# Development of Predictive Models using Machine Learning Algorithms for Food Adulterants Bacteria Detection

Timothy M. Amado[1], Ma. Rica Bunuan[2], Relamae F. Chicote[3], Sheila May C. Espenida[4], Honeyleth L. Masangcay[5], Camille H. Ventura[6]
*Electronics Engineering Department, College of Engineering*
*Technological University of the Philippines - Manila*
[1]timothy_amado@tup.edu.ph, [2]rica.bunuan@tup.edu.ph, [3]relamae.chicote@tup.edu.ph, [4]sheilamay.espenida@tup.edu.ph,
[5]honeyleth.masangcay@tup.edu.ph,[6]camille.ventura@tup.edu.ph

***Abstract*—One of the necessities of human to survive is food and meat is one of mainly consumed food by humans. Thus, a level of quality of food is a must to be safely consumed. There have been some cases of adulteration of meats, which can cause harm to consumers. Adulteration can lead to bacteria contamination which are difficult to determine the presence of bacteria without an instrument or food laboratory tests. Nowadays, the idea of applying machine learning in the field of food microbiology is becoming a trend. And one of these applications is on detection and classification of bacteria in food products. Hence, this study aims to apply machine learning algorithms to construct predictive models to detect the presence of bacteria such as Escherichia Coli and Staphylococcus Aureus in raw meat and determine which model is best through accuracy and cross-validation. In this study, five machine learning algorithms are used which are K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), Naïve-Bayes Classifier (NB), and Artificial Neural Network (ANN). All models are implemented effectively each having an accuracy of 94.97%, 91.84%, 97.57%, 61.46%, and 66.84% respectively. A web application is created using the shiny package in R to attain a standalone application used to show the detected bacteria.***

***Keywords: Adulteration, Escherichia Coli, Staphylococcus Aureus, Machine Learning, Cross-Validation, Random Forest (RF), Artificial Neural Networks (ANN), Support Vector Machine (SVM), K-Nearest Neighbors (h), Naïve-Bayes (NB)*** *(key words)*

## I. INTRODUCTION

Food is the necessity of human in order to survive in daily life and one of the most consumable food by humans are related to meat products. But meat products, especially when not proper handled and preserved will cause adulteration of meat resulting to diarrheal diseases or known as foodborne illness. Some cases of adulteration of meats are reported in which people suffers from diarrheal diseases. Adulteration can come to bacteria contamination which are difficult to determine the presence of bacteria without an instrument or food laboratory tests. Based on the Centers for Disease Control (CDC), Campylobacter, E. coli O157:H7, Salmonella, Staphylococcus aureus, Clostridium perfringens, Norovirus, Listeria monocytogenes and Toxoplasma gondii are the most common microorganisms causing foodborne illnesses [1]. The main purpose of food analysts is to determine and measure the bacteria that have useful and harmful effects on the quality and security of raw foods, including meat products. A food safety technique is emerging which utilizes a predictive model for the detection and classification of bacteria. It is possible to be a powerful tool for growth and assessment of methods to expand the stability and security of food products in meat industry [2]. The application of predictive modelling in the field of food microbiology is essential and beneficial to human. This improves the quality and security of food products.

The purpose of this research is to construct a predictive model such as ANN, KNN, SVM, RF, and NB that can be used to detect the presence of bacteria such as E. Coli O157:H7 and Staphylococcus Aureus, and determine which model is best through accuracy and cross-validation. Also, the researcher's intention is to reduce the time-consuming food laboratory tests and costs. This study will contribute to the knowledge of using predictive models in the field of food microbiology for safety and quality assurance of meat products in the industry.

## II. METHODOLOGY

The methods used for the development and application of the research is shown in this chapter.

