Future Computing and Informatics Journal

Volume 7
Issue 2 (2022) *Volume 7 Issue 2*

Article 4

2022

Key Performance Indicators Detection Based Data Mining

Fatma Abogabal
Faculty of Commerce and Business Administration, Future University in Egypt, fatma.abogabal@fue.edu.eg

shimaa mohamed ouf
BIS helwan university, shimaaouf@yahoo.com

Amira M. Idrees AMI
Faculty of Computers and Information Technology, Future University in Egypt, amira.mohamed@fue.edu.eg

Future Computing and Informatics Journal

Homepage: https://digitalcommons.aaru.edu.jo/fcij/

doi: http://Doi.org/10.54623/fue.fcij.7.2.4

Proposed framework for applying data mining techniques to detect key performance indicators for food deterioration

Fatma Abogabal [1, a] , Shimaa M. Ouf [2, b], Amira M. Idrees [3, c]

[1] Faculty of Commerce and Business Administration, Future University in Egypt, Egypt
[2] Faculty of Commerce and Business Administration, Helwan University, Egypt
[3] Faculty of Computers and Information Technology, Future University in Egypt, Egypt

[a] fatma.Abogabal@fue.edu.eg, [b] shimaaouf@yahoo.com,
[c] amira.mohamed@fue.edu.eg

ABSTRACT

One of the most prosperous domains that Data mining accomplished a great progress is Food Security and safety. Some of Data mining techniques studies applied several machine learning algorithms to enhance and traceability of food supply chain safety procedures and some of them applying machine learning methodologies with several feature selection methods for detecting and predicting the most significant key performance indicators affect food safety. In this research we proposed an adaptive data mining model applying nine machine learning algorithms (Naive Bayes, Bayes Net Key -Nearest Neighbor (KNN), Multilayer Perceptron (MLP), Random Forest (RF), Support Vector Machine (SVM), J48, Hoeffding tree, Logistic Model Tree) with feature selection wrapper methods (forward and backward techniques) for detecting food deterioration's key performance indicators. Therefore, results before and after applying wrapper feature selection methods have been compared, analyzed, and interpreted. In conclusion the proposed model applied effectively and successfully detected the most significant indicators for meat safety and quality with the aim of helping farmers and suppliers for being sure of delivering safety meat for consumer and diminishing the cost of monitoring meat safety.

Keywords: Data Mining, Machine Learning Techniques, Key Performance Indicators (KPI), Feature Selection Methods, Food Industry, Food Security, Food Hazards, Meat Safety and Quality, Food Deterioration.

Published by Arab Journals Platform, 2022

1. INTRODUCTION

Food safety and security is a significant factor in public health precedence to be considered because of its susceptibility to many hazards that could contaminate food products at any stage on its supply chain such as harvesting, processing, transporting, preparing, storing, and serving food. Food 'poisonings', causing death, raise alarm not only about the food served in restaurants and fast-food outlets also the food bought in supermarkets. Consumers, industries, as well as governments are together taking food safety as a serious issue. Food poisoning or any hazard issue like this may lead to a very fatal consequence for human health, hence it is essential to give more attention to food supply chain safety.

Food hazards can be categorized as chemical hazards which can be classified into two categories: chemical agents and toxic metals. Chemicals, both natural and manufactured, may have an adverse effect on people's health and cause illness. i.e. (cleaning or sanitizing agents or allergens and copper zinc) used in galvanized containers, physical hazards are those hazards which are not supposed to, nor likely to be in the food like wood, glass, bones grit or dust. Meanwhile physical and chemical hazards can induce foodborne sickness, biological hazards such as germs, bacteria, viruses, and parasites cause the bulk of these illnesses [1].

Data mining, machine learning algorithms and feature selection techniques have proven to be effective in a several fields, such as biology, finance, and marketing. Nevertheless, applying of data mining techniques in Food supply chain networks (FSNs) for quality monitoring and control has lagged that of other business (areas). One of the causes is that, historically, FSNs were less automated than other firms on a supply chain level. Nevertheless, in recent years, the food sector has started to develop information systems (IS) to collect data on various stages of FSNs. These IS now giving us the ability to utilize operational data and data mining technologies to explore the significant relationships for food quality issues [2].

High-dimensional data analysis is a challenge for researchers and engineers in the fields of machine learning and data mining. Feature selection provides an effective way to solve this problem by removing irrelevant and redundant data that do not contribute to the accuracy of a data mining model or may in fact decrease the accuracy of the model, as well as feature selection can reduce computation time, improve learning accuracy, and facilitate a better understanding for the learning model or data. Feature selection is also called feature-engineering process, variable selection, or attribute selection. Feature selection is the process of obtaining or selecting the most significant and smallest subset from an original feature set according to certain feature selection criterion, which selects the relevant features of the dataset on other meaning it is the process of selecting the most relevant features or the most important features for apply them in model construction [3].

https://digitalcommons.aaru.edu.jo/fcij/vol7/iss2/4
DOI: https://doi.org/10.54623/fue.fcij.7.2.4

As we defined Feature selection process. It plays an important role in compressing the data processing scale where the redundant and irrelevant features are removed. Redundant features duplicate or all the information contained in one or more other attributes. Irrelevant features contain almost no useful information for the data mining model construction at hand. Redundant and irrelevant features or attributes can reduce the accuracy of data mining model or do not contribute to the data-mining model. The ideal approach to feature selection is to try all possible subsets of features as input to the data mining Model and then take the subset that produces the best performance results [4].

