. Before the questionnaires were distributed, the hypothesis was created to assume the result of the survey which needed further study. In this survey, a total of five hypothesis were created. These are:

Performance expectancy is related to behavioral intention to use Near Field Communication technology.

Effort expectancy is related to behavioural intention to use Near Field Communication technology.

Social influence is related to behavioural intention to use Near Field Communication technology.

Facilitating condition is related to behavioural intention to use Near Field Communication technology.

Security value is related to behavioural intention to use Near Field Communication technology.

In order to prove the hypothesis, Pearson's Correlation and Linear Regression were used to prove the significance of the hypothesis. Once hypothesis and pre-test questionnaires were done, questionnaires were distributed to logistics operation staff in order to get feedback from them. A total of 30 respondents were involved in the survey and result is shown in Fig. 1.

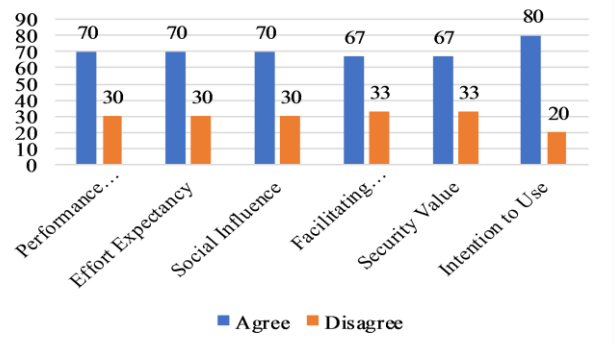

Fig. 1 Survey Result

From Fig. 1, it shows that most of the respondents agreed that performance expectancy, effort expectancy, social influence, facilitating condition, security value, and intention to use NFC could retrieve using the sensor-based method in the logistics industry. It shows that the respondents were willing to use the sensor-based method in their working lives. Next, Pearson's Correlation and Linear Regression explain how the data was being analysed and which hypothesis was supported by this survey.

2.2 Data Analysis via Pearson's Correlation Result

Pearson's Correlation also called Pearson Product Moment Correlation (PPMC) is a method which used to calculate the relationship between the two set of data [4]. In order to understand the relationship between performance expectancy, effort expectancy, social influence, facilitating condition, security value and intention to use researcher had used Pearson's Correlation to calculate it. Table 1 shows the result of six different datasets that used Pearson's Correlation to calculate.

Pearson's correlation analyses were performed and are reported in Table 1. The correlation matrix shows that there was a statistically significant correlation between all variables at significance level 0.01 ($p < 0.01$). It means performance expectancy, effort expectancy, social influence, facilitating condition, security value, intention to use were strongly positively correlated to each other. The variables values that were closer to 1 are strongly correlated. Hence, it can conclude the correlation is significant at the 0.01 level (2-tailed).

Table 1: Pearson's Correlation

Variable	*Mean*	*SD*	*PE*	*EE*	*SI*	*FC*	*SV*	*IU*
Performance Expectancy (PE)	5.300	0.950	1					
Effort Expectancy (EE)	5.242	1.014	**.949** (0.01)**	1				
Social Influence (SI)	5.150	0.993	**.917** (0.01)**	**.911** (0.01)**	1			
Facilitating Condition (FC)	5.133	1.127	**.875** (0.01)**	**.902** (0.01)**	**.879** (0.01)**	1		
Security Value (SV)	5.053	1.044	**.747** (0.01)**	**.796** (0.01)**	**.824** (0.01)**	**.860** (0.01)**	1	
Intention to Use (IU)	5.233	1.014	**.879** (0.01)**	**.902** (0.01)**	**.872** (0.01)**	**.945** (0.01)**	**.893** (0.01)**	1

Note: ** Correlation is significant at the 0.01 level (2-tailed); *Correlation is significant at the 0.05 level (2-tailed).