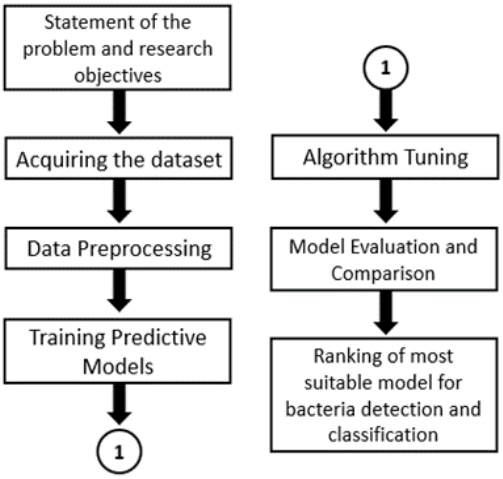


**Fig.1** Research Process Flow

Data pre-processing includes an order of phases to change raw data to form a clean and tidy dataset preceding to numerical analysis [3]. Tuning is typically a trial-and-error process by changing some hyperparameters and comparing its performance on the validation set in order to identify which set of hyperparameters obtained the most accurate model [4]. The system is composed of three sections. These are the data acquisition, development of predictive models, and selecting the most suitable model for bacteria detection and classification.

*a. Data Acquisition*

The dataset is acquired from the collected ppm values from the emitted gases of meat. These are sent to an open source terminal, Tera Term that serves as a serial monitor and saves directly to csv format required and uploaded in shiny web app for the prediction of bacteria.

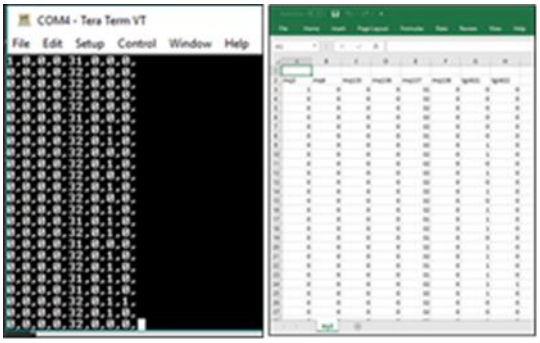


**Fig.2** Acquiring data using open source terminal Tera term and saved in Excel (.csv)

*b. Development of Predictive Models using Machine Learning Algorithms*

Five machine learning algorithms are utilized to build the predictive models. The algorithms that are used are Naïve-Bayes classifier (NB), Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), and K-Nearest Neighbors (KNN).

*i. Model training and algorithm tuning using Support Vector Machine*

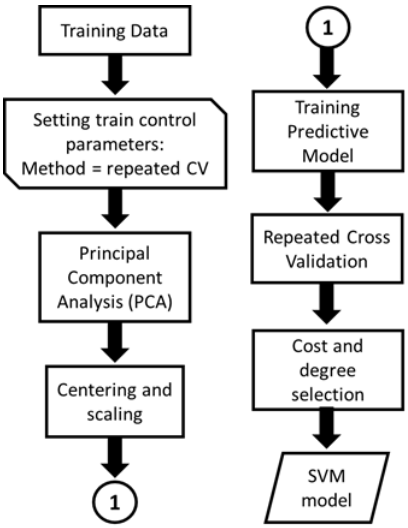


**Fig.3** Training ang tuning flowchart of SVM model

One learning algorithm used for regression and classification problems is the Support Vector Machine [5]. In training the model, training control parameters are first set. Repeated cross-validation technique as the method for algorithm tuning is done and a 10-fold, 10 repeat algorithms were set for the system. Centering and scaling are applied to normalize the data making non-uniform ranges to be standardized [6]. After setting the parameters, the training of SVM model proceeds with a repeated CV as mention, cost and degree selection that implies the largest value to select the optimal model, resulting to the developed SVM model.

*ii. Model training and algorithm tuning using K-Nearest Neighbors*

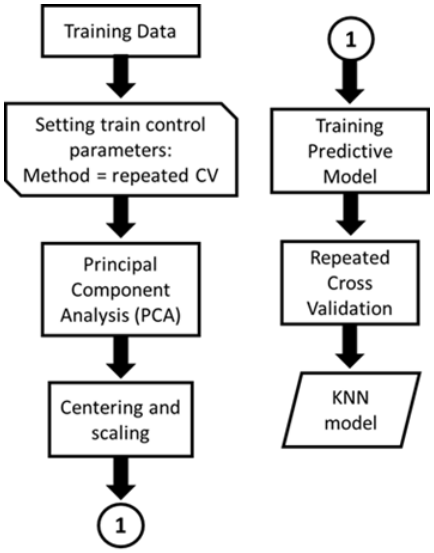


**Fig.4** Training ang tuning flowchart of KNN model

An algorithm that keeps all existing instances and categorizes new instances by means of similarity measure is the K-Nearest Neighbors. KNN can be applied for regression and classification problems in the industry [7].