The residual research will figure out the objectives of this paper as follow: section 2. Shows the related work of applying the concept of data mining with machine learning algorithms and feature selection techniques for enhancing food safety and security domain. Section 3. Illustrates the proposed framework. Section 4. Explains materials and tools for applying the proposed framework. Section5. Discuss the experimental structure for the research. Section6. Interprets the findings and results. Finally, Section 7. Presents the conclusion and future work.

2. RELATED WORK

This section presents the latest and most related previous articles that were conducted to clarify data mining algorithms with feature selection techniques and their accuracy in addition to the feature selection methods that have been applied for enhancing the safety of miscellaneous species of food supply chain in its various procedures. Finally, the articles are assessed and appraised based on of three factors, including (the year of publication, proposed methods for applying data mining techniques with their accuracy for enhancing food security domains and the proposed frameworks that combine data mining techniques with several feature selection methods with their accuracy to discover the most relevant and crucial key performance indicators (KPI) for controlling the food safety process in any stage food supply chain.

Findings and previous works will be classified as follow:

According to Nehaya, Khedr, Idrees, & Kholeif (2017) the key performance indicators (KPIs) are a vital aspect of business knowledge frameworks since the decisions made by KPIs are fundamental to development. In this research, they proposed a structure that would serve as an identification framework to detect KPIs from historical organizational data using data mining techniques, and to break down the relevance between factors that will influence the rendering to help organizations accomplish their business strategy, assisting decision makers to dictate the most convenient KPIs for their business objectives, without the need for the related knowledge in data mining. However, while their detection framework was only intended to be used with a certain set of qualities. They consider this and will concentrate on using this framework with big data to recognize KPIs dealing with more attributes in their future work to illuminate more important issues such as detecting anomalies and anticipating the deviation in KPIs. However, their detection framework has limitations in that it was applied specifically on a set number of attributes.

Published by Arab Journals Platform, 2022

According to Ilic, Ilic, Jovic, & Panic (2018) they applied a variety of data mining algorithms in this study to anticipate the risk of cherry fruit infection. By predicting when the environmental circumstances are suitable for the growth of fruit infection, one can predict when contamination incidents will occur. In this manner, chemical protection becomes more and more effective over time, saving farmers money while, more importantly, producing better food due to a reduction in chemical treatments. They are most likely going to make arrangements to upgrade the equipment so that it can automatically gather environmental data from the network of meteorological stations. Later, a programmed farmer alerting of potential contamination through the portable or web organization is another update.

In this article Based on the needy and unstable food supply in Egypt. The research developed a framework for estimating the appropriate production and import levels to meet the demands of necessary yields for Egyptian citizens concerned about population expansion through the use of data mining techniques. It illustrates the difference between the amount of food that is supplied and that is needed, as well as the abundance and lack of food that need be secured to end the circumstance of food insecurity. They aim to generalize and apply the proposed framework to various harvests, including those of animal origin, in their upcoming work (Milk, Eggs and Meat). Pioneers linked to food security locations and might greatly benefit from reports based on this concept [5].

According to Abdul Ghafoor & Sitkowska (2021) mastitis is one of the most costly and risky diseases for dairy cattle and prevalent among milk cattle due primarily to environmental infections. There are several preventative and curative measures that can be taken, but many of them can be prohibitively expensive, especially for smaller farms. To overcome this critical issue, supervised machine learning approaches were implemented to identify the most useful parameters that could be used to predict the risk of mastitis in cattle. To accomplish this, they presented a web application powered by 26 machine learning models for determining the likelihood of mastitis in calves based on the udder's inhale and exhale limits and temperature. The authors of this study sought to integrate these emerging trends in data science to derive a solution for predicting the risk of mastitis in cattle in advance, with the hopes of lowering the high cost of treatment, convincing farmers to forgo antibiotics as a preventative measure, and cutting down on unnecessary veterinary costs by making an open-source tool available online for free.

In this study A proposed framework for detecting the most significant indicators quality food status, determining the suitability of the current conditions compared with the required conditions, and alerting users of near-threshold conditions was introduced. The framework predicts the available parameters for maintaining the food's acceptability and includes a plan to follow. Finally, this research is a step which needs further investigation which the authors target to apply on a real dataset with more advanced targets including exploring relations between KPIs and apply the proposed framework in different critical domains [6].

This research introduced a novel technique for minimizing data dimensionality. The introduced technique is constructed based on two primary factors, the first one is using the modified Saaty method to determine the consistency of the qualities while also suggesting additional modifications aimed at measuring accuracy more precisely. The second factor involves using clustering algorithms on the same attributes to remove the attributes that have the lowest weighted and the least be consistent measurements from each

https://digitalcommons.aaru.edu.jo/fcij/vol7/iss2/4
DOI: https://doi.org/10.54623/fue.fcij.7.2.4

cluster. By integrating the two stages, the most important dataset attributes are highlighted. Finally, the proposed technique has been successfully applied on the Gastrology dataset which attributes have been reduced from 62 attributes to 31 attributes and the selected subset of 31 attributes was applied for a set of classification models and successfully enhance classification models accuracy [7].