2.3 Data Analyse via Linear Regression

Linear Regression is the basic common tools that used to predictive analyse [5]. It is a method that attempts to model the relationship between two variables by fitting a linear equation to observe the data. The function of linear regression as it denotes a variable as a dependent variable whose value that wish to predict while another variable to set as an independent variable from which wish to predict it [5]. In this research, Intention to Use NFC technology marked as dependent variable while other five variables like performance expectancy, effort expectancy and so on are marked as the independent variable. Table 2 shows the result of a linear regression which used intention to use NFC technology as a dependent variable to calculate other five variable independent data. Based on result, it shows that improvement on facilitating condition ($\beta = 0.462$, $p < 0.10$), and security value ($\beta = 0.339$, $p < 0.10$), have a positive and significant impact on influence logistics operation staff using NFC technology. Hence, facilitating condition increase by one Beta (β), then intention to use NFC technology increase by 0.462. With regard to security value, one Beta (β) increase in security vale, there is 0.339 increase in intention to use NFC technology. Therefore, it able to attract operation staff uses NFC technology in logistics industry where there is an improvement in facilitating condition and security value. However, performance expectancy ($\beta = 0.250$, $p > 0.10$), effort expectancy ($\beta = 0.100$, $p > 0.10$), and social influence ($\beta = -0.134$, $p > 0.10$) were not significant toward intention to use NFC technology. These results base on three significance level which are * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.01$. Hence, facilitating condition and security value were positive significant toward intention to use NFC technology at 90% confidence interval ($p < 0.10$). Therefore, it concludes that facilitating condition and security value are related to behavioural intention to use NFC technology as both had reached more than 90% confidence interval in linear regression.

Table 2: Linear Regression

Variables	*Intention to Use NFC Technology*
Performance Expectancy (PE)	0.250 (0.205)
Effort Expectancy (EE)	0.100 (0.614)
Social influence (SI)	-0.134 (0.407)
Facilitating Condition (FC)	**0.462* (0.05)**
Security Value (SV)	**0.339* (0.06)**
Constant	0.203 (0.537)

Notes: Dependent Variable "Intention to Use NFC technology." p-values are in parentheses; Significance levels are * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Manuscript received February 5, 2018
Manuscript revised February 20, 2018

2.4 Multi-Criteria Decision Making: Analytic Hierarchy Process (AHP)

Analytic Hierarchy Process (AHP) is the common method that been used for Multi-Criteria Decision Making (MCDM). The most advantage of Analytic Hierarchy Process is it easier to understand. Analytic Hierarchy Process using hierarchy structure method which easier to decision maker to understand the whole process of Analytic Hierarchy Process [6]. Besides, Analytic Hierarchy Process more flexible and easy to check inconsistency compare with other Multi-Criteria Decision-Making method [6]. In this research, Analytic Hierarchy Process used in this research to understand factor of influencing sensor-based method in logistics. In Analytic Hierarchy Process, it separated into four different part, which are:

Define the problem and the criteria.
Structure decision hierarchy table from the top until the goal of the decision.
Construct a set of pairwise comparison matrix.
Use priorities from comparison matrix to determine which criteria had the most priorities.

All the four-different part had their aim and objective in order to get the most priorities criteria towards decision maker. Hence, AHP used to analyse on survey result, and the result shows that security value is the most priorities of the influence sensor-based method in the logistics industry. Table 3 shows the priorities table that been calculated through AHP method.

Table 3: Priorities Table

Factor of influencing sensor-based method in logistics							
Criteria	***Performance Expectancy***	***Effort Expectancy***	***Social Influence***	***Facilitating Condition***	***Security Value***	***Total***	***Priority***
Performance Expectancy	0.06	0.01	0.02	0.05	0.15	0.29	0.06
Effort Expectancy	0.28	0.07	0.02	0.04	0.15	0.56	0.11
Social Influence	0.28	0.35	0.11	0.05	0.11	0.90	0.18
Facilitating Condition	0.22	0.35	0.43	0.21	0.15	1.36	0.27
Security Value	0.17	0.21	0.43	0.64	0.44	1.88	0.38
Total	1.00	1.00	1.00	1.00	1.00	5.00	

Table 3 showed priorities table result of factor influencing sensor-based method in logistics. The result of each criteria factor value based on the pairwise comparison that been made by researcher and the result of each criterion weighted by a researcher in pairwise comparison based on Pearson's Correlation result. Criteria factor value calculated based on the value of criteria divide with the total result of other criteria compared to the pairwise comparison. For example, the total value of performance expectancy in pairwise comparison is 18 while the value of effort expectancy is 5. To calculate criteria factor value of effort expectancy towards performance expectancy is five divides by 18 and the result of effort expectancy towards performance expectancy is 0.28 (5/18 = 0.28). Therefore, the result of effort expectancy towards performance expectancy in criteria factor value is 0.28. In priority column, it shows the priority result of each criterion. It calculated based on a total of criteria value divide with the total result of five criteria result. For example, total criteria value of security value is 1.88 while the total result of five criteria result is 5. Priority of security value result is 0.38 (1.88 / 5 = 0.38). If the priority result is lower, it means that criteria do not bring the most factor of the influence sensor based method in logistics. In another word, if the criteria priority result is higher, then it is the main factor of the influence sensor based method in logistics. From table result, Researcher understands that security value is the most priority criteria of logistics industry to using the sensor-based method in a daily routine that scored 0.38 mark. Second priorities are facilitating condition scored on 0.27 mark, third, fourth and fifth priorities are a social influence, effort expectancy and performance expectancy that all scored on 0.18, 0.11 and 0.06 marks. Therefore, it able to conclude that improvement of NFC security is needed so that it can influence operation staff using the sensor-based method in the logistics industry.

