

Web Based Application of Examination Question Classification

BY

TAN GUAN KHENG

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF INFORMATION SYSTEMS (HONS)

Business Information System

Faculty of Information and Communication Technology

(Kampar Campus)

MAY 2019

REPORT STATUS DECLARATION FORM

Title: _____

Academic Session: _____

I _____
(CAPITAL LETTER)

declare that I allow this Final Year Project Report to be kept in
Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

(Author's signature)

(Supervisor's signature)

Address:

Supervisor's name

Date: _____

Date: _____

Web Based Application of Examination Question Classification

BY

TAN GUAN KHENG

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF INFORMATION SYSTEMS (HONS)

Business Information System

Faculty of Information and Communication Technology

(Kampar Campus)

MAY 2019

DECLARATION OF ORIGINALITY

I declare that this report entitled “**METHODOLOGY, CONCEPT AND DESIGN OF WEB BASED APPLICATION OF EXAMINATION QUESTION CLASSIFICATION**” is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : _____

Name : _____

Date : _____

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to my supervisors, Mr Anbuselvan a/l Sangodiah who has given me this bright opportunity to engage in data mining project. A million thanks to you. The patience and advises given by him has created a motivation for me to accomplish this study. Finally, I must say thanks to my parents and my family for their love, support and continuous encouragement throughout the course.

Abstract

Examination can be considered as the most effective way to evaluate the student's performance on certain subjects. Normally, a standard examination paper will consist of six different categories of Bloom Taxonomy questions. Classifying questions based on Bloom Taxonomy cognitive will improve the overall quality of examination paper as Bloom Taxonomy was invented to promote higher levels of thinking of students in education by building up from lower level cognitive skills. However, it maybe inefficient and time consuming for lecturers to classify and label the questions manually since such application that exist yet to help the lecturer to classify and label the question automatically. Hence, this project proposed a web based examination question classification application that will classify and label the difficulty of question automatically by using machine learning and Python programming languages. In this project, supervised machine-learning model, Support Vector Machine was used in the web application by training it with a data set consisting of questions predefined with labels or categories of BT and difficulties to classify the unseen examination questions into appropriate level or category of BT and respective difficulties. Overall, the web application is very useful and convenient for lecturer and student since they just need to browse the website in order to categorize the examination questions.

TABLE OF CONTENTS

REPORT STATUS DECLARATION FORM	i
TITLE PAGE	ii
DECLARATION OF ORIGINALITY	iii
ACKNOWLEDGEMENTS	iv
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	viii
LIST OF TABLES	ix
LIST OF ABBREVIATIONS	x
CHAPTER 1 INTRODUCTION	1
1.1 Problem Statement	1
1.2 Background Information	2
1.3 Project Objectives	4
1.4 Project Scope	5
1.5 Impact, Significance and Contributions	6
1.5.1 Impact and Contributions	6
1.5.2 Significance	6
CHAPTER 2 LITERATURE REVIEW	7
2.1 Literature Review	7
2.1.1 Theory Review	7
2.1.2 Product Review	9
2.2 Critical Remarks of Previous Works	13
CHAPTER 3 SYSTEM DESIGN	14
3.1 Training of SVM Classifier	14
3.2 Design of Web Question Classification Tool	18
3.3 Use Case of Web Based Examination Question Classification	21

CHAPTER 4 EVALUATION	22
4.1 k-fold Cross Validation Method	22
4.2 Evaluate with dataset from same subject	23
4.3 Evaluate with Real Data Type	24
4.4 Graphical User Interface	33
CHAPTER 5 CONCLUSION	38
5.1 Implementation Issue and Challenge	39
5.2 Future Work	39
BIBLIOGRAPHY	40
APPENDIX A TRAINING DATASET	A-1
APPENDIX B POSTER	B-1
PLAGIARISM CHECK RESULT	
CHECK LISTS	

LIST OF FIGURES

Figure Number	Title	Page
Figure 1.2.1	Bloom Taxonomy	3
Figure 3.1.1	Training of SVM Classifier	14
Figure 3.1.2	Web Based Question Classification Tool Design	18
Figure 3.1.3	Use Case Diagram	14
Figure 4.4.1	Home Page	33
Figure 4.4.2	Navigation Bar	33
Figure 4.4.3	Excel Page	34
Figure 4.4.4	Choose File	34
Figure 4.4.5	Upload	35
Figure 4.4.6	Attachment	35
Figure 4.4.7	Sample Input	35
Figure 4.4.8	Sample Output	36
Figure 4.4.9	Question Page	36
Figure 4.4.10	Sample Input of Question	37
Figure 4.4.11	Sample Output of Question	37