*iii. Model training and algorithm tuning using Random Forest*

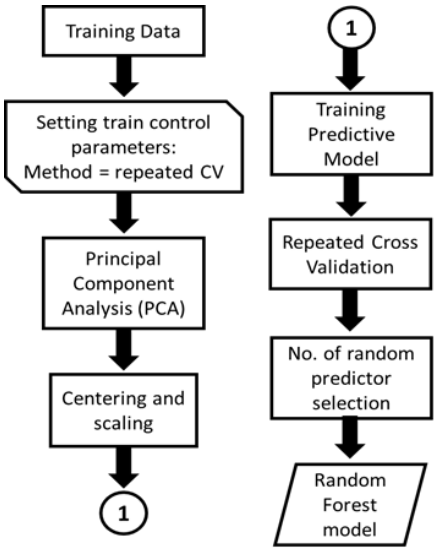


**Fig.5** Training ang tuning flowchart of RF model

Random forest algorithm uses tree as a basis of prediction. Using additional trees within the forest will lead to high accuracy outcomes. It is usually applicable for

regression and classification tasks, handle missing values and manage overfitting model [8]. The training and tuning of Random Forest model are the same as the previous models the only difference is that this model depends on the assumption of decision trees tested erratically as subsets of the training set.

*iv. Model training and algorithm tuning using Artificial Neural Network*

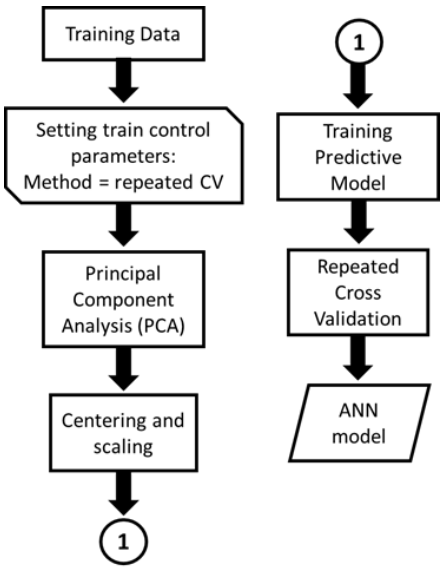


**Fig.6** Training ang tuning flowchart of ANN model

Artificial Neural Network is another machine learning algorithm and was compared to the process of making a decision of the human mind. One disadvantage of Neural Network is the slow training times, and it involves an outsized quantity of dataset to be correct [6]. The training and tuning of ANN model are the same with the previous models, hence the parameters set were different from the other models. The resampling outcome across tuning parameters are size, decay, accuracy and kappa.

*v. Model training and algorithm tuning using NB*

Based on Bayes Theorem, Naïve Bayes which has the independence prediction between predictors. A Naïve Bayesian model is uncomplicated to make, and its repetitive parameter predictions make it efficient for huge training sets [9]. The validation, training and tuning of the Naïve-Bayes model is like the previous models, however preprocessing methods like PCA, centering and scaling is not required on this model which may result in confusion to the system.

*c. Model selection*

After establishing and validating each predictive model, the comparison of each model will undergo by means of CV Accuracy and CM Accuracy that will be discussed in the next chapter.

## III. RESULTS AND DISCUSSION

The outcomes of the different models used, and methods done as well as the discussion and interpretation are shown in this chapter.