3. Background Study

In this chapter, the researcher had conduct background study that related to this research project that contains on security challenges of Near Field Communication (NFC) and Data Mining technique.

3.1 Near Field Communication Security Challenges

Near field, communication is a method which allows people to transfer information in within the shortest range. As near-field communication uses wireless technology, it surely has some security challenges. Security challenges which happen in near field communication are eavesdropping attack, data modification attack, data manipulation attack, counterfeiting attack, information injection attack, relay attack, and tag cloning. Further explanation of Near Field Communication security challenges is documented below.

Eavesdropping Attack

Eavesdropping attack is one of the security challenges in all wireless network technology. It can happen when there are two or more devices transfer data with each other [7]. In other words, eavesdropping attack is an attack which can intercept all data transmission between two or more devices. The aim of the attacker in performing eavesdropping attack is to steal users' personal information like bank account password and user confidential data.

Data Modification Attack

Data modification attack is also one of the security challenges in near field communication. Data modification occurs when there are third party people who change all data information without any authorised approval. In other words, the attacker attempts to intercept data packet from the network during a network transmission [8]. Once the data packet has been intercepted, the data information is modified and put back into the network, and the network continues the transmission to the recipient [8].

Data Manipulation Attack

Data manipulation attack is one of the security challenges that happen in near field communication. Data manipulation is also called corruption, which means the attacker intercept all information that is sent out by the sender, corrupt the data and then send them back to the recipient [8]. Data manipulation attack aims to destroy all the information or prevent a recipient from receiving correct data information that has been sent out by the sender [8].

Counterfeiting Attack

Counterfeiting Attack is one of the possible attacks that happen in near field communication. Counterfeiting attack occurs when there are third party people who want to include counterfeiting products within the original items sold into the global market. It causes the organisations to lose profit by it. Such thing happens when the attacker uses intercept method. In order to get product information on the original items, the attacker uses intercept method to sniff all information from NFC tags. Once the product information is obtained, then the attacker put the information on counterfeit products and sold them into the global market. It is difficult for consumers to identify the counterfeit product as all the product information is similar to the original items.

Tag Cloning

Tag Cloning mostly occurs in near field communication. Tag cloning occurs as it is interrelated to counterfeiting. In order to perform tag cloning, an attacker needs to get product information that denotes inside the tags. Therefore, interception is used to get data information [9]. When the user wants to get information on the tags, a tag reader is required to read the information on it. As the reader uses wireless technology to transfer information from tag to reader, it gives the attacker a chance to intercept the information.

Relay Attack

Relay attack is also called as Man in Middle Attack (MIMA). The function of relay attack is to intercept information when there are two devices communicate with each other [10]. When there is data transmission between both devices, relay attack intercepts data information from the network, make changes to the data information, and then send it back to the recipient. Relay attack aims to modify the information of data so that the recipient does not receive the correct information [10].

Information Injection Attack

Information injection attack is also called buffer overflow. Information injective happens when there is a coding mistake inside the software [11]. When there is a mistake in the software code, it gives a chance for an attacker to modify the code inside the software. Information injection aims to overwrite local application program and run on attacker program [11]. When the victim uses attacker's application, then the attacker can get all information from the victim, like login information and personal information. In order to compromise the whole victim system, the attacker is only able to use programming languages like C, C++, or Fortran [11].

3.2 Data Mining Techniques

As the weakness of Intrusion Detection System is not able to detect counterfeit items, Data Mining was used in this research project. Data Mining is also known as knowledge discovery in the database [12]. It is a method that uses a large dataset of information to identify interesting and useful patterns to solve selected problems [12]. Currently, data mining is being widely used in business, science research, government, and so on. In this research, the researcher explained the type of classification data mining like J48, JRip, Naïve Bayes, Logistics Model Tree, and Decision Tree which are commonly used by scientists or researchers on problem-solving. Further explanation of these five different types of classification data mining is documented as below.

J48

J48 is one of the decision trees that is categorised as a type of Data Mining technique. J48 is an extended version of the C4.5 algorithm. In the J48 algorithm, it creates a structured tree method which is from the topmost root node until the leaf node. Besides, the J48 algorithm uses Iterative Dichotomiser 3 (ID3) algorithm [13]. The biggest advantage of the J48 algorithm is the model generated by decision tree which helped the researcher to predict new instances of data.