LIST OF TABLES

Table Number	Title	Page
Table 2.3.1	Comparison between previous works	13
Table 3.1.1	Sample of Pre-processed Data	16
Table 3.2.1	Sample of Pre-processed input	20
Table 4.1.1	5-fold Cross Validation (SVM)	22
Table 4.1.2	5-fold Cross Validation (Decision Tree)	22
Table 4.1.3	5-fold Cross Validation (Naïve Bayes)	22
Table 4.2.1	5-fold Cross Validation (SVM)	23
Table 4.2.2	5-fold Cross Validation (SVM)	23
Table 4.3.1	Knowledge Level Result	25
Table 4.3.2	Comprehension Level Result	26
Table 4.3.3	Application Level Result	27
Table 4.3.4	Analysis Level Result	28
Table 4.3.5	Evaluation Level Result	29
Table 4.3.6	Synthesis Level Result	30
Table 4.3.7	Comparison and Total Accuracy	32

LIST OF ABBREVIATIONS

<i>DM</i>	Data Mining
<i>BT</i>	Bloom's Taxonomy
<i>QA</i>	Question Answering
<i>IR</i>	Information Retrieval
<i>SVM</i>	Support Vector Machine
<i>FS</i>	Feature Selection
<i>BoW</i>	Bag-of-Words
<i>ML</i>	Machine Learning

Chapter 1 Introduction

1.1 Problem Statement

Examination is defined as a formal test that we need to take in order to test our knowledge, skill and understanding on certain subjects. Examination can be considered as the most effective way to evaluate the student's performance on certain subjects. Normally, a standard examination paper nowadays will be designed according to Bloom Taxonomy's concept which consist of six categories of questions that will be knowledge, comprehension, application, analysis, evaluation and synthesis. The level of difficulty will increase start from knowledge to synthesis. However, the problem arises when the student and lecturer want to classify the difficulty of questions manually. According to Yusof and Chai cited in Omar et al. (2011), normally all the lecturers have the knowledge and skills to classify the question difficulty manually, but not all of them can identify the question difficulty correctly. Even the lecturers have the problems in classifying the question, not to mention the students because mostly all the students don't have adequate skills and knowledge about the question difficulty. In addition, even some lecturers are able to classify the questions manually, but it could be inefficient and very time consuming when there are huge amount of questions that need to be classified.

It is good for both student and lecturer if there is a system exists that will help the lecturer and student to classify the examination questions automatically just by inputting a list of examination questions. According to Sangodiah et al. (2017), there are some question classifier system that exists at the past but they have poor performance for the questions that come from different areas or domains. Furthermore, there are few researchers (Sangodiah et al. 2017; Fattah et al. 2015; Omar et al. 2011) who had done research on how to use the question classifiers to classify the question based on cognitive level of Bloom Taxonomy. However, all of the research done are only limited to the theory level. There are no such complete system exists yet to help the student and lecturer to classify the exam question from different areas based on difficulty. Hence, a web based application that can classify the examination questions based on Bloom Taxonomy concept accurately will be developed in this paper.

1.2 Background Information

In order to understand how this system works, there are few terms that need to be understood first. There are text mining, text classification, supervised learning, question classification and Bloom Taxonomy. According to Hearst (2003, p1), ‘Text Mining is the discovery by computer of new, previously unknown information by automatically extracting information from different written resources. A key element is the linking together of the extracted information together to form new facts or new hypotheses to be explored further by more conventional means of experimentation.’ Determine the unknown information that are not found by others and not yet written will be the goal of the text mining. Text mining actually is a subset of data mining that will work to discover the pattern or trend based on historical data. (Hearst, 2003). For instance, Econsave maybe found that when a customer buy a beer, they will also have high chance to buy a diaper. This is the interesting pattern that the text mining or data mining want to look for from the databases.

Text mining normally have three processes such as text clustering and classification, information extraction and retrieval and also trend detection. In this paper, we need to understand more about the text classification compared to other processes. According to Sriram (2010), ‘Classification is a supervised data mining technique that involves assigning a label to set of unlabelled input objects.’ Han (2006) stated that “supervised learning is a learning that the training data are accompanied with labels indicating the class of observations. For supervised learning, the new data is classified based on training set.” Basically, text classification is using supervised learning method to create a model using large amount of existing data with the predefined labels and categories. The model will be trained and based on the likelihood inferred, later it will use the model to label the other unlabelled documents (Cao et al. cited in Sangodiah, A et al., 2017).

