**USING RPA AND AI TO DESIGN AN AUTOMATIC EMAIL**

**CLASSIFYING AND FORWARDING SYSTEM IN BUSINESS**

***Author: Nguyen Quang Hoc[[1]](#footnote-1), La The Anh, Vu Luu Hoang Lan, Gia Ngoc Thao Ly, Phan Ngoc Bao Tam***

***Mentor: Ph.D Nguyen Thon Da[[2]](#footnote-2)***

*University of Economics and Law - Vietnam National University Ho Chi Minh City*

|  |
| --- |
| **ABSTRACT** |
| Enterprises may receive thousands of emails every day. It must be processed briefly to respond to customers and partners. However, this process is time-consuming in classifying emails to be handled by the right departments. Therefore, a model to automate the email classification process in the enterprise is proposed in this study. This model will rely on Artificial Intelligence (AI) techniques with supervised Machine Learning algorithm and Nature Language Process (NLP) to classify emails - right for the specialized work of each department in the business - and build Robotic Processing Automation (RPA) bot to receive and forward them to the department. The experimental data set is collected from Song Linh Trading and Service Co., Ltd. It includes 4270 emails from 3 departments: Sales, Customer care, and External Relations. We have tested and compared four text classification machine learning techniques (SGD Classifier, Linear SVC, Gradient Boosting Classifier, Extra Trees Classifier) ​​combined with two classification methods (Multi-class, Multi-tasking) to determine the most suitable classification model. Through model testing, the results show that Linear SVC with the Multi-class method gives the best accuracy and the automation system saves ten times more time than manual classification. RPA email forwarding also results in high accuracy. This model will benefit the time, accuracy, and security of email processing.  **Keywords:** Email, Artificial Intelligence AI), Robotic Process Automation (RPA), Email Classification, UiPath. |

1. Introduction

In recent years, with the development of technology, electronics, and networks, the demand for communication via email has been increasing. According to the report of Statista, 306,4 billion emails were sent in 2020. The predicted sent email will grow up to 376,4 billion in 2025 (Source: Statista). Enterprises are constantly faced with a large number of emails every day, along with the task of classifying them into the right specialized departments. In reality, enterprises are always specialized by dividing into many departments. Therefore, manually categorizing emails by department will consume a lot of time and effort. In addition, manual sorting is difficult to guarantee the time that emails are classified. There may even be the possibility of an incorrect classification, and missing emails due to the subjectivity of the classifier. Therefore, we have solved this email classification problem using Artificial Intelligence ([Kang, Cai, Tan, Huang, & Liu](#_ENREF_15)) techniques and Robotic Processing Automation (RPA) to propose an automatic classification model. Supervised Machine Learning Algorithms - Support Vector Machine (SVM) and Natural Language Processing (NLP) are used to classify and label departmental emails, then Robotic Processing Automation (RPA), built on UiPath software, will automatically direct the emails to the correct department.

Emails written by humans in natural language. This is the name for unstructured language, used by humans to communicate with each other. Therefore, in order for computers to understand the semantics of emails, NLP techniques are used to get them to structured data - a low-level language or computer language. Thereby, the computer can analyze and classify them according to the appropriate context. Computers rely on Machine Learning to learn and recognize the characteristics of each email and each category to determine which category the email belongs to. The Supervised Machine Learning model applied in this study will be based on the classified email data set from Song Linh Trading and Service Co., Ltd to learn the characteristics. Then, for each new email sent to the business, the RPA model will receive, read and pass it to the trained Machine Learning model for classification and labeling. After completing the classification, the RPA bot forwards the email to the department that has been labeled. However, each email has a confidence index calculated after the classification process, if this index does not reach a certain confidence threshold, RPA bot will notify them to humans for manual classification.

There is a lot of research related to email classification. However, according to Mujtaba et al.'s research overview report on email classification (2017), most of the research is spam classification. The topic of multi-folder categorization of email has not been noticed in the world, especially in Vietnam. Besides, most of the studies stop at "classification" without any other special treatment such as forwarding or feedback [Mujtaba, Shuib, Raj, Majeed, and Al-Garadi (2017)](#_ENREF_21). Some other studies such as ([Alghoul, Al Ajrami, Al Jarousha, Harb, & Abu-Naser, 2018](#_ENREF_3); [Patel, Shukla, Porwal, & Kotecha, 2019](#_ENREF_23)) have proposed an email processing model, but there are still some shortcomings such as not applying Machine Learning to increase accuracy and no practical application in business. RPA has little applied research and has not been popularized in Vietnam. Therefore, this study proposes a solution and application in the field of "Segmentation of emails by subject", combined with spam classification to provide the best solution for the topic of email classification. The study compares many classification techniques and algorithms in Machine Learning to build classification models with high accuracy. The data set was collected at a real company (Song Linh) to ensure the applicability of the research. Moreover, this is a new study in RPA and proposes a practical application of this field in Vietnam.

1. Theoretical framework
   1. E-mail

According to ([Mujtaba et al., 2017](#_ENREF_21)), the strong development of technology and the Internet worldwide has led to the electronic mail system gradually replacing the traditional mail system. Electric mail (e-mail) is one of the indirect communicative methods between electronic device users. E-mails are divided into 4 main types, including introduction mail, news mail, advertising mail, and pure document mail ([Alghoul et al., 2018](#_ENREF_3); [Duc](#_ENREF_11)).

E-mails gradually become a popular contact method for most users. It is seen as a communication channel similar to mobile-phone due to its familiarity with Internet users, so millions of emails are sent every day ([Dürscheid, Frehner, Herring, Stein, & Virtanen, 2013](#_ENREF_12)). Processing these emails demands company so much time and cost. In addition, ([Toàn, Lâm, Nghị, & Trung, 2011](#_ENREF_26)) has pointed out that the rapid development of email is also a double-edged sword for users. Taking advantage of the benefits of email, some individuals and organizations have used email for improper purposes such as sending too many promotional messages, reactionary letters, and even malicious code, which we call that email spam.

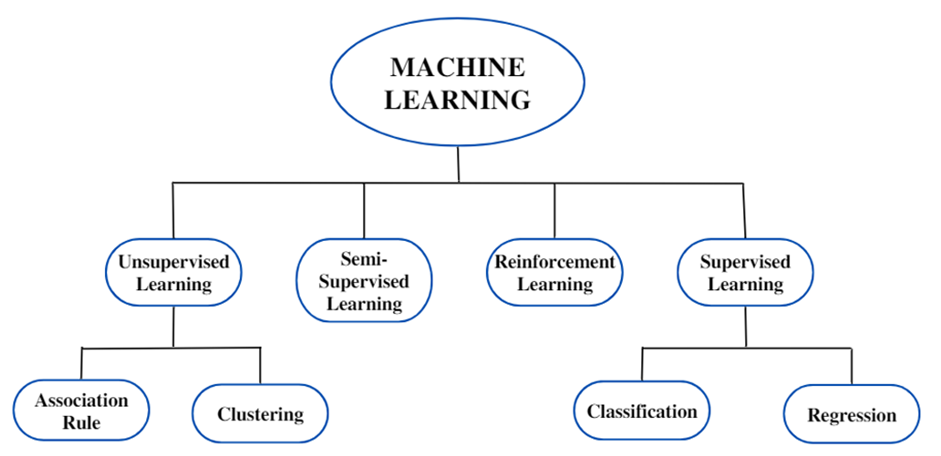
* 1. Natural Language Processing

([Fernandes, 2008](#_ENREF_13)) defined that “Natural language can be broadly defined as the opposite of artificial or constructed language”. Natural language is the type of unstructured data, which does not conform to any data models. Natural language processing (NLP) may be divided into 2 non-independent fields completely, including speech and text processing. Companies have also used this technology for automated tasks, such as processing, analyzing, and storing large documents; customer feedback analysis; and running chatbots for automated customer service.