3. PROPOSED FRAMEWORK

In this section we will explain the proposed adaptive Framework that will contribute to determine the most critical attributes which may lead to food deterioration. Therefore, these attributes can be taken into consideration as a crucial key indicator for the detection of probable food spoilage incidents (detection process of the key performance indicators for food safety quality).

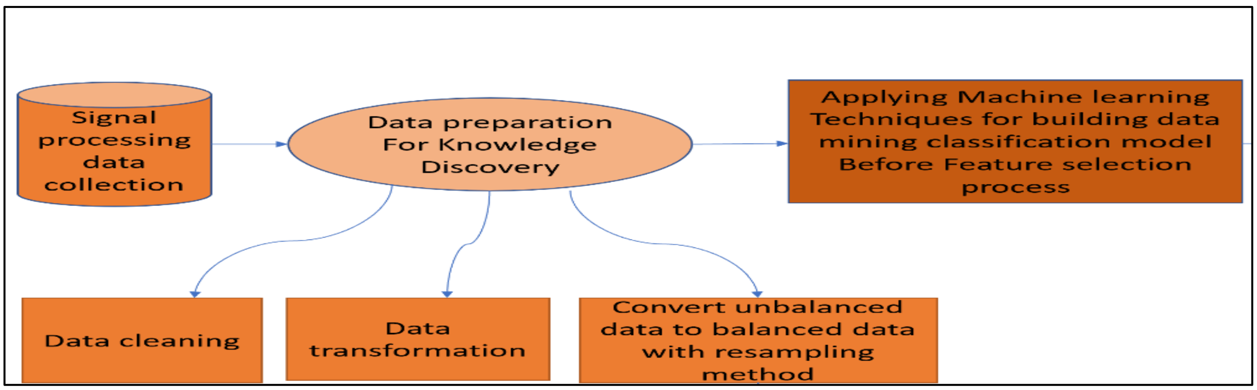

Fig 1. Proposed Framework

4. Materials and tools

In this section we will discuss the data collection, dataset description and Experimental Structure for E-nose meat signal processing dataset as follow:

4.1 Data Collection

The data are collected via electronic nose sensors for meat samples in storage unit at uncontrolled ambient environmental circumstances, as well as the variations in ambient environmental conditions tracked over 36 hours applying time series data. The E-nose sensors applied MQ gas sensors, MQ temperature sensors and MQ humidity sensors for measuring the fluctuations on the environmental factors for meat samples during the storage process to monitor meat quality and safety.

4.2 Data Set Description

There are 4553 records in the unbalanced e-nose meat dataset under uncontrolled ambient environmental factors in storage unit. The Data set consists of 14 attributes (13 MQ sensors and TVC value) and class label. **As shown in table 1:**

Table 1: List of E-nose meat data set attributes

Attribute serial	Attribute Name (MQ Sensor Name)	Attribute Description
1	Minute (time of capturing the status of any feature)	Time series for monitoring the environmental factors for stored meat during the experiment
2	TVC	Total Viable count (size of bacteria in meat samples) It means the percent of microbial population in meat samples
3	MQ135 sensors	High sensitivity to measure Ammonia, sulfide, and Benzes steam
4	MQ136 sensors	High sensitivity to measure Hydrogen Sulfide
5	MQ2 sensors	High sensitivity to measure (H_2, LBG, CH_4, CO, Alcohol, propane)
6	MQ3 sensors	High sensitivity to Alcohol and small sensitivity to Benzine
7	MQ4 sensors	High sensitivity to (CH_4, Natural Gases) and small sensitivity to Alcohol
8	MQ5 sensors	High sensitivity to (H_2, LBG, CH_4, CO, Alcohol
9	MQ6 sensors	High sensitivity to (LBG, Iso-butane, propane) and small sensitivity to alcohol
10	MQ7 sensors	High sensitivity to Natural gases
11	MQ8 sensors	High sensitivity to H_2 -Hydrogen
12	MQ9 sensors	High sensitivity to Methane, Propane and CO
13	Humidity (DHT22) sensors	High sensitivity to moister in the storage unit
14	Temperature (DHT22) sensors	High sensitivity to temperature in the storage unit

https://digitalcommons.aaru.edu.jo/fcij/vol7/iss2/4
DOI: https://doi.org/10.54623/fue.fcij.7.2.4

Data source: [8]

The Class Label consist of discrete values for beef quality [“excellent”, “good”, “acceptable”, “spoiled”] class label scattered as follow:

- Excellent: 6.25%
- Good: 4.63%
- Acceptable: 4.01 %
- Spoiled: 85.08

The data in the E-nose meat dataset, which was generated from a literature review, is unbalanced, meaning that one class label has a high number of observations while the other has a tiny percentage. The class spoiling comprises 85.08% of the total sample size in the E-nose meat dataset, which is a skewed proportion and an inaccurate classification. Result. How effectively can we truly predict both majority and minority classes when dealing with an unbalanced dataset? Let's use our e-nose meat data set for meat Quality as an example and assume that there are just two groups (Good, Excellent). Assume that out of 4553 entries in an existing dataset, only 5 samples were determined to be excellent samples. This dataset will be used to predict the grade of meat quality. Therefore, the majority class has a spoiling rate of 95%, whereas the minority class only has a decent sample grade of 5%. Let's say that 4553 out of 4553 meat samples are all predicted by our model to be spoilage. Consequently, it is essential to accurately determine the minority classes in rare circumstances like detection or prediction. Therefore, the model should not be skewed to merely identify the majority class but must also give the minority class similar weight or significance. An algorithm trained on the same data will be biased toward the same class if the dataset is biased toward that class.