| | MQ3 | MQ6 | MQ135 | MQ136 | MQ137 | MQ138 | TGS821 | TGS822 | Meat | Bacteria |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Beef | None |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Beef | None |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Beef | None |
| 4 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | Beef | None |
| 5 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | Beef | None |
| 635 | 19 | 21 | 16 | 7 | 16 | 20 | 1 | 14 | Beef | E.Coli |
| 636 | 19 | 21 | 16 | 7 | 16 | 21 | 1 | 14 | Beef | E.Coli |
| 637 | 19 | 21 | 16 | 7 | 17 | 21 | 1 | 15 | Beef | E.Coli |
| 638 | 19 | 21 | 16 | 7 | 17 | 21 | 1 | 15 | Beef | E.Coli |
| 639 | 19 | 22 | 16 | 7 | 17 | 21 | 1 | 15 | Beef | E.Coli |
| 640 | 19 | 22 | 16 | 7 | 17 | 21 | 1 | 15 | Beef | E.Coli |
| 1115 | 12 | 19 | 8 | 3 | 6 | 11 | 3 | 6 | Beef | S.Aureus |
| 1116 | 12 | 19 | 8 | 2 | 6 | 10 | 4 | 6 | Beef | S.Aureus |
| 1117 | 12 | 19 | 8 | 2 | 6 | 10 | 4 | 6 | Beef | S.Aureus |
| 1118 | 13 | 19 | 8 | 2 | 5 | 10 | 5 | 6 | Beef | S.Aureus |
| 1119 | 13 | 19 | 8 | 2 | 5 | 10 | 5 | 6 | Beef | S.Aureus |
| 1120 | 13 | 19 | 8 | 2 | 5 | 10 | 5 | 6 | Beef | S.Aureus |

**Fig.7** Structure of dataset used in the study

Figure 7 shows the structure of the dataset used in the study. A total of 1920 observations with 10 variables is obtained in the given dataset. Nine of this are used as predictor variables which are 8 sensors and meat while the last column is used as class variable which is the bacteria.

*a. Results of Predictive Models*

The results in validation, tuning and training of the five predictive models are shown in this section.

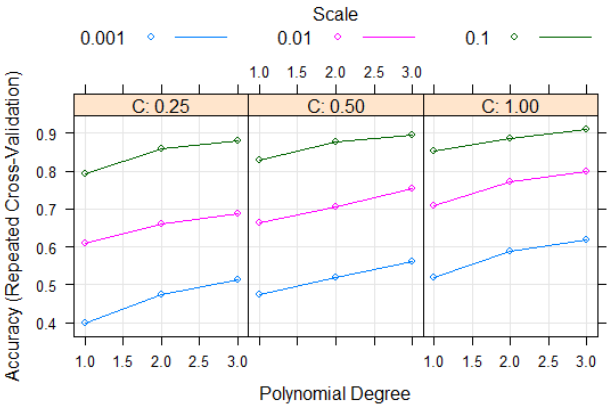


**Fig.8** Tuning graph of SVM model

SVM model gives the best tune at degree = 3, scale = 0.1 and cost = 1. Those parameters result to 0.90952 or approximately 90.95% accuracy in training. Table 1 presents the confusion matrix of the SVM model.

**Table 1** Confusion Matrix of SVM model

| | | Predicted Class | | | |
|---|---|---|---|---|---|
| | | E.Coli | Ecoli_Staph | None | S.Aureus |
| Actual Class | E.Coli | 124 | 1 | 0 | 0 |
| | Ecoli_Staph | 0 | 138 | 1 | 1 |
| | None | 19 | 5 | 143 | 19 |
| | S.Aureus | 1 | 0 | 0 | 124 |

From the confusion matrix shown in table 1, the model is somehow accurate which obtained 0.9095 accuracy. The overall statistics and the summary of statistics by class is shown in table 2 and 3.

**Table 2** Overall Statistics of SVM model

| | |
|---|---|
| Accuracy | 0.9184 |
| 95% Confidence Interval | (0.893, 0.9394) |
| No Information Rate | 0.25 |
| P-Value [Acc > NIR] | < 2.2e-16 |
| Kappa | 0.8912 |
| Mcnemar's Test P-Value | 8.607e - 08 |

The overall statistics shows that the model's performance is good in classifying data outside the training data. To know if the model's accuracy has a significant difference, the p-value is check. Because the p-value is small, the accuracy of the model doesn't have significant difference whether if it came from a majority of class or not.

**Table 3** Summary of Statistics by Class of SVM model

| | E.Coli | Ecoli_Staph | None | S.Aureus |
|---|---|---|---|---|
| Sensitivity | 0.8611 | 0.9583 | 0.9931 | 0.8611 |
| Specificity | 0.9977 | 0.9954 | 0.9005 | 0.9977 |
| Pos. Pred. Value | 0.992 | 0.9857 | 0.7688 | 0.992 |
| Neg. Pred. Value | 0.9557 | 0.9862 | 0.9974 | 0.9557 |
| Prevalence | 0.25 | 0.25 | 0.25 | 0.25 |
| Detection Rate | 0.2153 | 0.2396 | 0.2483 | 0.2153 |
| Detection Prevalence | 0.217 | 0.2431 | 0.3229 | 0.217 |
| Balance Accuracy | 0.9294 | 0.9769 | 0.9468 | 0.9294 |

Table 3 shows the summary of the statistics per class of SVM model. The tests conducted are the basic tests for evaluating the statistics of a predictive model. The output obtained indicates that the model used is somehow accurate.