According to ([Kang et al., 2020](#_ENREF_15)), NLP processes can be divided into 4 main stages: Data preprocessing; Data representation; Model training; Evaluation of project results and implementation. Data preprocessing aims to achieve “clean” data to improve model efficiency and accuracy by removing meaningless objects. Then, choose the best way to represent the data in a form that the computer can understand. Applied algorithms to train and test a model to ensure that it has optimal generalizability.

* 1. Machine Learning
     1. Text classifier machine learning

Machine Learning is a field of artificial intelligence that deals with the research and construction of techniques that allow systems to "learn" automatically from data to solve specific problems ([Mahesh, 2020](#_ENREF_17)).



**Figure 1. Machine learning taxonomy**

*Source: (*[*Bi, Goodman, Kaminsky, & Lessler, 2019*](#_ENREF_4)*)*

Supervised Learning is a form of machine learning that predicts the output for a new data set based on a given training data set in which each data sample has a known output or value destination. Supervised learning is divided into two main categories, including Classification and Regression. Classification is the case where the output values of the input data are divided into a finite number of classes (discrete value domain), for example determining whether the animal image is a dog or a cat. The process of building a machine learning system usually has 3 main steps, such as collecting and processing data; algorithm selection and training for the model; actual testing and evaluation. Data collected is very important to determine the predictability of the system, it is necessary to perform processing to use it easily. Some machine learning solutions incorporate NLP such as: Automated Algorithms ([Zong & Hong, 2018](#_ENREF_27)), Sentiment ([Mejova, 2009](#_ENREF_19)), Chatbots ([Shah, Lahoti, & Lavanya, 2017](#_ENREF_25)), Automatic Text Classification ([Dalal & Zaveri, 2011](#_ENREF_8)), ... These technologies will help organizations and businesses analyze data, mine information, automate processes and gain competitive advantage.

* + 1. Text classification methods
       1. Multi-class classification

Multiclass classification is the classification of more than two classes. Each sample can only be labeled as a class. The higher the number of classes, the more difficult it is to have high accuracy.

One-vs-Rest (abbreviated as OvR, also known as One-vs-All or OvA) is a technique that uses binary classification algorithms for multiclass classification. Accordingly, each binary classifier will classify between a class and a set of remaining classes.

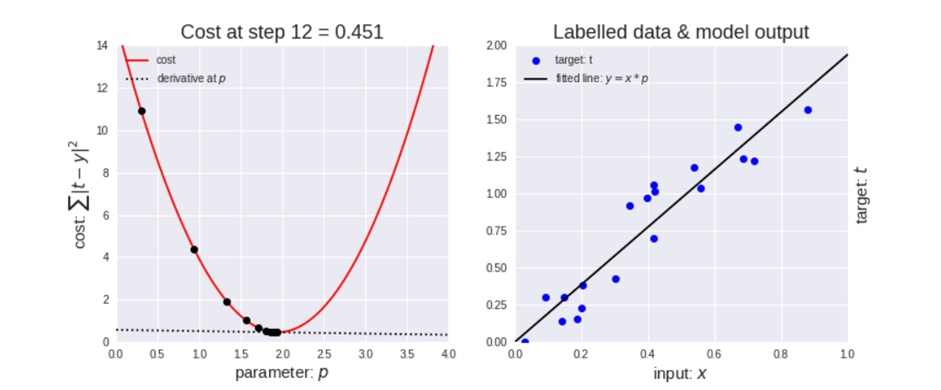
One-vs-One (OvO) is a technique that uses binary classification algorithms for multi-class classification. The OvO divides the multi-class classification dataset into binary classifiers. But instead of comparing a class with a set of classes, OvO classifies each class against each other.

* + - 1. Multiclass – multioutput classification

Multiclass - multioutput classifier (also known as multitasking classifier) is a classification task that labels each sample with a set of non-binary attributes. This is both a generalization of the multi-label classification task, which considers only binary attributes, and a generalization of the multi-class classification task, in which only one attribute is considered. For example, email is classified into two categories: spam email and non-spam email. Non-spam emails carry a lot of content such as business, foreign affairs, or marketing. So, there will be 2 labels that need to be correctly identified.

* + 1. Text Classification algorithms
       1. Stochastic Gradient Descent Classifier

Literally, descending means moving downhill to reach the lowest point on the curve. This is done repeatedly until the minimum point is reached. The three variants of gradient descent include batch (Batch), random (Stochastic), and mini-batch (Mini-Batch) gradient descent approaches ([Deepa et al., 2021](#_ENREF_9)).



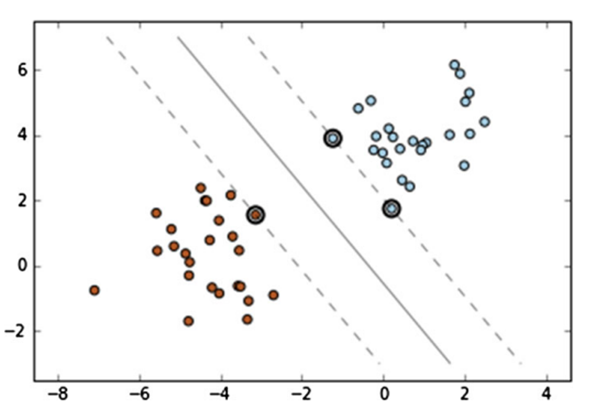
**Figure 2. Gradient descent optimisation**

*Source: Onfido Product and Tech, 2018*

SGDClassifier is a simple but effective optimization algorithm that provides the best results for the data. Instead of using the entire training set, this method will randomly take an element in the training set and recalculate the slope vector based on a data point, then iterate until the end. SGDClassifier is faster than conventional gradient descent because the updates are performed immediately after training each sample.

* + - 1. Linear Support Vector Classification

Support Vector Machine (SVM) is a monitoring algorithm that can be used for either classification or regression but is mainly used for classification. The goal of SVM is to find a hyperplane in multidimensional space to divide the data into parts corresponding to their number of layers, one of the proposed methods is a Linear Support Vector Classifier (Linear SVC).



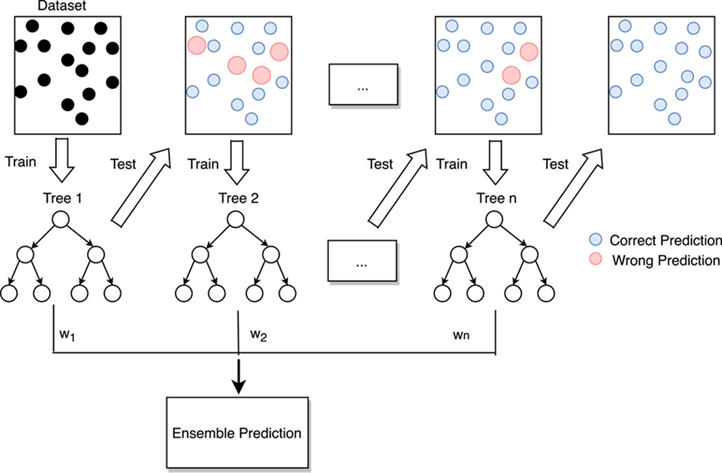
**Figure 3. An infinite number of classifiers can be drawn for the given data but SVM finds the classifier with the largest gap between support vectors. Circles represents the support vectors**

*Source: Chauhan et al (2019), Artificial Intelligence Review*

LinearSVC applies a linear multiplication function to perform the classification and it works well with a large number of samples. The training of linear SVM is much faster than non-linear SVM due to their difference in computational complexity ([Chauhan, Dahiya, & Sharma, 2019](#_ENREF_7)).

* + - 1. Gradient Boosting Classifier

Gradient Boosting is a method used to develop classification and regression models to optimize model learning, which is mostly non-linear in nature and known. More commonly known as decision trees or regression trees ([Abraham, Dutta, Mandal, Bhattacharya, & Dutta, 2018](#_ENREF_2); [Chakrabarty, Kundu, Dandapat, Sarkar, & Kole, 2019](#_ENREF_6)).