4.3 Data preparation

Most of Data sets are generated with missing values, inaccuracies or other blunders and separate data sets often have different formats that need to be harmonized when they are combined. Correcting data errors, validating data quality, and consolidating data sets are being a big part of data preparation process. In this research to prepare the data set to ensure that the data being readied and ready for applying machine learning algorithms for better accurate results we applied 3 steps sequentially as follow:

4.3.1 Data Balancing

In this research the dataset was imbalanced dataset where the target class has an unbalanced distribution of observations. Therefore, there was a need for converting the imbalanced dataset to balanced dataset to get more accurate results for machine learning models for classification.

4.3.2 Data cleaning

The dataset of this research has been cleansed by filling missing values, erasing the noisy data or irrelevant data, resolving the inconsistency, and removing outliers

4.3.3 Data Transformation

After cleaning and validating data, the next step is to ensure that the data is correctly formatted or not. If data is

Published by Arab Journals Platform, 2022

formatted incorrectly, it will help build a high-quality model.

4.4 Experimental Structure for E-Nose meat Data set under uncontrolled environmental factor

All Experiments are implemented with classification models and feature selection techniques a classification technique is a Machine Learning systematic approach that applies a set of algorithms to build a classification model from the input datasets by learning a target function F that maps each attribute set x to one of the predefined classes labels**.** The target function is also known informally as a classification model [9]. Classification models determines how data set should be classified under a set of labels, classes, or categories. Each technique applies a learning algorithm to determine which best model that define the associations between the class label of the input data set. The classification Model generated by a learning algorithm must accurately anticipate the class labels of instances and the input data sets effectively [10].All experiments are implemented via weka tool, and all Data mining classification models applied with 9 machine learning algorithms (Naive Bayes, Bayes Net Key -Nearest Neighbor (KNN), Multilayer Perceptron (MLP), Random Forest (RF), Support Vector Machine (SVM), J48, Hoeffding tree, Logistic Model Tree). All Experiments for the 9 classification models are evaluated with performance matrix of 6 measurements as shown in fig. 2:

Performance Matrix for Classification Models	
S.N	**Evaluation measurement**
1	Accuracy
2	Error Rate
3	Precision
4	F-measure
5	Recall
6	Time to build model

Fig. 2. Performance Matrix for evaluating 9 classification models

4.4.1 Applying Resampling Technique for converting the imbalanced datasets to balanced datasets

To overcome the problem of imbalanced dataset for getting more accurate results for performance matrix. There are many methods for Face this critical issue. One of the most effective methods for converting imbalanced dataset to balanced dataset is Resampling techniques has two methods: the oversampling and under sampling techniques. Over sampling vs. under sampling we will discusses the difference between these 2 techniques and when to use each one.

https://digitalcommons.aaru.edu.jo/fcij/vol7/iss2/4
DOI: https://doi.org/10.54623/fue.fcij.7.2.4

Over sampling technique: When one class of dataset is the understated minority class in the data sample, over sampling techniques maybe applied to duplicate these results for more balanced number of positive results in training (on other meaning in the process of random oversampling, samples from the minority class are randomly duplicated and then added to the training dataset). One of the most common and effective techniques for oversampling is SMOTE (Synthetic Minority Over-sampling Technique), which generates synthetic samples by taking features from instances in the minority class at random. The oversampling technique is applied when the amount of data assembled is inadequate. Inversely under sampling technique means when a class of dataset is the overrepresented majority class, under sampling can be applied to balance it with the minority class (in other meaning in the process of random under sampling samples randomly selected from majority class are deleted from the training dataset. The under sampling is applied when the amount of data assembled is adequate. As mentioned below according to the number of records for E-nose meat dataset is not sufficient. In our Research we applied the oversampling technique for balancing our dataset to get more accurate results for the performance results for classification models [11].

4.4.2 Building the classification Model and interpreting the classification results with applying imbalanced E-nose meat data set and balanced dataset after applying Smote technique.

In the first experiment for building 9 machine learning classification models as we mentioned for the 9 applied classification models in section 4.3, all classification models are applied twice with imbalance and balanced dataset to analyze the effect of applying SMOTE over sampling technique. The e-nose meat dataset has been split with cross validation k5 technique for all classifiers. In the first iteration, the first fold is used to test the model and the rest are used to train the model. In the second iteration, 2nd fold is used as the testing set while the rest serve as the training set. This process is repeated until each fold of the 5 folds have been used as the testing set. All classification models with balanced and imbalanced dataset are evaluated with the performance matrix which is mentioned in section 4.3.