JRip

JRip is one of the data mining classifier techniques. It is the basic and most popular algorithm. In JRip, all the classes are examined by increasing the size and the first set of rules for the class. They will be generated by using

Repeated Incremental Pruning to Produce Error Reduction (RIPPER) to reduce error in the JRip algorithm [14].

Naïve Bayes

Naïve Bayes algorithm is one of the data mining classifier techniques. Naïve Bayes classifier uses simple probability classifier method based on Bayes theorem with independence assumptions between predictors [15]. It is suitable for use when the dimensionality of input is high as Naïve Bayes algorithm is only able to predict which probability that a given tuple belongs to a particular class [15].

Logistics Model Tree

Logistics Model Tree (LMT) is one of the data mining classification techniques. Logistics Model Tree is the combination of linear regression and induction tree [16]. Logistics Model Tree (LMT) uses cost-complexity pruning which is significantly slower than other algorithms as most of the other algorithms like J48 and JRip use ten different random seeds to perform all the test runs while Logistics Model Tree is only restricted to five different random seeds [16]. Therefore, logistics model tree uses more time compared to other algorithms.

Decision Tree

Decision Tree is one of the Data Mining classification techniques. In a decision tree, it includes a root node, branches, and leaf nodes. The root node is the topmost node in the whole decision tree structure. The topmost node separates into branches and contains outcome of a test. From the branches, it separates into leaf node and contains on a class label on it [17]. The advantage of implementing decision tree is it does not need any knowledge to implement it [17].

Table 4 shows the comparison of five different data mining classifier techniques.

Table 4: Comparison of five different classifier techniques

Classifier	***Description***
J48	It is extended from a C4.5 algorithm that uses tree structure method from the topmost root node until the leaf node.
JRip	It uses Repeated Incremental Pruning to Produce Error Reduction (RIPPER) method to find out a set of rules to cover all the number of training data.
Naïve Bayes	It uses simple probability classifier method that is based on Bayes theorem with independence assumptions between predictors.
Logistics Model Tree	It relies on simple regression models with a complex tree structure to create the whole structure of Logistics Model Tree.
Decision Tree	It uses tree structure method, which begins from the topmost root node to branches that contain outcome of a test than to a leaf node which contains a class label.

4. Classifier Benchmarking and Proposed Detection Technique

The researcher had chosen five common data mining classifiers to experiment. They were Naïve Bayes classifier, J48 classifier, JRip classifier, LMT classifier, and Decision Tree Classifier. Throughout the experiment, J48 classifier provided the most optimum result in detecting counterfeit items. Hence, the researcher chose J48 classifier and enhanced it with rules created by the researcher that specifically detected counterfeit items. Further explanation of the experiment on the current data mining classifier and the enhanced data mining classifier technique is documented in the following section.

4.1 Experiment on Data Mining Classifier

Data Mining Classifier can provide the optimum result to tackle counterfeiting issue in the logistics industry. In order to understand which data mining classifier provides the most accurate detection of counterfeit items, the researcher chose five common Data Mining classifiers to be experimented on. The five Data Mining classifiers were Naïve Bayes algorithm, J48 algorithm, JRip algorithm, LMT algorithm, and Decision Tree algorithm. Steps in experimenting with the five different data mining classifiers are documented in the following section.

4.1.1 Process of Experiment Data Mining Classifier

In order to perform Data Mining classifier experiment, a total of four phases were done by the researcher. These included creating a dataset, Weka software, import dataset into Weka, and generate a result. Each phase had its aim and objective in order to generate the final result of the Data Mining classifier.

4.1.2 Results of the Experiment

As a result of the different data mining classifiers, the researcher used a dataset which consisted of 1000 data instances. Table 5 shows the five different classifier results which used Weka software to generate results.

Table 5: The result of different data mining classifiers

Classifier	***Correct Classifier***	***Incorrect Classifier***
Naïve Bayes	87.9%	12.1%
J48	99%	1%
JRip	98.5%	1.5%
LMT	98.8%	1.2%
Decision Tree	98%	2%

Based on Table 7, it shows that J48 algorithm performed as the best Data Mining classifier to tackle counterfeiting issue in the logistics industry. Based on these results, J48 algorithm provided the best detection as it scored 99% accuracy within the dataset. The second-best classifier was

LMT algorithm. It was able to provide 98.8% accuracy while the remaining JRip, Decision Tree, and Naïve Bayes algorithms were only able to provide 98.5%, 98%, and 87.9%% accuracy in the dataset. Hence, the researcher chose J48 algorithm to be enhanced in order to increase the accuracy on detect counterfeit items to prevent counterfeit issues from happening in the logistics industry.