Question classification is the process which a set of question will be analysed and labelled by the system based on the predefined category. (Panicker et al., 2012) To classify text documents, text classification will have a higher accuracy. However, it will be a very difficult and diligent task to classify the exam classification which called question classification (Sangodiah et al., 2017). Question classification will work in similar ways as the text classification. However, a unique challenge which request a separate study of its own is facing by the former. When classifying the questions, only little information or words can be used due to the shorter length of questions compared to documents' content (Xin and Dan cited in Sangodiah et al., 2017).

For the question classification to work well and efficient, Bloom Taxonomy guideline is implemented. And this implementation of Bloom Taxonomy has received much attention from others. (Sangodiah et al., 2017) According to Fattah et al (n.d., p204), 'Bloom's taxonomy is a classification system of educational objectives based on the level of student understanding necessary for achievement or mastery.' Bloom Taxonomy consists of 6 levels starting from the simplest, knowledge, followed by comprehension, application, analysis, synthesis and the most difficult, evaluation.

The New Version of Bloom's Taxonomy

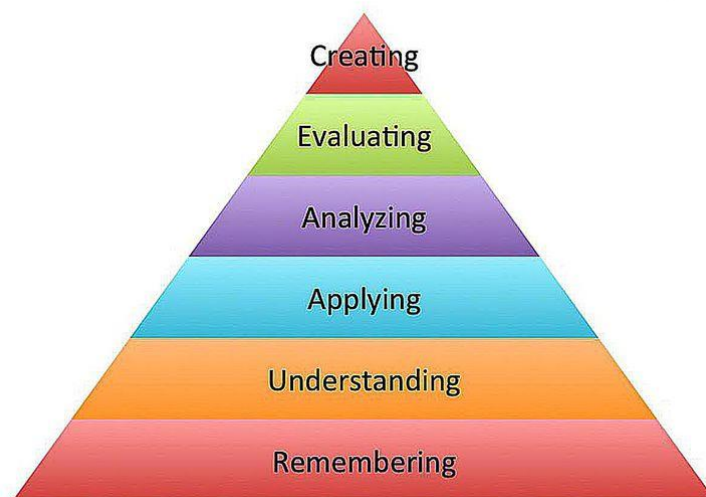


Figure 1.2.1 Bloom Taxonomy (Melissa, 2019)

1.3 Project Objectives

In this project, there are three objectives that need to be achieved:

1. To analyse the research work of existing question classification based on BT

In this project, previous research work of classifying question based on BT will be studied and review in order to get a better understanding that will aid in developing a question classification application with satisfied accuracy successfully.

2. To develop a web based application that apply the BT's concept and Support Vector Machine Classifier.

In this project, a web application that will analyse the inputted examination question and classifies them based on Bloom Taxonomy's categories and reaches approximately 70% of accuracy using Natural Processing Technique, Machine Learning and Support Vector Machine classifier.

3. To evaluate the developed web based application.

In this project, Support Vector Machine will be trained by a data set which consists of examination questions that are labelled with Bloom Taxonomy's level in order to classify the new examination questions into correct level of Bloom Taxonomy. The performance of model trained will be evaluated by comparing the correct label and predicted label by the model whether an accuracy of at least 70% will be achieved or not.

1.4 Project Scope

In this project, a web based application that can classify the examination question based on BT level will be developed and it is only for the education use. This application is mainly developed for the lecturers and students. The application is useful for the lecturers in helping them to build a well-structured exam questions based on the cognitive level of Bloom Taxonomy. While for the students, this application can help them to clearly understand how the structure of the examination questions look like and they can be well prepared for the difficult questions. Support Vector Machine will be the classification techniques that used to classify the questions in this project. Questions that include the mathematical operations, diagram or figures will not be able to be classified by the application. Only text based documents such as practical, tutorial, midterm or past year examination questions will be analysed and classified in the application. Furthermore, the application will only work for the exam questions that contain Bloom Taxonomy's (BT) keywords. In other words, only BT questions will be successfully classified in this application. Moreover, the web application will be built by using Python as the main programming language and Django as the development framework. The software used will be Python IDE and Sublime Text Editor. The concepts used will be Natural Language Processing (NLP), Question Classification and Bloom Taxonomy.

1.5 Impact, Significance and Contributions

1.5.1 Impact and Contributions

The project will develop a web based examination question classification application which can benefit the lecturers and students in the education sector. When a lecturer finish designing a set of new exam questions, they can easily evaluate the questions whether the new exam questions meet the BT standard or not. For students, before they take a new course in next semester, they can easily know that whether they are able to handle the course or not by viewing the past year exam questions first. Besides that, by having this system, student can easily visualize how will the structure of future final exam, midterm and quiz look like and get well prepared for the assessment.