Published by Arab Journals Platform, 2022

Table 2. Results of classifiers measurements to build each of the selected 9 models with imbalanced E-nose meat dataset before applying oversampling (No. of attributes = 14

Classifier Name	Accuracy	Error Rate	Precision	F-measure	Recall	Time To Build the Model
Naïve Bayes	98.46%	1.53%	98.7%	98.5%	98.5%	0.03 s
Bayes Net	99.75%	0.24%	99.8%	99.8%	99.8%	0.09 s
MLP	99.80%	0.19%	99.8%	99.8%	99.8%	3.84 m/s
SVM (SMO)	99.36%	0.63%	99.4%	99.4%	99.4%	0.08 s
KNN(IBK)	99.82%	0.17 %	99.8%	99.8%	99.8%	0 s
Hoeffding tree	98.50%	1.49 %	98.8%	98.6%	98.5%	0.06 s
J48	99.89%	0.10 %	99.9%	99.9%	99.9%	0.02 s
LMT	99.86%	0.15 %	99.8%	99.8%	99.8%	1.09 m/s
RF	99.93%	0.06 %	99.9%	99.9%	99.9%	0.24 s
Results of dataset classifiers measurements to build each of the selected 9 models after applying oversampling technique with balanced E-nose meat (No. of attributes = 14)						
Classifier Name	**Accuracy**	**Error Rate**	**Precision**	**F-measure**	**Recall**	**Time to build the model**
Naïve Bayes	97.7%	2.3	97.7%	97.7%	97.7%	0.01 s
Bayes Net	99.7%	0.3	99.8%	99.8%	99.8%	0.05 s
MLP	99.8%	0.2	99.8%	99.8%	99.8%	13.88 s
SVM(SMO)	99.0%	1.0	99.0%	99.0%	99.0%	0.05 s
KNN(IBK)	100%	0	100%	100%	100%	0 s
Hoeffding trees	97.7%	2.3	97.7%	97.7%	97.7%	0.08 s
J48	99.9%	0.1	99.9%	99.9%	99.9%	0.03 s
LMT	100%	0	100%	100%	100%	4.4 s
RF	100%	0	100%	100%	100%	0.78 s

https://digitalcommons.aaru.edu.jo/fcij/vol7/iss2/4
DOI: https://doi.org/10.54623/fue.fcij.7.2.4

After applying Smote algorithm to balance the data we found that results of classification are changed and getting more accurate results for classifying electronic nose meat samples and preparing the data set for feature selection process. After cleaning and validating data, the next step is to ensure that the data is correctly formatted or not. If data is formatted incorrectly, it will help build a high-quality model.

4.4.3 Applying feature selection wrapper techniques with the balanced E-nose meat dataset

For the E-Nose meat dataset the applied feature selection method for the nine machine learning classification models is wrapper method which consist of (Forward Selection technique and backward selection technique).

The wrapper techniques were applied in this study since they generated more accurate predictive results than filter Techniques. Based on a particular machine learning algorithm that we are intending to fit for a specific dataset, this feature selection method was developed. It follows a greedy search approach by evaluating all the possible combinations of features against the evaluation criterion. All best subsets that have been generated from applying forward and backward techniques separately generated based on the concept of threshold 50% which means the number of repeated each attribute per the nine classifiers. In our Research we applied the concept of union subset which means taking a combination of a forward technique subset and backward technique subset after applying each technique separately with the nine classifiers. This concept is applied to be sure that we applied the new subsets of both forward and backward techniques. All results of applying the two applied techniques for wrapper feature selection method will be compared, analyzed, and interpreted as follow:

After applying forward technique for the balanced dataset, we found that the best data set as follow:

Published by Arab Journals Platform, 2022

Table 3. Results of the best selected new subsets with building each of the selected 9 models after applying Wrapper Method (Forward Technique)
(No. of attributes = 14)

Classifier name	Best subsets
Naïve Bayes	**3 Attributes** **{1,8,11} -Merit of best subset (Accuracy): 99.5%**
Bayes Net	**2 Attributes** **{2,6} -Merit of best subset (Accuracy): 100%**
SVM (SMO)	**4 Attributes** **{2,3,6,12} -Merit of best subset (Accuracy): 99.3%**
IBK (KNN)	**2 Attributes** **{1,2} -Merit of best subset (Accuracy): 100%**
Hoeffiding	**1 Attributes** **{1} Merit of best subset (Accuracy): 99.9%**
LMT	**2 Attributes** **{2,10} -Merit of best subset (Accuracy): 100%**
J48	**13 Attributes** **{1,2,3,4,5,7,8,9,10,11,12,13,14} -Merit of best subset (Accuracy): 99.9%**
RF	**3 Attributes** **{2,9,14} -Merit of best subset (Accuracy): 100%**

Attribute serial	Attribute Name	No. of repeated times per subset/ applied classifiers (9 classifiers)	Percentage of repeated folds
1	**Minute (time of capturing the status of any feature)**	**5/9**	**55.5%**
2	**TVC**	**7/9**	**77.7%**
3	**MQ135**	**3/9**	**33.3%**
4	**MQ136**	**1/9**	**44.4%**
5	**MQ2**	**1/9**	**11.1%**
6	**MQ3**	**2/9**	**22.2%**
7	**MQ4**	**2/9**	**22.2%**
8	**MQ5**	**2/9**	**22.2%**
9	**MQ6**	**2/9**	**22.2%**
10	**MQ7**	**1/9**	**11.1%**
11	**MQ8**	**3/9**	**33.3%**
12	**MQ9**	**2/9**	**22.2%**
13	**Humidity (DHT22) sensors**	**1/9**	**11.1%**
14	**Temperature (DHT22) sensors**	**2/9**	**22.2%**

https://digitalcommons.aaru.edu.jo/fcij/vol7/iss2/4
DOI: https://doi.org/10.54623/fue.fcij.7.2.4