4.1.3 Statistical Test on the Experiment

In order to prove the experiment that was conducted, a statistical test was performed to make sure J48 classifier perform the best accuracy in detecting counterfeit items in the logistics industry. To perform the statistical test, Weka experimenter was used with other five common classifiers with a single dataset. The statistical test aims to prevent bias cases from happening. In a Weka experimenter, 95% confidence interval level was set to analyse all five common classifiers. If classifier was less than 95% confident interval than it was not significant enough to detect counterfeit items. Table 6 and Fig 2 show the experiment results of five common classifiers through a statistical test.

Table 6: Statistical test of five common classifiers

Classifier	*Correct Classifier*	*Incorrect Classifier*
Naïve Bayes	87.9%	12.1%
J48	98.93%	1.07%
JRip	98.77%	1.23%
LMT	98.87%	1.13%
Decision Tree	98.27%	1.73%

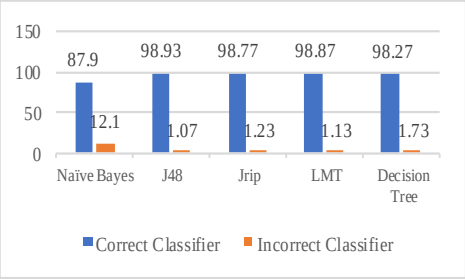

Fig. 2 Statistical test of five common classifiers

Based on Table 8 and Fig 2, J48 classifier still maintained the best classifier as it scored 98.93% accuracy in detecting counterfeit items in the logistics industry. The second-best classifier was LMT classifier which scored 98.87% accuracy. The remaining classifiers which were JRip classifier, Decision Tree classifier and Naïve Bayes classifier scored 98.77%, 98.27%, and 87.9% accuracy, respectively. Throughout the statistical test, it was able to conclude that J48 classifier provides the best accuracy in detecting counterfeit items in the logistics industry. Next, enhanced J48 detection technique was introduced by improving the rules to detect counterfeit items specifically.

4.2 Enhanced J48 Algorithm Detection Technique

As the researcher experimented with various data mining classifiers, J48 algorithm provided the most accurate of the dataset. Hence, the researcher decided to use J48 classifier and enhanced it to become a new classifier to detect counterfeit items in the logistics industry specifically. The improved current data mining classifier aims to prevent counterfeit items from flowing into the global market. Further explanation of enhanced J48 algorithm is documented in the following section.

4.2.1 Enhanced J48 Algorithm Process

In order to provide an understanding of how enhanced J48 algorithm detection technique works in the logistics industry. Once the dataset is loaded into the enhanced J48 algorithm detection, refining the dataset takes place on it. The function of refining the dataset is to recheck data information in the dataset that has been inputted by the user [18]. In this research project, rules were created to make sure it did not contain any misclassified data information in the dataset. Rules created by enhanced J48 classifier are:

If NFC tag in dataset shows two similar identifier numbers, then it is remarked as a counterfeit item.

If product NFC tag number is scanned more than three times, then it is remarked as a counterfeit item.

If the mean value of time taken is more than 120 minutes, then it is remarked as a counterfeit item.

If the date of delivering items is more than one day, then it is marked as a counterfeit item.

Based on the rules created, the enhanced J48 classifier can differentiate counterfeit items in the dataset. Once the dataset was refined, train the classifier occurred as cross-validation to calculate the accuracy of the dataset was implemented.

4.2.2 Result

Once the dataset goes through all the three steps, the result will be shown in the final step. As a result, the researcher can understand the accuracy of data information in the dataset. If the result showed some incorrect classified data, then there might be chances that the counterfeit items occur in the logistics industry. Moreover, there is some other result information like kappa statistics, mean absolute error, and root mean squared error. In this research project, the result was generated using enhanced J48 algorithm detection technique, showing that the total of 99.2% accuracy was the highest accuracy compared with the existing J48 classifier. Even though enhanced J48 algorithm did not provide full accuracy to detect all counterfeit items, but it helped in reducing counterfeit items which flowed into the global market and reducing organisation profit loss.

4.3 Result of the Enhanced J48 Classifier Detection Technique

In order to test the enhanced J48 classifier detection technique, the researcher infused 1000 data instances into the dataset as the more the data instances; the more accurate the data detection will get from enhanced J48 algorithm detection technique. The result of the 1000 data instances dataset is tabulated in Table 7.