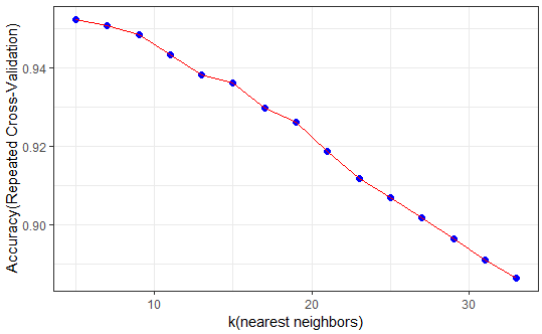


**Fig.9** Tuning graph of KNN model

Algorithm tuning in the KNN model yields the best tune at k = 5 and kappa = 96.65%. Those parameters result to 0.95236 or 95.24% accuracy in training.

**Table 4** Confusion Matrix of KNN model

| | | Predicted Class | | | |
|---|---|---|---|---|---|
| | | E.Coli | Ecoli_Staph | None | S.Aureus |
| Actual Class | E.Coli | 130 | 1 | 4 | 0 |
| | Ecoli_Staph | 2 | 140 | 0 | 0 |
| | None | 10 | 3 | 137 | 4 |
| | S.Aureus | 2 | 0 | 3 | 140 |

From the confusion matrix of KNN model, it can be seen model is accurate with few errors upon validation resulting in 0.9497 or approximately 94.97% CM accuracy.

**Table 5** Overall Statistics of KNN model

| | |
|---|---|
| Accuracy | 0.9497 |
| 95% Confidence Interval | (0.9285, 0.966) |
| No Information Rate | 0.25 |
| P-Value [Acc > NIR] | < 2.2e-16 |
| Kappa | 0.9329 |
| Mcnemar's Test P-Value | NA |

The overall statistics indicates that the model's performance is good in classifying data outside the training data. The p-value is the same as the previous model meaning the accuracy doesn't have a significance difference.

**Table 6** Summary of Statistics of KNN model

| | E.Coli | Ecoli_Staph | None | S.Aureus |
|---|---|---|---|---|
| Sensitivity | 0.9028 | 0.9722 | 0.9514 | 0.9722 |
| Specificity | 0.9884 | 0.9954 | 0.9606 | 0.9884 |
| Pos. Pred. Value | 0.963 | 0.9859 | 0.8896 | 0.9655 |
| Neg. Pred. Value | 0.9683 | 0.9908 | 0.9834 | 0.9907 |
| Prevalence | 0.25 | 0.25 | 0.25 | 0.25 |
| Detection Rate | 0.2257 | 0.2431 | 0.2378 | 0.2431 |
| Detection Prevalence | 0.2344 | 0.2465 | 0.2674 | 0.2517 |
| Balance Accuracy | 0.9456 | 0.9838 | 0.956 | 0.9803 |

The model has a very good performance which yields to 0.9497 accuracy in confusion matrix and 0.9524 or 95.24% in CV accuracy. Compared to the previous models, KNN is very good in classifying data outside the training data.

**Fig.10** Tuning graph of RF model

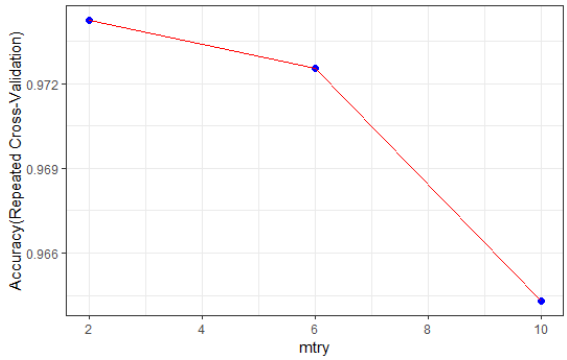


Fig.10 shows the tuning graph of RF model. Mtry = 2 was the final value used for the model which yields to the highest accuracy of 0.9746 or approximately 97.46% CV accuracy. This also shows that the model was excellent in classifying data outside the training data.