The no. of selected attribute in the Final Set with Forward technique applying Greedy -Stepwise method according to the highest percentage of repeated folds with threshold 50% is 2 attributes from the overall 14 attributes will be as follow:

-Set (A) is the final best subset of attributes selected by applying wrapper feature selection method -Forward technique Based on the high percentage of repeated folds with threshold 50%

Attribute serial	Attribute Name	Percentage of repeated folds
1	**Minute**	**55.5%**
2	**TVC**	**77.7%**

5. RESULTS AND FINDINGS

After interpreting the results we compared and analyzed and the results of applying data mining with machine learning the 9 classifiers before and after feature selection for E-nose meat dataset we found that: applying the new subset of attributes (C) with only 8 attributes {Minute , TVC, MQ135, MQ136, MQ3, MQ4, MQ5, MQ8 } and neglecting the other 7 attributes {MQ2, MQ6, MQ7, MQ9, Humidity, Temperature} Enhanced the results of building the datamining and machine learning models and the results are evaluated as follow :

The results of these 5 classifiers (Naïve Bayes, Bayes Net, MLP, Hoeffding and J48) have increasing changes in accuracy, precision, F-measure and recall values & decreasing changes in error rates and time to build model values after feature selection with wrapper-Greedy Stepwise- forward technique. The results of (LMT, RF) classifiers have no change in all measurements except a tiny change a decreasing in time to build the model measurement value which is a little enhancement. The results of classifier SVM (SMO) with the green color have decreasing changes with very tiny values in accuracy, precision, F-measure and recall values & increasing changes with very tiny values in error rates and time to build the model values after feature selection with wrapper-Greedy Stepwise-forward technique. The results of classifier KNN (IBK) with the green color have decreasing changes with very tiny values in accuracy, precision, F-measure and recall values & increasing changes with very tiny values in error rates and no change in time to build the model values after feature selection with wrapper-Greedy Stepwise- forward technique. Finally, the J48, LMT and RF classifiers effectively achieved the best results for the accuracy, error rates, precession, F-measure, Recall and time to build model. In conclusion Applying the concept of applying the integration concept of machine learning algorithms with feature selection wrapper methods successfully detected the most significant key indicators which can affect the meat quality and safety hence farmers and suppliers can depend on these only 2 attributes {minute ,TVC} and 6 sensors {MQ135, MQ136, MQ3, MQ4, MQ5, MQ8 } as key indicators for monitoring meat quality and safety status and minimizing the cost of monitoring sensors of meat quality and safety. Furthermore, the proposed framework successfully removed the attributes with no advantages for meat safety and security which can affect the performance matrix of data mining classification models as a noisy data.

6. CONCLUSION AND FUTURE WORK

In this paper the proposed framework of combining machine learning algorithms with wrapper feature selection techniques has been successfully detected the most significance indicators for monitoring meat safety and assisting food suppliers for diminishing and optimizing the cost of sensors array factors which are needed to trace the safety of food products. The proposed framework of data mining is applied on electronic nose meat data sets stored on chambers and monitored under various 14 factors. After applying this framework, the 14 attributes are optimized to 8 attributes based on two factors: applying wrapper feature selection forward and backward techniques with applying the concept of 50%threeshold to select the most significant combination of attributes for monitoring meat quality and the results of building the nine classification models after using the new best subset. For future work the applied framework can be applied to more food signal processing data sets with more feature selection techniques (hybrid methods) to compare the results with wrapper methods and selecting the best subsets for enhancing the results and more machine learning algorithms. It needs to be applied with big size of signal processing data sets with more features for enhancing the accuracy of classification models results.

REFERENCES

[1] "Summary of Food Hazards and Contamination," 2016. [Online]. Available: http://www.masslocalinstitute.org/.

[2] M. Bortolini, R. Accorsi, M. Gamberi and F. Pilati, "A model to enhance the penetration of the renewables to power multistage food supply chains," in *Sustainable Food Supply Chains*, 2019, pp. 305-315.

[3] M. Attia, M. Farghaly , M. Hamada and A. M. Idrees, "A Statistical-Mining Techniques' Collaboration for Minimizing," *Future Coomputing and informatics journal*, 2021.

[4] A. Idrees, A. A. Almazroi and A. E. Khedr, "Utilizing Mining Techniques for Attributes' Intra-Relationship Detection, a Collaborative Approach," 2021.

[5] A. E. Khedr, M. Kadry and G. Walidb, "Proposed framework for implementing data mining techniques to enhance decisions in agriculture sector," in *International Conference on Communication, Management and Information Technology (ICCMIT)*, 2015.

[6] F. Abogabal, S. M. Ouf, . A. M. . Idrees and . A. E. . Khedr, "AN ARCHITECTURAL FRAMEWORK FOR GENERATING FOOD SAFETY," *Computer and Information Science*, 2019.

[7] A. M. Idrees and W. H. Gomaa, "A Proposed Method for Minimizing Mining Tasks' Data Dimensionality," *International Journal of Intelligent Engineering and Systems,*, 2020.