1.5.2 Significance

Manual classification of examination questions will be inefficient and time consuming when the lecturers have large amount of questions need to be classified. By having this application, it can save a lot of times for lecturers by reduce the need of them to classify the exam question manually since the application will do for them. Furthermore, this application will also help the lecturers who do not have adequate understanding about the BT to design a good examination paper accordance to BT level and further improve the student academic performance.

Chapter 2 Literature Review

2.1 Theory Review

There are several techniques that exist that can be used in question classification such as Artificial Neural Network (ANN), Support Vector Machine (SVM), Naïve Bayes, Rule Based Classifier and Nearest Neighbor (NN). According to Sangodiah et al. (2017), ANN will automatically classify the questions into three difficulty levels which are hard, medium and easy. Five features which are query-text relevance, mean term frequency, length of Q&A, term frequency and distribution of Q&A in text had been considered by the technique. 78% of accuracy were achieved when using this technique. The limitation of this approach will be lacking of semantic analysis and only working on small set of data. According to Yusof cited in Sangodiah (2017), Bloom taxonomy has been applied on Artificial Neural Network classifier to classify questions. Feature reduction methods are very focused on by researchers to reduce the dimensionality of the feature space. Based on Bloom taxonomy, good performance in terms of speed of classifying questions are found in the reduction methods. Furthermore, bag-of-words concept is applied in the feature type used in the research.

Support Vector Machine (SVM) will be one of the high accurate machine learning method for question classification. An optimal hyperplane within the input space will be found by SVM to correctly classify the binary classification problem. SVM is less vulnerable to other technique. Best result for both testing and training data set will be produced by SVM. However, SVM will be more difficult to implement and perform well in large collection of documents (Shrihari, C & Desai, A, 2015). SVM with bag-of-word feature type to classify questions based on Bloom taxonomy that used by researcher showed a good performance in terms of accuracy and precision (Yahya and Osman cited in Sangodiah et al., 2017). Some linguistic feature types such as POS Tagging, unigram, bigram, trigram, POS Bigram and Trigram perform with several classifiers are investigated by the researchers. The best result obtained is using SVM with unigram. The linguistic features can improve classification accuracy are shown (Addin and Osman cited in Sangodiah et al., 2017).

Naïve Bayes is one of the most common approaches for question classification. It has been used widely in many areas that includes text and question classification. The main possibility, which is the word connected to the named entity classes will be pulled by the algorithm of Naïve Bayes with the help of feature vector table (Abdulhadi, 2015). The main advantages of Naïve Bayes will be it is very efficient and easy to learn compared to other machine learning method. (Kim, SB et al., 2006). According to Abdulhadi (2015), Naïve Bayes is easy to implement and have very good performance on solving text classification problems. One of the limitation of Naïve Bayes will be it not perform well for the short text classification (Yuan, Q et al., 2012). According to Kim et al. (2006), the limitation will be the rough parameter estimations. Parameter estimation is affected more by long documents compared to short documents. Next limitation will be difficulty in handling rare categories that contain only a few training documents. If there is insufficient amount of training examples or set data, NB will not work well. This statement can be proved by Yadav, R & Mishra, M, when they trained Naïve Bayes classifiers with 5500 of training data, they achieved 83.2% of accuracy, but the accuracy dropped to merely 53.8% when they trained with only 1000 of training data.

According to Haris and Omar cited in Sangodiah et al. (2017), the question can be classified by using rule based approach. “Rule-based classifier classifies the question in straightforward method using a set of predefined heuristic rules based on taxonomy. “ (Jayalakshmi, 2015). Jayalakshmi (2015) stated that the questions are classified by the approaches using rules that are manually created by the experts. However, a lot of rules are needed to improve in classifying for this technique although the work shows some reasonable findings. Furthermore, other domains or different language will not be supported by the rule based approach since it is difficult to create a new set of rules. Another limitation of rule-based will be it is not perform well in the new dataset and not accessible to a large number of questions (Jayalakshmi, 2015). According to Omar et al. cited in Sangodiah et al. (2017), in order to categorize question based on Bloom taxonomy, rule based system was proposed to capture sequence patterns.