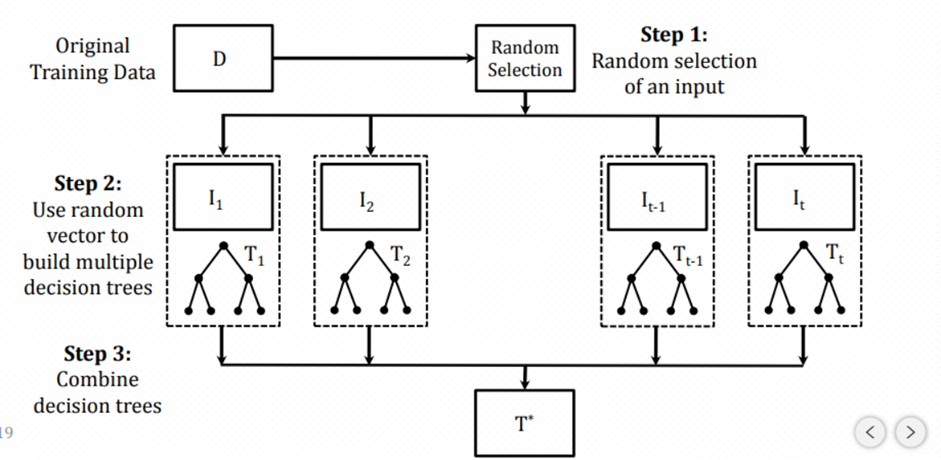
**Figure 4. Flow diagram of gradient boosting machine learning method**

*Source: ResearchGate*

Gradient Boosting Classifier (GBC) is a group of machine learning algorithms that combine multiple weak learning models together to create a strong predictive model. GBC tries to reduce errors by resampling and changing the weights of each weak “learner” to increase classification accuracy. ([Bowd et al., 2020](#_ENREF_5)) pointed out that “GBC had the best overall performance”. Compared with gradient descent, the slope is by introducing changes to the parameters, while gradient boosting reduces the gradient by introducing new models. Then the parameters of the tree are modified to reduce the residual loss.

* + - 1. Extra Trees Classifier

Extra Trees Classifier is a synchronous machine learning method that trains multiple decision trees and aggregates the results from a group of decision trees to make predictions ([Abhishek, 2020](#_ENREF_1)). Extra Trees Classifier uses the entire data set to train the decision tree. As such, to ensure enough difference between individual decision trees, it randomly selects values for analysis. This helps the model reduce bias and save time, checking the computational cost, Extra Trees is much faster than Random Forest.



**Figure 5. Visual Explanation of Extra Trees Classifier**

*Source: Navoneel et al (2018), ICACCCN*

* + 1. Algorithm evaluation parameters

Classification is an example of supervised learning so evaluating the models will help improve overall predictability before we deploy this model for production on unseen data.

**Table 1. Confusion Matrix for Binary Classification and the Corresponding Array Representation used**

|  |  |  |
| --- | --- | --- |
|  | **Actual Positive Class** | **Actual Negative Class** |
| **Predicted Positive Class** | True positive (TP) | False negative (FN) |
| **Predicted Negative Class** | False positive (FP) | True negative (TN) |

*Source: Authors*

From Table 1, several commonly used metrics can be generated as shown in Table 2 to evaluate the performance of classifiers with different focuses of evaluations.

**Table 2. Threshold Metrics for Classification Evaluations**

|  |  |  |
| --- | --- | --- |
| **Metrics** | **Formula** | **Evaluation Focus** |
| Precision *(*[*Mujtaba et al.*](#_ENREF_21)*)* |  | P measures the positive patterns that are correctly predicted from the total predicted patterns in a positive class. |
| Recall *(R)* |  | R measures the fraction of positive patterns that are correctly classified. |
| F1-Measure *(F1)* |  | F1 represents the harmonic mean between recall and precision values. |
| Kappa () |  | |

*Source: Authors*

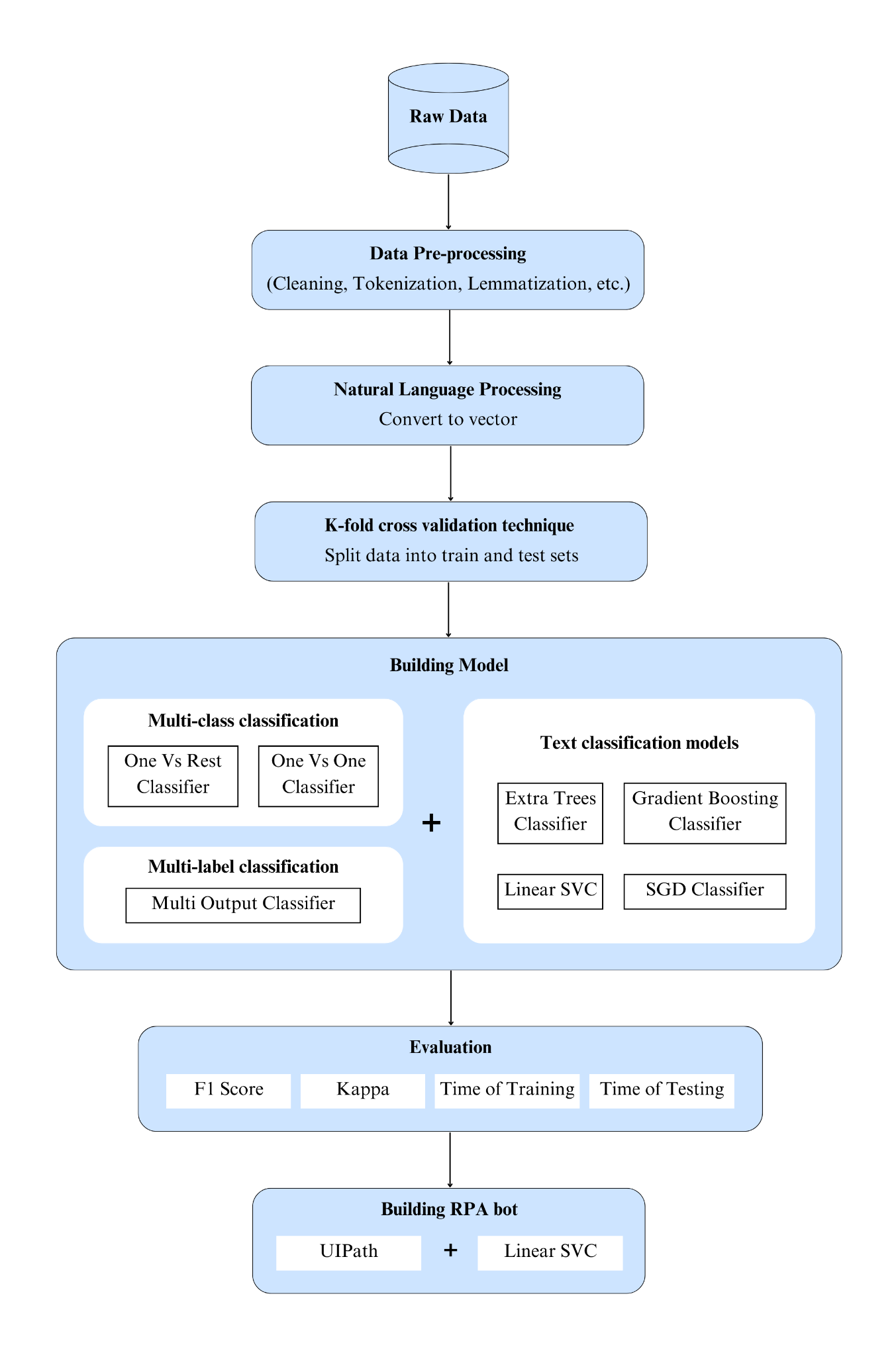
Cohen's kappa coefficient ([Mohammadi, Malekian, Nosrati, & Karimi](#_ENREF_20)) measures the agreement between two raters who each classify N items into C mutually exclusive categories. It is considered an efficient method in multi-class classification to calculate the percentage of consensus ([McHugh, 2012](#_ENREF_18)).