Table 7: The Result of the Enhanced J48 Classifier Detection Technique

Classified	***Enhanced J48 Classifier***		***Normal J48 Classifier***	
Correctly Classified Instances	992	99.2%	990	99%
Incorrectly Classified Instances	8	0.8%	10	1%
Kappa Statistics	0.9521		0.9401	
Mean absolute error	0.0084		0.0127	
Root mean squared error	0.0902		0.1003	
Relative absolute error	5.1084 %		7.7125%	
Root relative squared error	31.5214 %		35.04%	
Total Number of Instances	1000		1000	

Table 9 shows the result of the enhanced J48 algorithm detection technique. Based on the result, the enhanced J48 algorithm detection technique was able to provide 99.2% accuracy in the dataset. As compared with existing J48 algorithm detection technique, enhanced provide more accuracy in the detect counterfeit item. It can prevent more counterfeit items sell into the global market because counterfeit items do not have quality control on product items that able to affect consumer health and safety. Even though the enhanced J48 algorithm detection technique was not able to provide 100% accuracy in the dataset, but at least it was able to detect about 99.2% accuracy, which was almost near to 100% accuracy. This helps logistics industry to detect counterfeit items more accurately and at the same time reduce counterfeit items from flowing into the global market.

5. Evaluate on Enhanced Detection Technique

In order to evaluate the enhanced J48 classifier detection technique, the researcher had chosen five existing data mining classifier detection techniques to compare with the enhanced J48 algorithm detection technique to make sure it provides better detection compared to other detection techniques. Future explanation of the comparison with five detection techniques is documented in the following section.

5.1 Comparison with the Enhanced J48 Algorithm Detection Technique

In order to prove that the enhanced J48 algorithm detection technique can provide better detection than other detection techniques, the comparison is needed to perform by using the same dataset that contains 1000 data instances to analyse with different detection techniques. The researcher had chosen J48 algorithm, JRip algorithm, Decision Tree algorithm, Logistics Model Tree algorithm, and Naïve Bayes algorithm to compare with the enhanced J48 algorithm. Table 8 shows the overall comparison of five different data mining classifier detection techniques with the enhanced J48 algorithm.

Table 8: Comparison of Five Detection Techniques with Enhanced J48 Algorithm

Classified	***Correctly Classified Instances***	***Incorrectly Classified Instances***
Naïve Bayes	87.9%	12.1%
J48	99%	1%
JRip	98.5%	1.5%
LMT	98.8%	1.2%
Decision Tree	98%	2%
Enhanced J48	99.2%	0.8%

Table 8 shows the comparison of five detection techniques with enhanced J48 algorithm detection. It states that enhanced J48 algorithm detection technique provided better accuracy in detecting counterfeit items in the supply chain, in which it was able to detect about 99.2% accuracy. This makes the highest accuracy compared to other existing data mining classifiers.

5.2 Statistical Test on Enhanced J48 Technique

To prove that enhanced J48 technique provided better detection compared to other classifiers, a statistical test was used in this section. The function of the statistical test was to prove the algorithm technique which was significantly better than other classifiers. In this research project, Weka experimenter was used to performing the statistical test with other five common classifiers to prove enhanced J48 technique provided the best detection in counterfeit items. In Weka experimenter, 95% confidence interval level was set to analyse all five common classifiers. If classifier was less than 95% confident interval, then it was not significant enough to detect counterfeit items. Table 9 shows the result of enhanced J48 technique with other five common classifiers through the statistical test.

Table 9: Statistical Tests of Five Common Classifiers and Enhanced J48 Technique

Classified	*Correctly Classified Instances*	*Incorrectly Classified Instances*
Naïve Bayes	87.9%	12.1%
J48	98.93%	1.07%
JRip	98.77%	1.23%
LMT	98.87%	1.13%
Decision Tree	98.27%	1.73%
Enhanced J48	99%	1%

The statistical test shows that enhanced J48 technique was still the best detection to detect counterfeit items in the logistics industry. Enhanced J48 technique was able to score 99%, which was the best detection compared to common classifiers like JRip (98.77%) and LMT (98.87%).