The last technique reviewed will be the K-Nearest Neighbors (KNN) algorithm. According to Chen et al. (2014), 'KNN is an extension of the nearest neighbour algorithm. Based on the thinking of nearest neighbour, K nearest neighbours of the test samples are selected, and the type of K new nearest samples can be determined.' Chen et al. (2014) stated that KNN main advantage are widely used in the classification field and the simplicity of the principle of KNN to be understand by people. However, there are two main issues that KNN need to solve in order to improve its classification performance. How to determine the nearest neighbour number K of samples to be classified will be the primary issue that need to be solved. The next issue will be the need of calculation of the distance between each classified sample and the training sample. The classification performance will be affected when the classification sample data is too large because the complexity will increase. (Chen et al., 2014)

2.2 Product Review

For the system, there are several of them are similar to the exam question classification system. The first system is Information Retrieval System (IR). Due to there are huge amount of data generated daily, information is not a shortage nowadays. Nevertheless, there are only some of the available information will be useful for specific information need. Information Retrieval system will provide users with the subset of available information relevant to their needs (Micheal, 1998). Micheal (1998) also stated that unstructured information such as text, image, audio, video and others will be retrieved by using IR system. Extra techniques such as relevance feedback, unsupervised query expansion and structured queries will be implemented to improve the IR system effectiveness. A language modelling approach to IR that enable for a simple formulation of problem is proposed by (Micheal, 1998).

The next system will be Information Retrieval Agent. According to Marko (1995), discrete cycle will be run by the Information Retrieval Agent system. What the IR agent can perform will be the system will use some search heuristic to search the web, taking some time. Next, the best pages will be presented to the user by using some selection heuristic. Evaluation from user for each page presented will be received by the system. According to the feedback, search and selection heuristics will be updated. Marko (1995) also stated that before they evaluate the search heuristic, they will assign a word weights to the words in the documents using TFIDF scheme. By using this measure, the best keywords with the highest weights can be retrieved from each document to represent the words that user are interested to search. The future work will be an improved version that will allow communication and cooperation between instances of the system running on behalf of different users. (Marko, 1995)

A set of programs composing a fully automatic document retrieval system is defined as SMART information retrieval package. Before retrieving the document, a large collection of online documents such as online article, journal article, online newspaper and others will be automatically indexed with a document representative. The representative contain of a list of important concepts that found in the document, and each of the concepts will be given a weight. If the weight is higher, this mean that the concept is more important. Hence, when the user come to the SMART system and input their desired documents which also called as query representative, the SMART system will calculate the similarity of the query representative with the document representative and order the result with the highest to lowest similarity. Future improvement of this system will be the algorithms used in the implementation will be improved. The algorithm with higher efficiency will be implemented. (Chris, 1985)

Another type of similar system rather than information retrieval system will be question answering system. In other words, “Question answering is a difficult form of information retrieval characterised by information needs that are at least somewhat expressed as natural language statements or questions, and was used as one of the most natural type of human computer communication”. (Gupta, P & Gupta, V, 2012) Question answering system will be able to find out the correct answer of the questions by given a set of documents. Information technology, artificial intelligence, natural language processing, knowledge and database management and cognitive science are involved in question answering. Question answering system normally has question classification, information retrieval, and answer extraction as the components. While the answer extraction will be the most difficult and challenging among these three components due to the need of the system to validate and rank the user’s answer. (Gupta, P & Gupta, V, 2012) Research has shown that correct and accurate way of question classification will lead the question answering system to be successful (Sundblad, 2007).

Student Question Classification is a system that will automatically classify the student questions by matching the questions to previous one with similar meanings but different phrase to provide automated answer for the challenging questions. There are some researcher work on how to reduce the natural language of the questions to a vector space, and cosine similarity is utilized to identify similar previous questions. The system will be trained by using the questions that normally a student will ask. The system will tokenized the questions based on spaces and special keywords. The stop words such as you, me, the etc. were removed to leave a word stems. For instance, “extracting” become “extract”. Classification accuracy of the questions will be approximately 55%. (Heiner, C & Joseph, LZ, 2009) According to Heiner & Joseph (2009), the future work will be “first, the existing corpus could be leveraged as a starting point for designing common error detectors and appropriate interventions. Second, a human TA could supervise the classification algorithm, and override any incorrect decisions that it makes, until the number of incorrect decisions decreases.”

According to Douglas cited in Prasad et al. (n.d.), qa42 is an open domain question answering system that builds to return specific answers to factoid questions rather than summaries that are done by traditional QA system. Questions that are related to persons, organizations, locations, dates and quantities will only answered by qa42. First, the system will undergo parse analysis to categorize the questions into groups of noun, verb and subject. Qa42 will generate a list of query template and for each query template, it will generate a list of answer models for the purpose of predicting and scoring summaries. The query template will be send to google search engine and the returned summaries will be scored. The three answers with the highest score will be presented by qa42. (Prasad, J et al., n.d.)