* 1. Robotic Processing Automation

Robotic Process Automation (RPA) is a software technology that can simulate how humans interact with software to perform high-volume, repeatable tasks. The use of RPA helps reduce costs for human resources, save time, as well as limit the risks of errors in the business process through workflow automation. Users can build bots using RPA automation by analyzing human behavior on IT applications. Show the bots what to do, and leave them to it. For businesses, RPA can bring immediate value to core business processes including the payroll, hiring of new employees, receipts and payables, invoice processing, inventory management, report generation, software installation, data, vendor migration onboarding, and so on ([Madakam, Holmukhe, & Jaiswal, 2019](#_ENREF_16)).

1. Research method

The diagram in Figure 6 provides an outline of the proposed model's architecture and highlights the different phases involved in its development.

****

**Figure 6. Research method**

*Source: Authors*

The main purpose of this model is to build a bot that can forward emails automatically to departments. The email department classification model is built by constructing machine learning models and comparing them to find the optimal model for classification. The study then integrated the optimal model with UiPath software to automate the classification and forwarding of emails within the enterprise. It incorporates the preprocessing techniques, natural language processing techniques, and combines them with multi-label and multi-class classification models in order to obtain an optimal email classification model.

* 1. Data cleaning
     1. Remove missing values

Missing values occur when no storage is recorded in a variable in an observation. In this dataset, missing values may occur when an email is not labeled, a label is attached to an email with no content, or a row of data is empty.

* + 1. Checking duplicate data

Duplicate values are those identical values that are repeated in multiple rows of data. Removing duplicate values helps eliminate unnecessary values and improves the quality of the dataset. In the dataset used in the study, there were no duplicate values found after inspection; therefore, there was no need for a step to remove duplicate values.

* 1. Natural Language Processing
     1. Tokenization

Tokenization is a pre-processing method that separates a text into words, phrases, symbols, or other meaningful elements called tokens. Tokenization is an important step in natural language processing because it allows computers to understand the basic units of a text, making the analysis and processing of text easier. Tokens can be used to calculate the frequency of words, create feature vectors to feed into machine learning models, or perform other tasks such as translation or text summarization.

* + 1. Lowering capitalization

Although uppercase and lowercase letters do not differ in meaning, this can cause significant issues and affect the results of text classification tasks. The most common approach to handling this is to convert all letters to lowercase. This technique projects all words in the text into the same feature space, but it causes an important issue when interpreting some words, such as "US," which becomes "us" after being converted to lowercase ([Gupta & Lehal, 2009](#_ENREF_14)). A way to overcome this limitation is to use a slang and abbreviation conversion tool.

* + 1. Removing slang and abbreviation

An abbreviation is a shortened form of a word or phrase which contains mostly the first letters of the words. Slang is a subset of the language used in informal talk or text. A common method for dealing with these words is converting them into formal language ([Dhuliawala, Kanojia, & Bhattacharyya, 2016](#_ENREF_10)).

* + 1. Noise Removal

Most of the text and document data sets contain many unnecessary characters such as punctuation and special characters, which can be detrimental to classification algorithms ([Pahwa, Taruna, & Kasliwal, 2018](#_ENREF_22)).

* + 1. Lemmatization

Lemmatization is the NLP process that replaces the suffix of a word with a different one or removes the suffix of a word completely to get the basic word form (lemma) ([Sampson, 2005](#_ENREF_24)).

* + 1. Convert to vector

For various types of textual language or documents, computers cannot understand and process them in a conventional way, but rather need to represent them as vectors so that the computer can understand their meaning. Some commonly used models for text vectorization are: Bag of Words, TF-IDF, Distributional Embedding, etc. In this study, we used the TF-IDF model for implementation.

TF-IDF (Term Frequency - Inverse Document Frequency) is a basic technique for calculating the weight of a word in a text. The weight represents the level of importance of that word in the text, which is in a collection of many texts.

* 1. K-fold sampling technique

K-Fold is a method of randomly dividing data into training and testing sets to evaluate machine learning models. This is an extremely optimal method for datasets with few observations, making it suitable for our research. The value of k is fixed at 10, a commonly used value that has been proven to result in low errors and low variance. Our group also uses this index to divide the data.

* 1. Building model
     1. Text classification methods

Not all classification algorithms support multi-class classification. Some algorithms are designed to classify binary classes and therefore, naturally, these algorithms do not support multi-class classification problems. However, there is a method to apply binary classification algorithms to multi-class classification problems, which is to convert the multi-class classification problem into multiple binary classification problems, and then use binary classification algorithms on each small problem. Our research uses 2 methods to convert multi-class, multi-label classification problems into binary classification problems.

* + - 1. Multi-label classification with 2 steps

(1) Classify to filter out spam emails, non-spam emails will be passed to the next process.

(2) Classify the topics of non-spam emails to forward them to relevant departments for processing.

For this classification method, the research group will use the Multi-Output-Classifier technical to evaluate its effectiveness.

* + - 1. Multi-class classification

Combining the two processes mentioned in Multi-label classification into one to classify topics (including spam label and departments) at once. This means there will be 4 labels representing 3 departments and 1 spam email label. The study uses One-Vs-One Classification and One-Vs-Rest Classification.

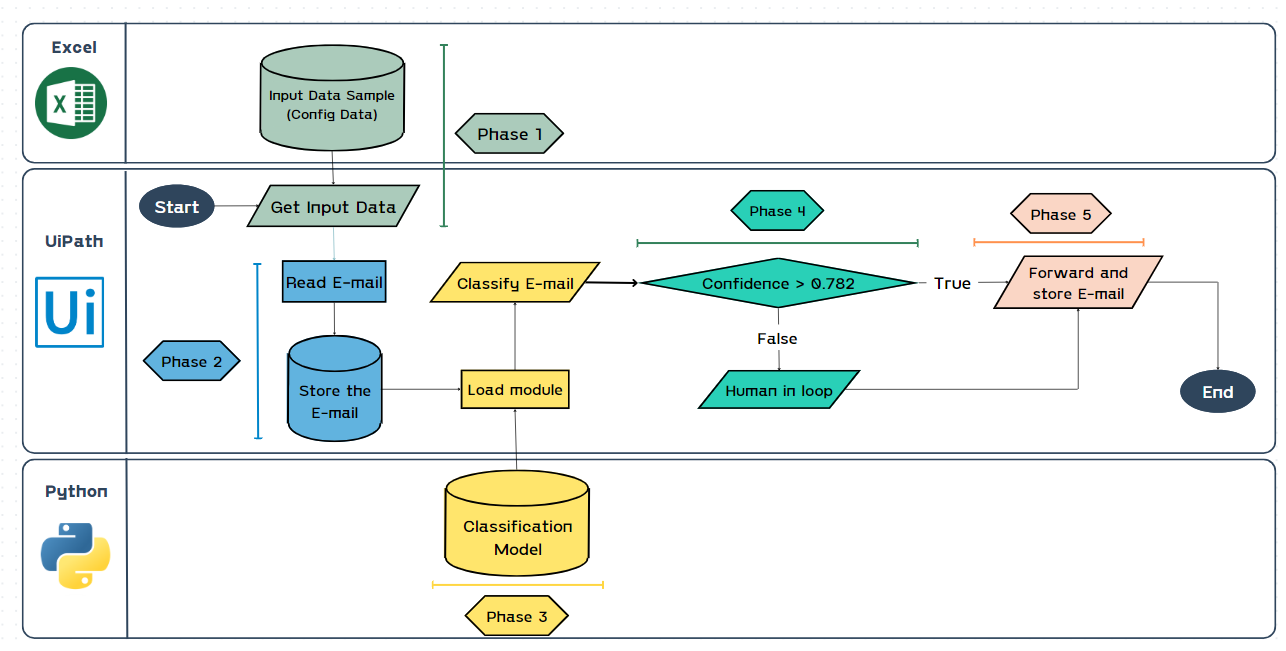
* + 1. Text classification algorithms

The study builds a classification model by combining 4 machine learning algorithms with the 3 text classification techniques mentioned above. In other words, the study constructs 12 text classification models to find the optimal model after evaluating the results, and uses that model to classify the departments.