[8] "Electronic nose dataset for beef quality monitoring under an uncontrolled environment," 2018. [Online].

https://digitalcommons.aaru.edu.jo/fcij/vol7/iss2/4
DOI: https://doi.org/10.54623/fue.fcij.7.2.4

[9] A. E. Khedr, A. M. Idrees and A. I. El Seddawy, "Enhancing Iterative Dichotomiser 3 algorithm for classification decision tree," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2016.

[10] A. E. Khedr, A. M. Idrees and A. I. El Seddawy, "Adaptive Classification Method Based on Data Decomposition," *Journal of Computer Science*, p. 2016.

[11] "over sampling and under sampling," 2018. [Online]. Available: https://www.techtarget.com/whatis/definition/over-sampling-and-under-sampling#:~:text=When%20one%20class%20of%20data,of%20data%20collected%20is%20insufficient..

[12] A. E. Khedr, M. kadry and G. Walid, "Proposed framework for implementing data mining techniques to enhance decisions in agriculture sector," in *International Conference on Communication, Management and Information Technology*, 2015.

[13] D. R. Wijaya, R. Sarno and E. Zulaika, "Sensor Array Optimization for Mobile Electronic Nose: Wavelet Transform and Filter Based Feature Selection Approach," *International Review on Computers and Software (IRECOS)*, 2016.

[14] [Online]. Available: https://hwpi.harvard.edu/files/chge/files/lesson_4_1.pdf.

[15] "Chapter 4: Food Safety Hazards," 2014. [Online]. Available: https://inspection.canada.ca/food-safety-for-industry/archived-food-guidance/non-federally-registered/product-inspection/inspection-manual/eng/1393949957029/1393950086417?chap=5.

[16] M. Maksimović, V. Vujović and E. O. Mikličanin, "A Low Cost Internet of Things Solution for Traceability and Monitoring Food Safety During Transportation," in *HAICTA 2015, 7th International Conference on Information and Communication Technologies in Agriculture, Food and Environment,*, At: Kavala, Greece, 2015.

[17] [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/food-safety.

[18] Y. Li, Application of data mining methods to establish systems for early warning and proactive control in food supply chain networks, 2010.

[19] G. Vlontzos and P. Pardalos, Data mining and optimisation issues in the food industry, vol. 3, 1, Ed., 2017.

[20] "Food safety: What you should know," 2015. [Online]. Available: https://apps.who.int/iris/handle/10665/160165.

[21] P. Farhana R., "Early detection of food pathogens and food spoilage microorganisms: Application of metabolomics," *Trends in Food Science & Technology-ELSEVIER*, vol. 54, 2016.

[22] G. Elmasry, D. F. Barbin, D. Wen Sun and P. Allen, "Meat Quality Evaluation by Hyperspectral Imaging Technique: An Overview," *Food Science and Nutrition*, vol. 52, no. 8, 2012.

[23] M. A. Bourlakis and P. W. Weightman, Food Supply Chain Management, 2008.

[24] M. Bortolini, R. Accorsi, M. Gamberi and F. Pilati, "Sustainable Food Supply Chains," in *Chapter 21 - A model to enhance the penetration of the renewables to power multistage food supply chains,* 2019.

[25] P. J. Kiger, "Supply Chain 101: What Happens When Our Food Supply Is Disrupted by a Pandemic?," 2020. [Online]. Available: https://money.howstuffworks.com/food-supply-chain-pandemic.htm.

[26] O. O. Ibrahim, "Introduction to Hazard Analysis and Critical Control Points (HACCP)," *EC Microbiology,* 2020.

[27] P. Ning Tan, M. Steinbach and V. Kumar, Introduction to data mining, 2016.

[28] A. O. Adebayo and M. S. Chaubey, "DATA MINING CLASSIFICATION TECHNIQUES ON THE ANALYSIS OF STUDENT'S PERFORMANCE," *Global Sceintific Journal,* vol. 7, no. 4, 2019.

[29] A. B. Annasaheb and V. K. Verma, "Data Mining Classification Techniques: A Recent Survey," *International Journal of Emerging Technologies in Engineering Research (IJETER),* vol. 4, no. 8, 2016.

[30] "Hazard analysis approaches for certain small retail establishments in view of the application of their food safety management systems," *European Food Sceince Autority,* 2017.

[31] L. Rokach and O. Maimon, "Decision Trees," in *Data Mining and Knowledge Discovery Handbook*, 2005.

[32] V. S. Kodogiannisa, . E. Kontogiannib and J. Lygouras, "Neural network based identification of meat spoilage using Fourier-transform infrared spectra," *Journal of Food Engineering,* 2014.

[33] S. K and S. Sasithra, "REVIEW ON CLASSIFICATION BASED ON ARTIFICIAL NEURAL NETWORKS," *International Journal of Ambient Systems and Applications (IJASA),* vol. 2, no. 4, 2014.

[34] L. Rokach and 0. Maimon, "Decision Trees," in *DATA MINING AND KNOWLEDGE DISCOVERY HANDBOOK*, 2005.

[35] P.-N. Tan, M. Steinbach, A. Karpatne and V. Vipin, Introduction to Data Mining, 2020.