“Jikitou is a question answering system that can provide an improved method for information retrieval that deals with specific biological questions.” Answers related to biological questions will be returned rather than lists of documents, which can reduces the need of user to go through the entire document (Micheal, 2013). Knowledge base, question analysis, answer agents and user interface will be the subsystems of the system architecture of Jikitou. Possible answers to the questions were found by the multiple software agents and the most appropriate answers are displayed to the user at the user interface. And the system will prompt a dialog that will require the user to give their feedback to refine the query. The system will marked up the answer automatically and linked to semantically related content in other databases. (Micheal, 2013)

2.3 Critical Remarks of Previous Works

Comparison System	Strength	Limitation
IR agent (Marko, 1995)	i) Burden of user will be reduced when the system will help to filter a stream of incoming information.	i) No search query provided by the user, no indexing of the collection of documents and longer periods of time is needed for the operation.
SMART IR (Chris, 1985)	i) A flexible experimental system for research in IR will be provided. ii) A fast, portable and interactive environment for actual users is provided.	i) Problems of user interface and concurrency.
Student Question Classification (Heiner, C & Joseph, LZ, 2009)	i) Quite good for student who desperately need the answer about the assignments or something.	i) System may not answer the student correctly. ii) Need to resolve the usability issues between the teachers and the students in order to deploy the system.
Qa42 (Prasad, J et al., n.d.)	i) Specific answer is returned to factoid questions rather than summaries.	i) Answers that fall into a handful of entity types can only be returned by qa42.
Jikitou (Micheal, 2013)	i) Multiple natural language processing techniques, data resources and technologies are combined to create a unique system to help researchers or students to understand the biomedical corpus.	i) Knowledge based data will be quickly outdated. ii) Periodically update of knowledge based data is required.
Web based Exam Question Classification	i) Able to categorize and label the difficulty of the BT questions.	i) Not working for the non-BT, graphic and mathematical questions.