* 1. Evaluation

The research uses F1-Score, Kappa coefficient, training time, and testing time as evaluation metrics to compare the 12 text classification models and find the most suitable and optimal model for email classification.

* F1-Score: Calculates the harmonic mean between the precision and recall of each label's predicted and actual results.
* Kappa coefficient: Measures the agreement between two classification results. It represents the proportion of the actual similarity between two times compared to the probability of them predicting the same.
* Training and testing time: Determines the performance of the model.
  1. Proposing an automation process flow for RPA bot



**Figure 7. Automation process flowchart**

*Source: Authors*

Figure 7 is the workflow proposed by our research to design an RPA automation process flow. It includes 5 phases.

Phase 1: Inputting configuration data

Phase 2: Reading and storing emails

Phase 3: Integrating UiPath with Python

Phase 4: The process of classifying

Phase 5: Forwarding and storing email

* + 1. Inputting configuration data

The objective of this phase is to provide essential input parameters, namely the Outlook email account, the folder name where the email is stored, the department names for classification, the corresponding email addresses, and the confidence index. This study is used to classify 4 labels: 3 departments Customer Care, Sales, External Relations, and Spam.

**Table 3. Input data table**

|  |  |  |
| --- | --- | --- |
| **Name** | **Description** | **Type** |
| Mail Account | *Outlook Account* | String |
| MailFolder | *Folder Name* | String |
| Department 1 | *Department Name* | String |
| Department 2 | *Department Name* | String |
| Department 3 | *Department Name* | String |
| Email Department 1 | *Department Email* | String |
| Email Department 2 | *Department Email* | String |
| Email Department 3 | *Department Email* | String |
| Confidence | *Evaluate the classification result* | Float |

Source: Authors

* + 1. Reading and storing emails

Next, this study uses the IMAP protocol and UiPath's Get IMAP Mail Message activity to retrieve emails from Outlook. The IMAP protocol offers many advantages such as efficient handling of large volumes of email and saving local storage space. Furthermore, business emails are stored on Mail Servers, improving storage and security capabilities.

* + 1. Integrating UiPath with Python

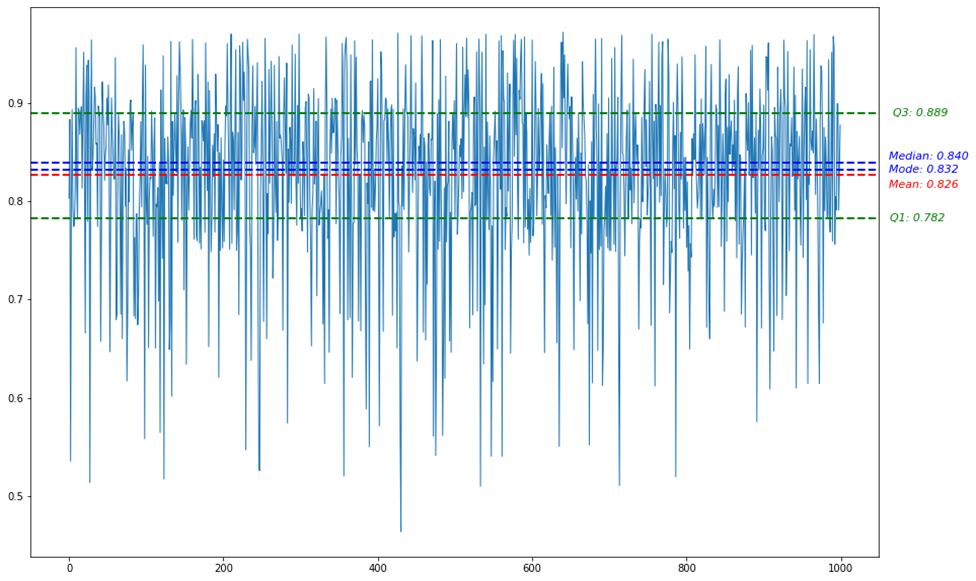
This study utilized the UiPath.Python.Activities extension to connect Python with UiPath. Then, the text classification model was passed to perform the classification.

* + 1. The process of classifying

After the text classification model has been integrated into UiPath, the next phase is to provide the content of each email to the model. Each email will be classified by the model and return two results: the department classification and the accuracy of the classification.

The accuracy score of each email will be compared with the confidence index mentioned in phase 1. If the accuracy score of that email is higher than the input confidence index, the email will be automatically forwarded to the corresponding department, otherwise it will do "human in loop" (requires human intervention for classification).

To determine the confidence index, the research team ran a test of 1000 emails and obtained the classification accuracy results of each email as follows:



**Figure 8. Test results for calculating confidence index over 1000 emails**

*Source: Authors*

Based on the results of the quartile analysis, it was found that over 50% of the emails had a classification accuracy of over 82.6%, and over 75% of the emails had a classification accuracy of over 78.2%. Prioritizing the concentration of classification values, the research team decided to set the model's confidence value at 78.2% or 0.782.

After the emails are classified with their accuracy, the process will proceed to compare the accuracy of each email with the confidence value (set at 0.782). If the accuracy of an email is equal to or greater than 0.782, it will be automatically forwarded to its classified department. Otherwise, if the accuracy is lower than 0.782, the "human in loop" step will be performed.

The "human in loop" step will notify the reason for low accuracy and require human intervention to select the appropriate department for the email. The selection dialog will have 5 options including 3 default departments (Customer Care Department, Sales Department, External Relation Department), Spam, and Other (allow the user to enter an email address to send to).

* + 1. Forwarding and storing email

The UiPath software will notify the user of the number of emails sent to each department and the number of unclassified emails requiring human intervention. However, this feature only helps users manage emails for each department and cannot track all emails of the business. To facilitate tracking and storing email content after the process is complete, the research team will manage and store emails as a dataset and save them in an Excel file including email content, forwarding method, department name, email address, and email status.

**Table 4. The components of the email storage system**

|  |  |
| --- | --- |
| **Name** | **Description** |
| **Email content** | *Content of forwarded email* |
| **Forwarding Method** | *Email forwarding methods, including automatic forwarding (Auto) and human intervention (Human in loop)* |
| **Department Name** | *The name of the department to which the email was forwarded* |
| **Email Address** | *Address of email* |
| **Email Status** | *Update email status whether or not forwarded* |

*Source: Authors*

After email storage, the automated email classification process is completed and the email data is added to the training data source to support the next email classification process.

1. Results and discussion
   1. Results
      1. Model processing and text classification (email)
         1. Evaluation and selection of classification methods

To determine the appropriate classification method, the research compared the effectiveness of two methods:

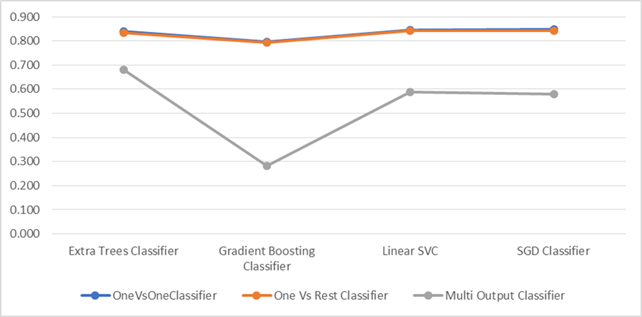
* Multi-class classification: Emails are classified into a single label, which includes four categories representing three departments and spam email (Customer Care, External Relationship, Sales, Spam Email). For this method, the research team used two techniques: One-Vs-One and One-Vs-Rest.
* Multi-label - multi-class classification (multi-task classification): Emails are classified into two labels. The first label includes two categories: spam (1) and non-spam (0). For non-spam emails, the second label includes three categories representing the three departments in the problem. For this method, the Multi-output method is used.