6. Discussion

In this discussion, it had separated into five different parts. There are:

6.1 Preliminary study of influence sensor-based tagging in the logistics industry.

A preliminary study conducted by the researcher to understand factor of the influence sensor-based method in the logistics industry. The researcher used survey method to collect feedback from logistics operation staff and analysed survey result through Pearson's Correlation and Linear Regression. As a result, facilitating condition and security value is significantly related to behavioural intention to use Near Field Communication technology in the logistics industry. The researcher are investigated which factors top priorities influence are a sensor-based method in the logistics industry, Multi-Criteria Decision Making used to calculate priorities of all factors. Throughout Multi-Criteria Decision Making, security value is the top priorities that influence sensor-based method in the logistics industry. Due to it, improvement of NFC technology needed to prevent the counterfeit attack and improve logistics operation staff confidence using Near Field Communication in the logistics industry.

6.2 Comparison of Data Mining classifier

In order to prevent a counterfeit attack, happen in the logistics industry, data mining used to detect counterfeit items in the logistics industry. The experiment has been conducted by the researcher to understand which data mining classifier provide an optimum result on detect counterfeit items. During the experiment, the researcher created a dataset that contains 1000 data instances that injected with 150 counterfeit items and researcher had chosen five common data mining classifier that been used in the logistics industry. There are Navïe Bayes classifier, LMT classifier, J48 classifier, JRip classifier and Decision Tree Classifier. Throughout the experiment, it shows that J48 classifier provides the most optimum result which able to detect 99% detection accuracy in counterfeit items. As J48 classifier provide the most optimum result, the researcher chose J48 classifier and enhanced it to become a classifier that specifically detects counterfeit items in the logistics industry.

6.3 Enhance Data Mining Classifier Technique

To improve the current J48 classifier, the researcher had created rules that specifically detect on counterfeit items through the dataset. Rules that created are:

If NFC tag in dataset shown two similar identifier number, then it remarked counterfeit items.

If product NFC tag number been scanned more than three times, then it remarked counterfeit items.

If the mean value of time taken more than 120 minutes, then it remarked counterfeit items.

If the date of delivering items more than one days, then it marked as counterfeit items.

Based on the result, enhanced J48 classifier improved to 99.2% accuracy detection in detect counterfeit items in the logistics industry. Hence, the first objective of this research archived.

6.4 Evaluation of enhanced data mining classifier

Once enhance J48 classifier created, evaluation performed to prove to enhance J48 provide better detection compare with other data mining classifier. The researcher had chosen five common classifiers to compare with enhancing J48 classifier. In order to perform the evaluation, Weka software used by analyse the dataset and provide a result on which classifier provide the most accurate detection. Based on the result, enhanced J48 technique provide better detection (99.2%) compare with another classifier like J48 classifier (99%), JRip classifier (98.5%), LMT classifier (98.8%), Decision Tree classifier (98%) and Navïe Bayes classifier (87.9%).

6.5 Statistical test on enhancing data mining classifier

Statistical test performed in enhanced J48 technique. The statistical test used to prevent bias cases happen in comparison result. Hence, the statistical test set at 95% confidence interval which is the result less than 95% accuracy, then it would not be able to prevent a counterfeit attack in logistics. Based on the statistical result, enhanced J48 technique still maintain the best detection accuracy (99%) compare with another classifier like J48 classifier (98.93%), LMT classifier (98.87%), JRip classifier

(98.77%), Decision Tree classifier (98.27%) and Navïe Bayes classifier (87.9). It can conclude that enhanced J48 technique provides the best detection to detect counterfeit items in the logistics industry. Therefore, the second objective of this research project been archived.

7. Conclusion

Counterfeiting attack is attacker using intercept method to intercept all the information during the data information transmit and transfer data information into the duplicated tag. In order to solve counterfeiting issue, data mining classifier had been used to solve the counterfeit issue. It covered on the basic introduction and background study of logistics, Near Field Communication, security challenges of Near Field Communication and Data Mining technique. Besides, the preliminary study also covered to understand factor of the influence sensor-based method in the logistics industry. The result shows security value is the most factor that able to influence operation staff to convert to the sensor-based method. TO understand which data mining technique provides the most optimum result, experiment on various data mining classifier had to be performed. The researcher used Weka software which helps to calculate out which classifier provide most optimum result in the dataset. Based on the result, the J48 algorithm provides the most optimum result that provides 99% accuracy on the dataset.

References

[1] Arise Report 2016. Counterfeit Products Are Becoming an Epidemic In The US, viewed 15 February 207, Retrieved from http://www.arismoving.com/shocking-consequences-counterfeit-products/. >

[2] Culzoni, M.J., Dwivedi, P., Green, M.D., Newton, P.N. and Fernández, F.M., 2014. Ambient mass spectrometry technologies for the detection of falsified drugs. MedChemComm, 5(1), pp.9-19.