Table 2.3.1 Comparison between previous works

Chapter 3 System Design

3.1 Training of SVM Classifier

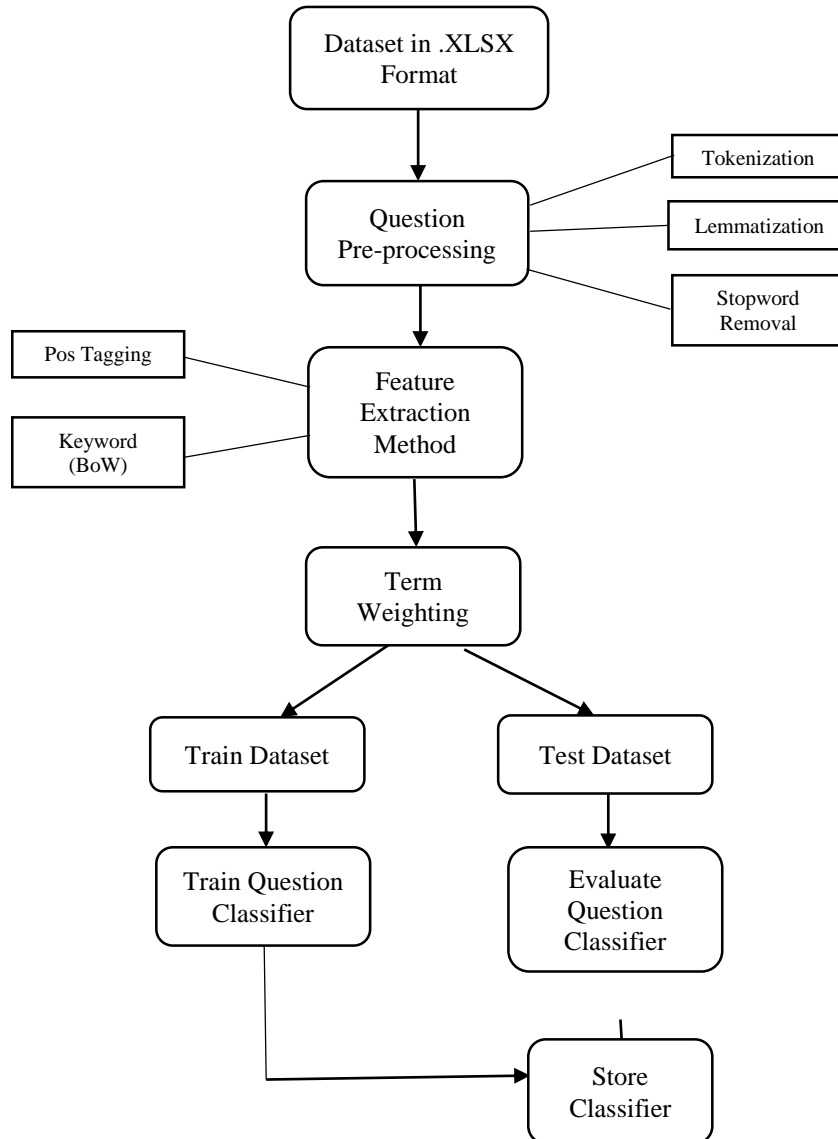


Figure 3.1.1 Training of SVM Classifier

Dataset

Two set of sample data will be prepared and stored in excel (.xlsx) format. The first dataset consists of 163 questions that had been labelled with MOTS level and BT level. This questions in this dataset are collected from different subjects in business domain. The second dataset consists of 95 questions that had been labelled with MOTS level and BT level but those questions are collected from a same subject in the business domain. All the labels are correct and verified by the educational experts. However, only BT level will be used as the predefined label and MOTS level label will just be ignored as the training process proceeds.

Question Pre-processing

The dataset will be imported into the web application and loaded by using openpyxl function. Openpyxl is one of the library that can read and write the contents in the excel file. Firstly, the first character of every questions will be converted to lower-case character because the upper case character will affect the result of lemmatization later. Special noun such as a person's name, a place, a brand or other words with all upper-case characters will be removed from the questions. Then, the application will start to remove the punctuation characters since the punctuation characters are not important in determining the BT level. Next, the application will start to remove the digits such as "1, 2, 3..." in the questions because digits will decrease the accuracy of the classified results. Next stage will be tokenization process. In this process, the questions will be broken down into separate pieces that called tokens and stored in a new created list. The following stage will be lemmatization stage.

In this process, WordNetLemmatizer and Part-of-Speech (PoS) tagging from NLTK library will be imported and used. WordNetLemmatizer will convert every token in the list to its original form by inputting the form of the token. And with the help of PoS tagging, the form of the word whether it is verb, noun, adverb or adjective will be determined. For instance, the result of word, "presenting", from PoS tagging will be "VB" which represents the verb. Hence, when lemmatizing the "presenting", a result of "present" will be get eventually. Move to next stage, the number texts such as "one, two, three ..." will also be removed from the list. Lastly, stopword such as "a", "the", "an" and others will also be removed from the list while there are some stopword such as "how", "why" or "which" will be keep because they considered as BT keywords.

Feature Selection

After the question pre-processing, bag-of-words model will be used to create a feature set. Every token in the list will be joined to form a vocabulary list. For an example, question 1 = ["explain", "advantage", "read"] and question 2 = ["define", "brand", "audit"]. The feature set extracted will be ["explain", "advantage", "read", "define", "brand", "audit"]. After the looping of all the questions in the list, a final feature set that consists of every token of all questions will be extracted. Then, the feature set also called vocabulary list will be sorted and set to make sure that there is no any duplicated token within the list.

Term Weighting

Next, the application will import the vocabulary list and the pre-processed question in list form to a function. The function will score each token in each question based on the vocabulary list. The purpose of this function is to convert the text into vector so that the machine can understand it. If the word presents in the vocabulary list, 1 will be scored and 0 will be scored if the word is absent.

Feature Set	advantage	audit	brand	define	explain	read	BT Level
Question							
explain advantage read	1	0	0	0	1	1	Comprehension
define brand audit	0	1	1	1	0	0	Knowledge

Table 3.1.1 Sample of Pre-processed Data

The final output of this stage will be a preprocessed list: [[1,0,0,0,1,1], [0,1,1,1,0,0]].

Evaluate Question Classifier

In this stage, the preprocessed list from the previous stage will be split into training dataset and test dataset by using `train_test_split` function from sklearn library. The function will split the dataset into training dataset and test dataset with the ratio of 7:3. Support Vector Machine (SVM), Decision Tree and Naïve Bayes classifier will be imported along with the sklearn library. The training dataset will be fitted into the model created to train the model. The trained model will be evaluated by k-fold cross validation method in order to measure the accuracy and performance of the trained model among those classifiers.

Store Classifier

Lastly, the trained model with satisfied performance and accuracy will be stored into a file with “.sav” format by using pickle function. So, this will avoid the necessity of user to train the model every time when they use the web application. Hence, the performance of the web application will be much faster and resource-efficient since the stored classifier can be imported to classify the question directly.