[36] O. Yamini and S. Ramakrishna, "A Study on Advantages of Data Mining Classification Techniques," *International Journal of Engineering Research & Technology (IJERT),* 2015.

[37] G. Vlontzos and P. M. Pardalos, "Data mining and optimisation issues in the food industry," *International Journal of Sustainable Agricultural Management and Informatics,* 2017.

[38] S. A. A. Shah, "A Comparative Study of Feature Selection Approaches: 2016-2020," *International Journal of Scientific & Engineering Research,* vol. 11, no. 2, 2020.

[39] N. Sultan's, A. E. Khedr, A. M. Idrees and S. Kholeif, "Data Mining Approach for Detecting Key

https://digitalcommons.aaru.edu.jo/fcij/vol7/iss2/4
DOI: https://doi.org/10.54623/fue.fcij.7.2.4

Performance Indicators," *Journal of Artificial Intelligence* , 2017 .

[40] I. Sa, Z. Ge's, F. Dayoub's and B. Upcroft's, "DeepFruits: A Fruit Detection System Using Deep Neural Networks," *MDPI (Sensors journsl)*, 2016.

[41] S. Li,, Y. Xu, Y. He, Z. Geng, Z. Jiang and Q. Zhu, "Research on public opinion warning based on analytic hierarchy process integrated back propagation neural network," *Chinese Automation Congress (CAC)*, 2017.

[42] A. Zakeri, M. Saberi, O. K. Hussain and E. Chang, "An Early Detection System for Proactive Management of Raw Milk Quality: An Australian Case Study," *IEEE*, 2018.

[43] Q. Dai, D.-. Wen Sun, Z. Xiong, J.-. Hu Cheng and X.-. An Zeng, "Recent Advances in Data Mining Techniques and Their Applications in Hyperspectral Image Processing for the Food Industry," *Journal of food safety* , 2014.

[44] R. Tardío and J. Peral, "Obtaining Key Performance Indicators by Using Data Mining Techniques," in *International Conference on Conceptual Modeling*, 2015.

[45] . T. Tiao Pan, D. Wen Sun, J. Hu Cheng and H. Pu, "Regression Algorithms in Hyperspectral Data Analysis for Meat Quality Detection and Evaluation," *IFT Scientific Journals*, 2016.

[46] Y. Y. Y, H. K. Choy,, G. Ho and H. Lam, "A fuzzy association Rule Mining framework for variables selection concerning the storage time of packaged food," in *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2016.

[47] H. Zhong, C. Lin, C. Wu and G. Zhang, "Research on dairy supply chain traceability system and its pre-warning model," in *2016 International Conference on Logistics, Informatics and Service Sciences (LISS)*, 2016.

[48] M. Bortolini, R. Accorsi, . M. Gamberi and . F. Pilati, "A model to enhance the penetration of the renewables to power multistage food supply chains (Chapter 21)," in *Sustainable Food Supply Chains*, 2019, pp. 305-315 .

[49] J. wang, and H. Yue, "Food safety pre-warning system based on data mining for a sustainable food supply chain," *Food Control*, 2016.

[50] T. Liu and A. Hu, "Model of Combined Transport of Perishable Foodstuffs and Safety Inspection Based on Data Mining," *Food and Nutrition Sciences*, 2017.

[51] J. Cai, J. Luo, S. Wang and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, 2018.

[52] A. Lukyamuzi, J. Ngubiri and W. Okori, "Tracking food insecurity from tweets using data mining techniques," in *Proceedings of the 2018 International Conference on Software Engineering in Africa*, 2018.

[53] D. R. Wijaya, R. Sarno and E. Zulaika, "Noise filtering framework for electronic nose signals: An application for beef quality monitoring," *Computers and Electronics in Agriculture (COMPUT ELECTRON AGR)*, 2019.

[54] L. Zhou, C. Zhang, Z. Qiu and Y. He, "Application of Deep Learning in Food: A Review," *Comprehensive Reviews in Food Science and Food Safety journal of (Institute of food technologies -JFT Journals)*, 2019.

[55] Z. R. Sapautra, H. Pratiwi, A. e. Windarto and F. Wiza, "Utilization of Data Mining Techniques in National Food Security during the Covid-19 Pandemic in Indonesia," *Journal of Physics Conference Series*, 2020.

[56] M. Ilic, S. Ilic, S. Jovic and S. Panic, "Early cherry fruit pathogen disease detection based on data mining prediction," *Computers and Electronics in Agriculture* , 2018.

[57] N. Abdul Ghafoor and B. Sitkowska, "A Machine Learning Application to Predict Risk of Mastitis in Cattle from AMS Sensor Data," *AgriEngineering*, 2021.

[58] S. Nehaya, A. E. Khedr, A. M. Idrees and S. Kholeif, "Data Mining Approach for Detecting Key Performance Indicators," *Journal of Artificial Intelligence*, March 2017.

[59] D. R. Wijaya, R. Sarno and E. Zulaika, "Sensor Array Optimization for Mobile Electronic Nose: Wavelet Transform and Filter Based Feature Selection Approach," *International Review on Computers and Software*, 2016.

https://digitalcommons.aaru.edu.jo/fcij/vol7/iss2/4
DOI: https://doi.org/10.54623/fue.fcij.7.2.4