**Table 7** Confusion Matrix of RF model

| | | Predicted Class | | | |
|---|---|---|---|---|---|
| | | E.Coli | Ecoli_Staph | None | S.Aureus |
| Actual Class | E.Coli | 141 | 2 | 0 | 0 |
| | Ecoli_Staph | 0 | 138 | 0 | 1 |
| | None | 1 | 3 | 144 | 4 |
| | S.Aureus | 2 | 1 | 0 | 139 |

The confusion matrix of RF model signifies a very accurate model for classifying data even using test dataset outside with fewer errors upon validation compared to the previous models. This model able to reach 97.57% accuracy in confusion matrix.

**Table 8** Overall Statistics of RF model

| | |
|---|---|
| Accuracy | 0.9757 |
| 95% Confidence Interval | (0.9596, 0.9866) |
| No Information Rate | 0.25 |
| P-Value [Acc > NIR] | < 2.2e-16 |
| Kappa | 0.9676 |
| Mcnemar's Test P-Value | 6.20e - 02 |

From the overall statistics of RF model, it shows that the model obtained a CM accuracy of 0.9757. Compared to the previous models, RF model result to better and higher accuracy.

**Table 9** Summary of Statistics by Class of RF model

| | E.Coli | Ecoli_Staph | None | S.Aureus |
|---|---|---|---|---|
| Sensitivity | 0.9792 | 0.9583 | 1 | 0.9653 |
| Specificity | 0.9954 | 0.9977 | 0.9815 | 0.9931 |
| Pos. Pred. Value | 0.986 | 0.9928 | 0.9474 | 0.9789 |
| Neg. Pred. Value | 0.9931 | 0.9863 | 1 | 0.9885 |
| Prevalence | 0.25 | 0.25 | 0.25 | 0.25 |
| Detection Rate | 0.2448 | 0.2396 | 0.25 | 0.2413 |
| Detection Prevalence | 0.2483 | 0.2413 | 0.2639 | 0.2465 |
| Balance Accuracy | 0.9873 | 0.987 | 0.9907 | 0.9792 |

The summary of statistics by class of RF model is shown in Table 9 that implies higher chances of classifying data accurately. When compared to KNN model, it is easy to determine that the RF model produces superior results with 97.46% CV accuracy and 97.57% CM accuracy.

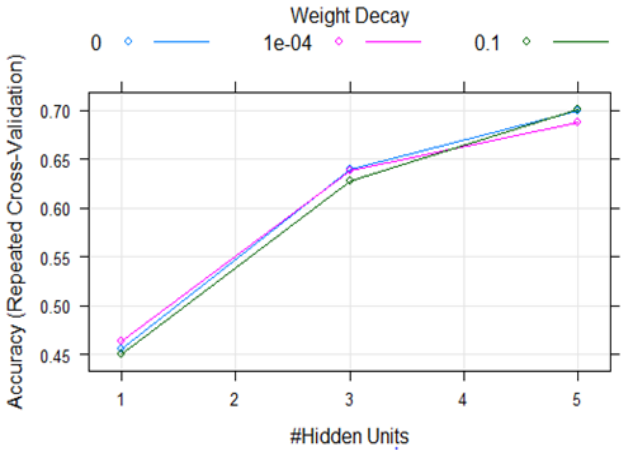


**Fig.11** Tuning graph of ANN model

The tuning graph of ANN model shows the cross-validation accuracy using the training dataset. Size = 5 and decay = 0.1 are the final values used in the model giving an accuracy of only 0.7007 in CV.

**Table 10** Confusion Matrix of ANN model

| | | Predicted Class | | | |
|---|---|---|---|---|---|
| | | E.Coli | Ecoli_Staph | None | S.Aureus |
| Actual Class | E.Coli | 126 | 98 | 4 | 6 |
| | Ecoli_Staph | 2 | 28 | 0 | 11 |
| | None | 11 | 9 | 108 | 35 |
| | S.Aureus | 30 | 9 | 32 | 92 |

**Table 11** Overall Statistics of ANN model

| | |
|---|---|
| Accuracy | 0.6146 |
| 95% Confidence Interval | (0.5735, 0.6545) |
| No Information Rate | 0.25 |
| P-Value [Acc > NIR] | < 2.2e-16 |
| Kappa | 0.4861 |
| Mcnemar's Test P-Value | < 2.2e-16 |

Confusion Matrix and Overall Statistics of ANN model in Table 10 and table 11 simply shows that the model used doesn't got a very good CM accuracy using the test dataset outside for which it only obtained 0.6146 accuracy.