[3] Shari, L. P. & Charles, P. P 2003. Security in Computing, 3rd eth, Prentice Hall, US.

[4] Deborah, J.R. 2007. 'Intermediate statistics for dummies'. Wiley & Sons, Hoboken, NJ.

[5] George, A.F. 2003. 'Linear regression analysis'. Wiley & Sons, Auckland, New Zealand.

[6] Martin, A. 2017. 'A survey on Multi Criteria Decision Making Methods and Its Applications'. American Journal of Information System, vol1. Viewed 20 February 2017. Retrieved from: http://pubs.sciepub.com/ajis/1/1/5/.>

[7] Wang, G. 2015. 'Intercept behaviour analysis of industrial wireless sensor networks in the presence of eavesdropping attack'. IEEE Transactions on Industrial Informatics, vol12, viewed 19 February 2017, Retrieved from: http://ezproxy.usm.my:2092/document/7029608/.

[8] Iman, M. 2014. 'A review of types of security attacks and malicious software in network security'. International Journal of Advanced Research in Computer Science and Software Engineering, vol4, viewed 19 February 2017. < https://www.ijarcsse.com/docs/papers/Volume_4/5_May2014/V4I5-0465.pdf>.

[9] Luke, M. 2013. 'Exposing Clone RFID tags at the reader'. In 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications. IEEE Conference Publications, Melbourne, VIC, Australia PP1669-1674.

[10] Zhao, W. 2016. 'Implementation and analysis of a practical NFC relay attack example'. In 2012 Second International Conference on Instrumentation, Measurement, Computer, Communication and Control. IEEE Conference Publications, Harbin, China, PP143-146.

[11] Paul. E.B. & Irena, B. 2016. 'Defeating buffer overflow: a trivial but dangerous bug'. IT Professional, vol18, viewed 19 February 2017, Retrieved from: http://ezproxy.usm.my:2092/document/7763738/.

[12] Charu, C. A. 2015 'Data mining: the textbook'. Springer, New York USA.

[13] Tina, R.P. 2013. 'Performance analysis of naïve bayes and J48 classification algorithm for data classification'. International Journal of Computer Science and Applications, vol6, viewed 20 May 2017, Retrieved from: http://researchpublications.org/IJCSA/NCAICN-13/189.pdf >.

[14] Anli, R. 2017. 'J48 and Jrip rules for E-Governance data'. International Journal of Computer Science and Security (IJCSS), vol5, viewed 20 May 2017. < http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.232.4477&rep=rep1&type=pdf>

[15] Tahira, M. 2016. 'A machine learning approach for student assessment in e-learning using Quinlan;s C4., naïve bayes and random forest algorithms'. In 2016 19th International Multi-Topic Conference (INMIC). IEEE Conference Publications, Islamabad, Pakistan, PP1-8

[16] Niels, L. 2003, 'Logistics Model Trees', Diploma thesis, University of Freiburg, viewed 20 May 2017. <http://www.cs.uni-potsdam.de/~landwehr/diploma_thesis.pdf>.

[17] Zhou, Z. 2017. 'A novel method of transformer fault diagnosis based on k-mediods and decision tree algorithm'. In 2017 1st International Conference on Electrical Materials and Power Equipment (ICEMPE). IEEE Conference Publications, Xi'an, China, PP369-373.

[18] Dai, W. 2004. 'Transferring naïve Bayes classifiers for text classification'. In AAAI'07 Proceedings of the 22nd national conference on Artificial intelligence. AAAI Press, Vancouver, British Columbia, Canada, PP540-545.

Manmeet Mahinderjit Sing Has received PhD in Data Security from The University of Queensland, Brisbane, Australia. She is and working Senior Lecturer in USM, Malaysia. To date, she has published more than forty (40) international refereed journals and conference proceedings (local and overseas). Her research interests fall in the areas of Information Security, IoT Sensors & Applications Security, Data Mining Security and Mobile Security.

Rohail Hassan is currently pursuing PhD in Management Sciences and working as Graduate Research Assistant at Universiti Teknologi PETRONAS (UTP), Malaysia. He obtained his Master of Philosophy (TQM) in 2012 and BS (HONS) in Business & Management Sciences from University of the Punjab, Pakistan in 2010 consecutively. He is also working as freelancer trainer. To date, he has published more than thirty (30) international refereed journals and conference proceedings (local and overseas). His research interests fall in the areas of Corporate Governance, Strategic Management, Women Empowerment, Gender and Diversity issues, Diversity Management, Financial Management, TQM, SCM, Multidiscipline Studies, i.e., IT and Computer Sciences, and Firm Performance.