The data was trained using each method, and accuracy was analyzed to determine the results:

**Table 5. Accuracy of classification methods**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **OneVsOneClassifier** | **One Vs Rest Classifier** | **Multi Output Classifier** |
| **Extra Trees Classifier** | 0,839 | 0,834 | **0,682** |
| **Gradient Boosting Classifier** | 0,797 | 0,793 | **0,282** |
| **Linear SVC** | 0,845 | 0,842 | **0,589** |
| **SGD Classifier** | 0,847 | 0,842 | **0,580** |

*Source: Authors*



**Figure 9. Comparison of classification accuracy results**

*Source: Authors*

Based on the analysis results, the multi-class classification method is more effective with an average accuracy of 83.2%, which is 1.56 times higher than the multi-label classification method (only 53.3%). Therefore, the research team decided to choose the multi-class classification method. It was observed that in the multi-label method, both One-Vs-One and One-Vs-Rest techniques give almost the same results (a difference of 0.48%). Therefore, it is necessary to use other evaluation metrics to choose the algorithm that provides the highest classification efficiency instead of solely relying on accuracy.

* + - 1. Evaluation and selection of classification model

The research team proposed 4 specific models: Stochastic Gradient Descent (SGD) Classifier; Linear Support Vector Classification (Linear SVC); Gradient Boosting Classifier; and Extra Trees Classifier. After training and evaluating the results, the following outcomes were achieved:

**Table 6. Results of the experiments with each model**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No.** | **Model** | **F1 Score** | ***Coefficient K*** | **Training time** | **Testing time** |
| 1 | Extra Trees Classifier | 0,8301 | 0,7726 | 27,0909 | 0,4232 |
| 2 | Gradient Boosting Classifier | 0,8033 | 0,7488 | 26,9546 | 0,0164 |
| **3** | **Linear SVC** | **0,8445** | **0,7908** | **0,1122** | **0,0032** |
| 4 | SGD Classifier | 0,8385 | 0,7857 | 0,0877 | 0,0037 |

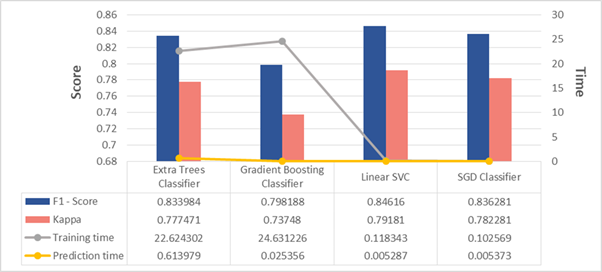
*Source: Authors*

When evaluating each model, the Linear SVC model achieved the highest result with an F1 score of 84,45% and a Kappa ([Mohammadi et al.](#_ENREF_20)) score of 79,08%. In terms of training and testing time, the Linear SVC model had the shortest prediction time of 0,0032 seconds. In terms of training time, the Linear SVC model (0,1122s) ranked second after the SGDClassifier model (0,0877s), although there was a difference in training time, the accuracy evaluation scores of the Linear SVC model were still higher. To have an overall evaluation, the research team will analyze the evaluation of the model combined with the multi-class classification algorithms.

**Table 7. Results of the classification model experiments**

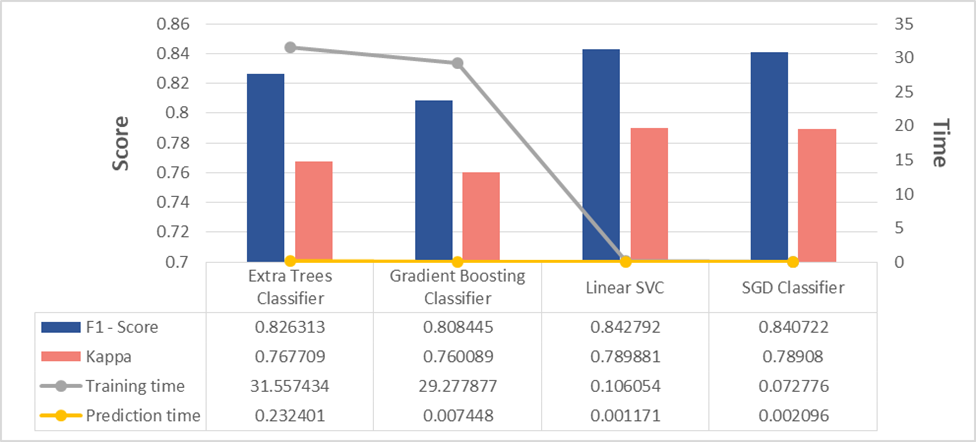
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Model** | **Multi-label classification** | **F1 Score** | **Coefficient K** | **Training time** | **Testing time** |
| **1** | SGD Classifier | *One Vs Rest* | 0,840722 | 0,78908 | 0,072776 | 0,002096 |
| *One Vs One* | 0,836281 | 0,782281 | 0,102569 | 0,005373 |
| **2** | **Linear SVC** | *One Vs Rest* | 0,842792 | 0,789881 | 0,106054 | 0,001171 |
| ***One Vs One*** | **0,84616** | **0,79181** | **0,118343** | **0,005287** |
| **3** | Gradient Boosting Classifier | *One Vs Rest* | 0,808445 | 0,760089 | 29,277877 | 0,007448 |
| *One Vs One* | 0,798188 | 0,73748 | 24,631226 | 0,025356 |
| **4** | Extra Trees Classifier | *One Vs Rest* | 0,826313 | 0,767709 | 31,557434 | 0,232401 |
| *One Vs One* | 0,833984 | 0,777471 | 22,624302 | 0,613979 |

*Source: Authors*



**Figure 10. Applying One-Vs-One multiclass classification method**

*Source: Authors*

****

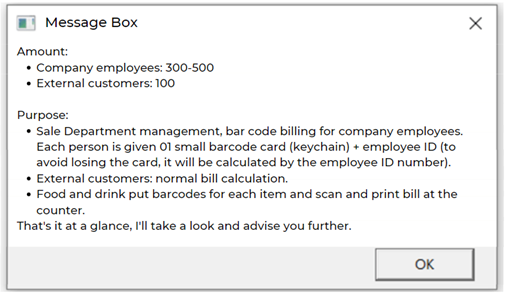
**Figure 11. Applying One-Vs-Rest Multiclass classification method**

*Source: Authors*

When evaluating the 4 models with the One-Vs-One and One-Vs-Rest techniques, the Linear SVC model continued to achieve the highest scores in both algorithms, with an F1 score of 84.616% for OvO and 84.279% for OvR, and a k score of 79.18% for OvO and 78.99% for OvR. Regarding training time, although it was lower than the SGD Classifier model, the difference was very small (less than 0.5 times), and the training time was the shortest among the 4 models. Regarding the training-evaluation time of the Linear SVC model with 2 algorithms, the OvO technique was longer than OvR with an average difference of 1.815, but in terms of F1 and k evaluation scores, OvO was higher than OvR. As accuracy was prioritized, the research team decided to choose **The Linear SVC model combined with the One-Vs-One multi-class classification technique**.

* + 1. Automated email recognition and forwarding process

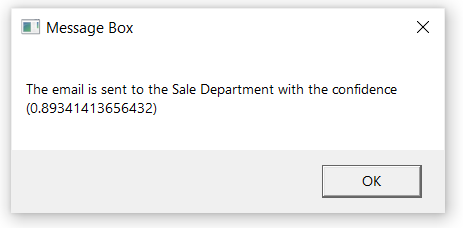
To start the process, the RPA system will read the subject and content of the email to be checked and display a notification box as shown in Figure 12.



**Figure 12. Reading email content and notification (using UiPath)**

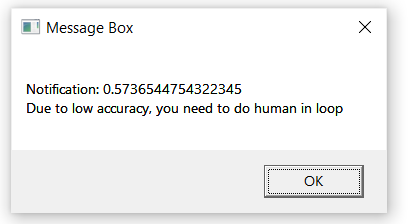
*Source: Authors*

The reliability of each department will be predicted and displayed in this dialogue box. Figure 13 shows that if the reliability is less than 0,782; the user will receive a notification to review and process the email. Then, the bot will operate in human in loop mode.