**Table 12** Summary of Statistics by Class of ANN model

| | E.Coli | Ecoli_Staph | None | S.Aureus |
|---|---|---|---|---|
| Sensitivity | 0.875 | 0.19444 | 0.75 | 0.6389 |
| Specificity | 0.75 | 0.96991 | 0.8727 | 0.8935 |
| Pos. Pred. Value | 0.5385 | 0.68293 | 0.6626 | 0.6667 |
| Neg. Pred. Value | 0.9474 | 0.78318 | 0.9128 | 0.8813 |
| Prevalence | 0.25 | 0.25 | 0.25 | 0.25 |
| Detection Rate | 0.2188 | 0.04861 | 0.1875 | 0.1597 |
| Detection Prevalence | 0.4062 | 0.07118 | 0.283 | 0.2396 |
| Balance Accuracy | 0.8125 | 0.58218 | 0.8113 | 0.7662 |

Table 12 shows ANN's statistics by class which implies poor performance in classifying data outside the testing data.

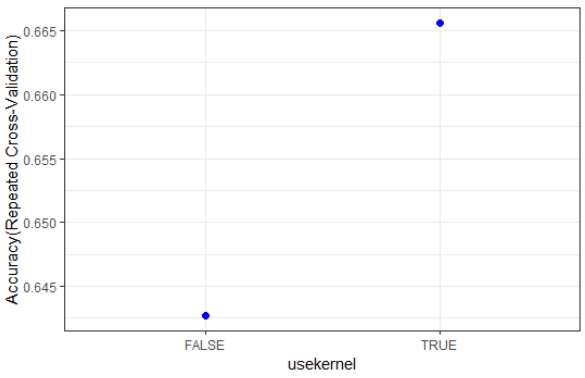


**Fig.12** Tuning graph of NB model

Algorithm tuning in the NB model produces a tuning parameter fL = 0, adjust = 1, and usekernel = TRUE. Those parameters yield only to 0.6697 CV accuracy. This implies that the model has a poor chance of classifying data.

**Table 13** Confusion Matrix of NB model

| | | Predicted Class | | | |
|---|---|---|---|---|---|
| | | E.Coli | Ecoli_Staph | None | S.Aureus |
| Actual Class | E.Coli | 112 | 52 | 8 | 19 |
| | Ecoli_Staph | 1 | 65 | 0 | 10 |
| | None | 30 | 17 | 134 | 41 |
| | S.Aureus | 1 | 10 | 2 | 74 |

The confusion matrix of the NB model implies that this model got the lowest accuracy upon validation compared to the previous models. This yields to higher chances of error in classifying data.

**Table 14** Overall Statistics of NB model

| | |
|---|---|
| Accuracy | 0.6684 |
| 95% Confidence Interval | (0.6283, 0.7068) |
| No Information Rate | 0.25 |
| P-Value [Acc > NIR] | < 2.2e-16 |
| Kappa | 0.5579 |
| Mcnemar's Test P-Value | < 2.2e-16 |

Table 14 shows the overall statistics of the NB model with only 0.6684 accuracy in confusion matrix. This means that the model is not good in data classification.

**Table 15** Summary of Statistics by Class of NB model

| | E.Coli | Ecoli_Staph | None | S.Aureus |
|---|---|---|---|---|
| Sensitivity | 0.7778 | 0.4514 | 0.9306 | 0.5139 |
| Specificity | 0.8171 | 0.9745 | 0.7963 | 0.9699 |
| Pos. Pred. Value | 0.5864 | 0.8553 | 0.6036 | 0.8506 |
| Neg. Pred. Value | 0.9169 | 0.842 | 0.9718 | 0.8569 |
| Prevalence | 0.25 | 0.25 | 0.25 | 0.25 |
| Detection Rate | 0.1944 | 0.1128 | 0.2326 | 0.1285 |
| Detection Prevalence | 0.3316 | 0.1319 | 0.3854 | 0.151 |
| Balance Accuracy | 0.7975 | 0.713 | 0.8634 | 0.7419 |

The summary of statistics by class of the NB model is shown in Table 15, indicates that the model is not good in classifying data since it only got 0.7007 in CV accuracy and 0.6146 accuracy in CM accuracy. The comparison of all the predictive models developed in this study is shown in Table 16. These models are compared with their respective cross-validation accuracy and confusion matrix accuracy.

**Table 16** Comparison of the different predictive models

| Model | CV Accuracy | CM Accuracy |
|---|---|---|
| Random Forest | 0.9746 | 0.9757 |
| K-Nearest Neighbors | 0.9524 | 0.9497 |
| Support Vector Machine | 0.9095 | 0.9184 |
| Artificial Neural Network | 0.7007 | 0.6146 |
| Naïve-Bayes Classifier | 0.6697 | 0.6684 |

Cross-validation accuracy states if the model performs well using own training dataset while confusion matrix accuracy gives the model performance when training using the dataset outside. Of all the models, the Random Forest predictive model reached the highest accuracy in CV accuracy and CM accuracy. While the KNN model being the second to the highest, it is also possible to use when classifying data. In the application of this study, Random Forest model is chosen for its better performance against overfitting and produces superior results.

## IV. CONCLUSION

The five predictive models are successfully implemented which arrived at a conclusion that the Random Forest predictive model is the leading model that obtained an excellent performance in cross-validation and confusion matrix accuracy with 97.46% and 97.57% respectively. Among the five models used, Random Forest predictive model is the most suitable model for detecting and classifying bacteria thus, making the model dependable for upcoming implementations.

## ACKNOWLEDGEMENT

The authors would like to acknowledge National Meat Inspection Service (NMIS) Quezon City for their cooperation in doing this project and Paco Market Manila vendors for allowing the proponents to deploy the project in the venue.

## REFERENCES

[1] Switaj, T., Winter, K. and Christensen, S. (2019). *Diagnosis and Management of Foodborne Illness*. [online] Aafp.org. Available at: https://www.aafp.org/afp/2015/0901/p358.html?fbclid=IwAR1jlr5gnLlWWLgFji3k-V9g1jaiPBWKvnlrXRB0oUi23FcPUnn2R6qvoRI [Accessed 15 Apr. 2018].

[2] Holley, R., Palanichamy, A. and Jayas, D. (2006). *ASABE Technical Library :: Abstract. Review of Microbial Modeling Techniques for Meat Industry* [online] Elibrary.asabe.org. Available at: http://elibrary.asabe.org/abstract.asp?fbclid=IwAR2HgNBYwrbXFTYMq7dmDdmmWnffSzZ6kWfJLEvKWaWLkhNvIYuDNTChRUk [Accessed 2 Feb. 2019].

[3] Malley, B., Ramazzotti, D. and Tzung-yu Wu, J. (2016). *Secondary Analysis of Electronic Health Records*. Springer, Cham, pp.115-141.

[4] Tariq, O. (n.d.). *Model Tuning | DataRobot Artificial Intelligence Wiki*. [online] DataRobot. Available at: https://www.datarobot.com/wiki/tuning/ [Accessed 5 Feb. 2019].

[5] Bambrick, N. (2016). *Support Vector Machines: A Simple Explanation*. [online] Kdnuggets.com. Available at: https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html [Accessed 14 Apr. 2018].

[6] Amado, T. (2017). A Smartphone Indoor Localization Algorithm Based on WLAN Location Fingerprinting with Feature Extraction and Clustering. *Sensors*, 17(6), p.1339.

[7] SRIVASTAVA, T. (2014). *Introduction to KNN, K-Nearest Neighbors : Simplified*. [online] Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/ [Accessed 14 Apr. 2018].

[8] Polamuri, S. (2017). *How the random forest algorithm works in machine learning*. [online] Dataaspirant. Available at: http://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learing/ [Accessed 14 Apr. 2018].

[9] Sayad, S. (n.d.). *Naive Bayesian*. [online] Saedsayad.com. Available at: https://www.saedsayad.com/naive_bayesian.htm [Accessed 14 Apr. 2018].