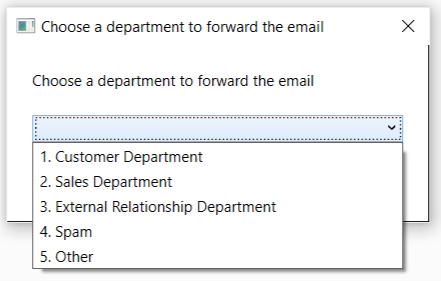
**Figure 13. Email notifications with reliability higher than 0,782 will be forwarded to the classified department**

*Source: Authors*

****

**Figure 14. Email notifications with reliability less than 0,782 will proceed to the “human in loop” step**

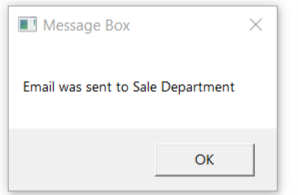
*Source: Authors*

****

**Figure 15. The operation requires human intervention "human in loop"**

*Source: Authors*

If the "Other" option is selected, the user will be prompted to enter an email address to send to. After completing the email classification and forwarding process, the system will notify the administrator.

****

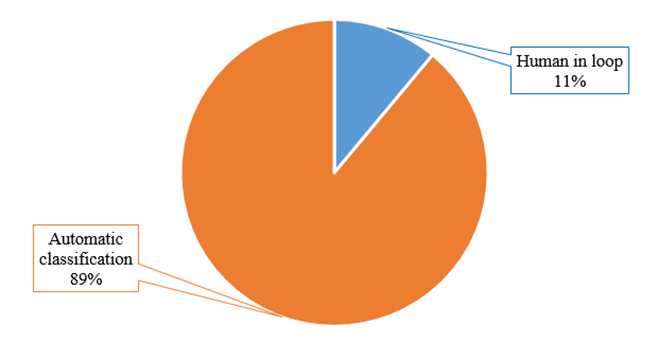
**Figure 16. Notification that email forwarding is complete**

*Source: Authors*

Once the email has been successfully forwarded, the system will store the email for easy retrieval and use when necessary.

Testing the process, the average effectiveness of the process is 82,6% (Figure 4-18). Applying this model to a dataset (1000 emails) taken from the team's data file, the results show that 890 emails were automatically classified and 110 emails required human intervention within 3 minutes and 30 seconds. According to statistics from the ServiceNow Community website, the average time from receipt to completion of processing actions for an email is 2-5 minutes (120-300 seconds) [1]. This means that to process 1000 emails, the classifier would take up to 33 hours. In contrast, based on the processing time results of the team on 1000 emails, the system only takes an average of 2,2 seconds to automatically process an email (including those that require human intervention). This shows that the automation process helps increase work efficiency by up to 10 times.

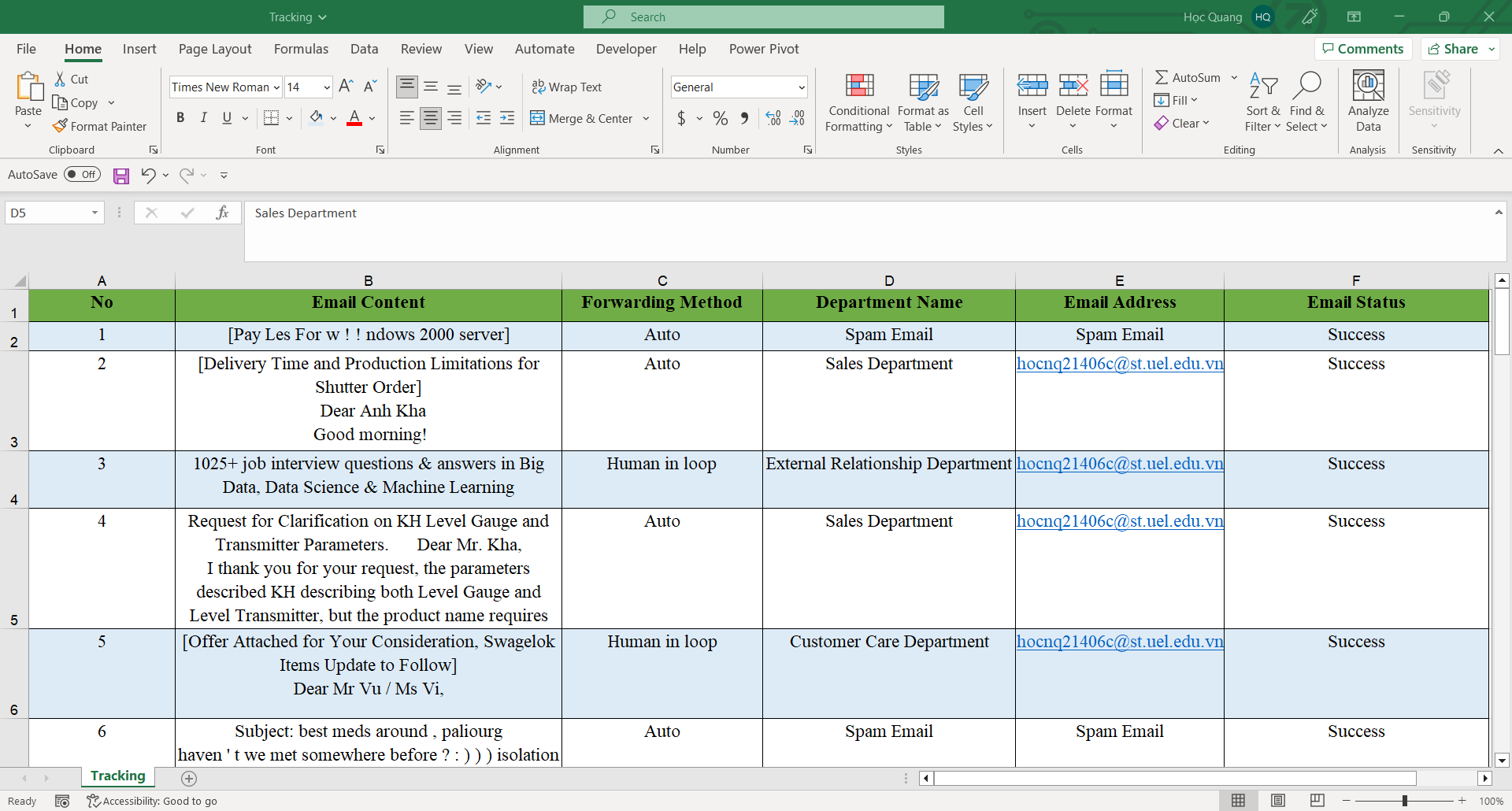
Therefore, the automation process has a fast classification speed and high accuracy compared to humans when performing the task.

****

**Figure 17. Comparison of the number of emails automatically classified and those requiring human intervention in the case of 1000 emails**

*Source: Authors*

After the email has been successfully forwarded, it will be stored in an excel file for easy tracking and can be used to supplement the training data to improve email classification results and enhance the efficiency of the forwarding process.



**Figure 18. Result of the process of storing forwarded emails**

*Source: Authors*

* 1. Discussion

Enterprises often have a single email to represent the entire company - a basic email with an address based on the enterprise's name. However, in reality, enterprises have many departments and representatives, so customers and partners are often required to contact specific emails to fasten the problem-solving process. This is the way enterprises today often apply to avoid overloading official email or avoid ambiguity in the process of receiving and processing emails. This method, although being agreed upon by customers and enterprises, still causes certain disadvantages such as customers not responding when sending emails to the wrong department, or employees of different departments having to deal with email confusion, etc. In contrast, for enterprises that only provide customers - partners with an official email, they face inadequacies in adding clerical staff. Both of these methods are wasteful of human resources and time for both enterprises and customers.

Therefore, the research team proposed an RPA bot with a fully automated email forwarding system that integrates Machine Learning - Linear SVC model with Multi-class method on OneVsOne technique. With this system, enterprises only need to provide an official representative email address for all their customers - partners. The system will replace the current manual email processing. Emails sent to the representative mailbox will be automatically forwarded to the respective department, limiting errors from both customers - partners and enterprises. This system will improve the customer experience for the enterprise as well as free up human resources for the enterprise.

In addition, the research could be developed in some directions:

1. Collecting data from multiple enterprises and other departments: In fact, enterprises often have many departments such as the Finance department, Accounting department, Human resources department, etc. depending on the field of operation. The dataset in this study includes only three departments: "Customer Care", "External Relations" and "Sales". Expanding the field and size of the dataset would bring more exact and applicable results.
2. Developing NLP to handle other languages: Expand the language range of NLP such as Vietnamese, bilingual English - Vietnamese, etc. to expand the scope of application to more enterprises.
3. Storing data on a professional system: Instead of storing data on excel, designing a database management system or cloud data storage is able to manage large data sources of the enterprise.
4. Tracking emails: Designing a system to track the status of emails such as received, forwarded, forward failed, etc.
5. Conclusion

The research comprises two fundamental components: constructing a classification method and establishing a classification process. The classification method involves NLP processing of the sample email dataset, utilizing a text classifier machine learning model in conjunction with multiclass classification techniques. The team compared four techniques (SGD Classifier, Linear SVC, Gradient Boosting Classifier, Extra Trees Classifier) combined with three techniques (OneVsOne Classifier, OneVsRestClassifier, Multi Output Classifier) on two methods (Multi-class, Multi-output) using F1 Score, Kappa, Training Time, and Testing Time indicators to select and construct the most appropriate classification model. The selected model will be trained on the NLP processed dataset, resulting in a successful construction of the Machine Learning classification method.

In the classification process, the RPA system, integrated with the Machine Learning classification method, will automatically open, read, extract, analyze, classify, evaluate, forward, and archive each email sent to the business. Moreover, using the quartile method of the sample dataset results, a confidence index of 0.782 is established to filter out emails with a low confidence index in the classification. Such emails will be notified automatically to humans for manual sorting.

The system contributes to the global trend of digital transformation and automation, optimizing enterprise resources, and enhancing customer experience. Due to its completeness, high applicability, and feasible development direction, the system has immense potential and can be directly applied to enterprise.

1. Appendix

***Appendix A. Data of Song Linh Trading and Service Co***

[***RPA\_Emails/DATASET\_SONGLINH.csv at main · NCKH-RPA/RPA\_Emails · GitHub***](https://github.com/NCKH-RPA/RPA_Emails/blob/main/DATASET_SONGLINH.csv)

***Appendix B. Enron company spam data***

[RPA\_Emails/DATASET\_ENRON.csv at main · NCKH-RPA/RPA\_Emails · GitHub](https://github.com/NCKH-RPA/RPA_Emails/blob/main/DATASET_ENRON.csv)

# REFERENCES

1. Abhishek, L. (2020). *Optical character recognition using ensemble of SVM, MLP and extra trees classifier.* Paper presented at the 2020 International Conference for Emerging Technology (INCET).
2. Abraham, A., Dutta, P., Mandal, J. K., Bhattacharya, A., & Dutta, S. (2018). *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2018, Volume 3* (Vol. 755): Springer.
3. Alghoul, A., Al Ajrami, S., Al Jarousha, G., Harb, G., & Abu-Naser, S. S. (2018). Email classification using artificial neural network.
4. Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is machine learning? A primer for the epidemiologist. *American journal of epidemiology, 188*(12), 2222-2239.
5. Bowd, C., Belghith, A., Proudfoot, J. A., Zangwill, L. M., Christopher, M., Goldbaum, M. H., . . . Weinreb, R. N. (2020). Gradient-boosting classifiers combining vessel density and tissue thickness measurements for classifying early to moderate glaucoma. *American journal of ophthalmology, 217*, 131-139.
6. Chakrabarty, N., Kundu, T., Dandapat, S., Sarkar, A., & Kole, D. K. (2019). *Flight arrival delay prediction using gradient boosting classifier.* Paper presented at the Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2018, Volume 2.
7. Chauhan, V. K., Dahiya, K., & Sharma, A. (2019). Problem formulations and solvers in linear SVM: a review. *Artificial Intelligence Review, 52*(2), 803-855.
8. Dalal, M. K., & Zaveri, M. A. (2011). Automatic text classification: a technical review. *International Journal of Computer Applications, 28*(2), 37-40.
9. Deepa, N., Prabadevi, B., Maddikunta, P. K., Gadekallu, T. R., Baker, T., Khan, M. A., & Tariq, U. (2021). An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier. *The Journal of Supercomputing, 77*, 1998-2017.
10. Dhuliawala, S., Kanojia, D., & Bhattacharyya, P. (2016). *Slangnet: A wordnet like resource for english slang.* Paper presented at the Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16).
11. Duc, T. M. SPAM EMAIL FILTERING BASED ON MACHINE LEARNING.
12. Dürscheid, C., Frehner, C., Herring, S. C., Stein, D., & Virtanen, T. (2013). Email communication. *Handbooks of Pragmatics [HOPS]*(9), 35-54.
13. Fernandes, K. (2008). *On the significance of speech: How infants discover symbols and structure.* New York University,
14. Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. *Journal of emerging technologies in web intelligence, 1*(1), 60-76.
15. Kang, Y., Cai, Z., Tan, C.-W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics, 7*(2), 139-172.
16. Madakam, S., Holmukhe, R. M., & Jaiswal, D. K. (2019). The future digital work force: robotic process automation (RPA). *JISTEM-Journal of Information Systems and Technology Management, 16*.
17. Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet], 9*, 381-386.
18. McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica, 22*(3), 276-282.
19. Mejova, Y. (2009). Sentiment analysis: An overview. *University of Iowa, Computer Science Department*.
20. Mohammadi, M., Malekian, K., Nosrati, M., & Karimi, R. (2013). Email Marketing as a Popular Type of Small Business Advertisement: A Short Review. *Australian journal of basic and applied sciences, 7*(4), 786-790.
21. Mujtaba, G., Shuib, L., Raj, R. G., Majeed, N., & Al-Garadi, M. A. (2017). Email classification research trends: review and open issues. *IEEE Access, 5*, 9044-9064.
22. Pahwa, B., Taruna, S., & Kasliwal, N. (2018). Sentiment analysis-strategy for text pre-processing. *Int. J. Comput. Appl, 180*(34), 15-18.
23. Patel, M., Shukla, A., Porwal, R., & Kotecha, R. (2019). *Customized Automated Email Response Bot Using Machine Learning and Robotic Process Automation.* Paper presented at the 2nd International Conference on Advances in Science & Technology (ICAST).
24. Sampson, G. (2005). *The'Language Instinct'Debate: Revised Edition*: A&C Black.
25. Shah, R., Lahoti, S., & Lavanya, K. (2017). An intelligent chat-bot using natural language processing. *International Journal of Engineering Research, 6*(5), 281-286.
26. Toàn, H. P., Lâm, N. V., Nghị, Đ. T., & Trung, N. M. (2011). PHÂN LOẠI THƯ RÁC VỚI GIẢI THUẬT BOOSTING CÂY QUYẾT ĐỊNH NGẪU NHIÊN XIÊN PHÂN ĐƠN GIẢN. *Tạp chí Khoa học Trường Đại học Cần Thơ*(19b), 1-9.
27. Zong, Z., & Hong, C. (2018). *On application of natural language processing in machine translation.* Paper presented at the 2018 3rd International Conference on Mechanical, Control and Computer Engineering (ICMCCE).

1. Corresponding author: Nguyen Quang Hoc; Tel: +84 862124950; Email: [hocnq21406c@st.uel.edu.vn](mailto:hocnq21406c@st.uel.edu.vn) [↑](#footnote-ref-1)
2. Mentor: Nguyen Thon Da; Tel: +84 906230232; Email: [dant@uel.edu.vn](mailto:dant@uel.edu.vn) [↑](#footnote-ref-2)