3.2 Design of Web Question Classification Tool

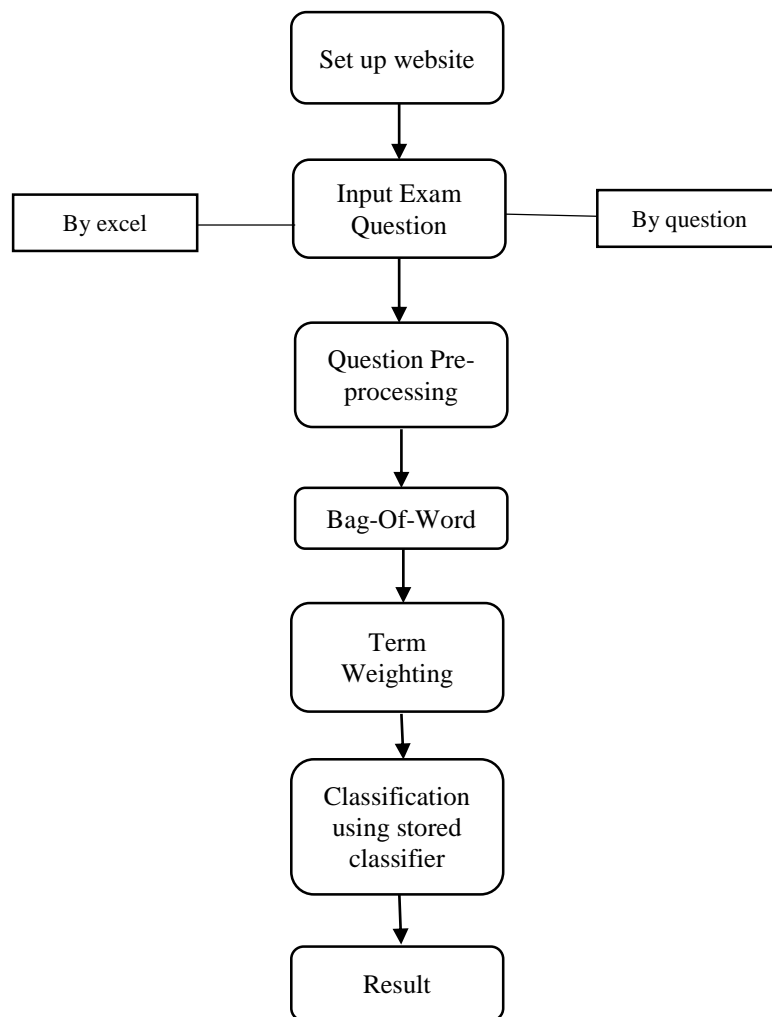


Figure 3.2.1 Web Based Question Classification Tool Design

Set up Website

In order to set up a website, Django framework is used in this project. This website consists of 3 pages which are the home page, question page and excel page. Home page is a page that provides some general introduction about this web page. Question page is the page that enable the user to classify their question's BT level but only limited to one question per time. The excel page will enable the user to classify multiple questions per time by inputting their questions in excel (.xlsx) format. After finished the design of the web page such as form, bootstrap, layout and some instruction in text format, the web page will be hosting at Python Anywhere which is an online integrated development environment and web hosting service based on the Python programming language.

Input Exam Question

The web application provides two options for the user to choose the method of inputting their questions which are by question or by excel. For question, there will be a textbox that allows the user to paste their question into it to be classified but only one question is allowed each time. For excel, user can prepare an excel file that contains a list of questions and input the excel file to be further processed.

Question Pre-processing

Before proceeds to question pre-processing stage, the question inputted by user will be checked whether the question consists of at least one BT keyword or not. If the question do not have any BT keywords, it will be skipped for excel page and for question page, an error message will pop out to inform the user this question is not a BT question. Next, the questions will be processed into a machine readable format such as [1,0,1 ...]. The web application will perform several steps such as punctuation character removal, digit number removal, tokenization, lemmatization and stopword removal.

Bag-Of-Word

The feature set that extracted previously when trained the classifier will be stored and imported into the web application. Then, the web application will compare the pre-processed words with the imported feature set to perform BoW. The word that doesn't exist in the feature set will be ignored.

Term Weighting

Next, the web application will perform term weighting for the pre-processed words, if the word present then 1 will be scored else 0 will be scored.

Feature Set Question	advantage	audit	brand	define	explain	read	BT Level
explain							
benefit	0	0	0	0	1	0	
code							

Table 3.2.1 Sample of Pre-processed input

Eventually, a result similar to Table 3.1.1 will be get but with unlabelled BT level.

Classification

Furthermore, the previous classifier that stored in .sav format will be imported. The imported model will get the input such as [1,0,0,1,1] from Table 3.2.1 and automatically label the BT level.

Result

If the user inputs by question, the label will be shown in the web instantly. If the user inputs by excel file, excel file with question and the label will be returned to the user in the form of attachment.

3.3 Use Case of Web Based Examination Question Classification

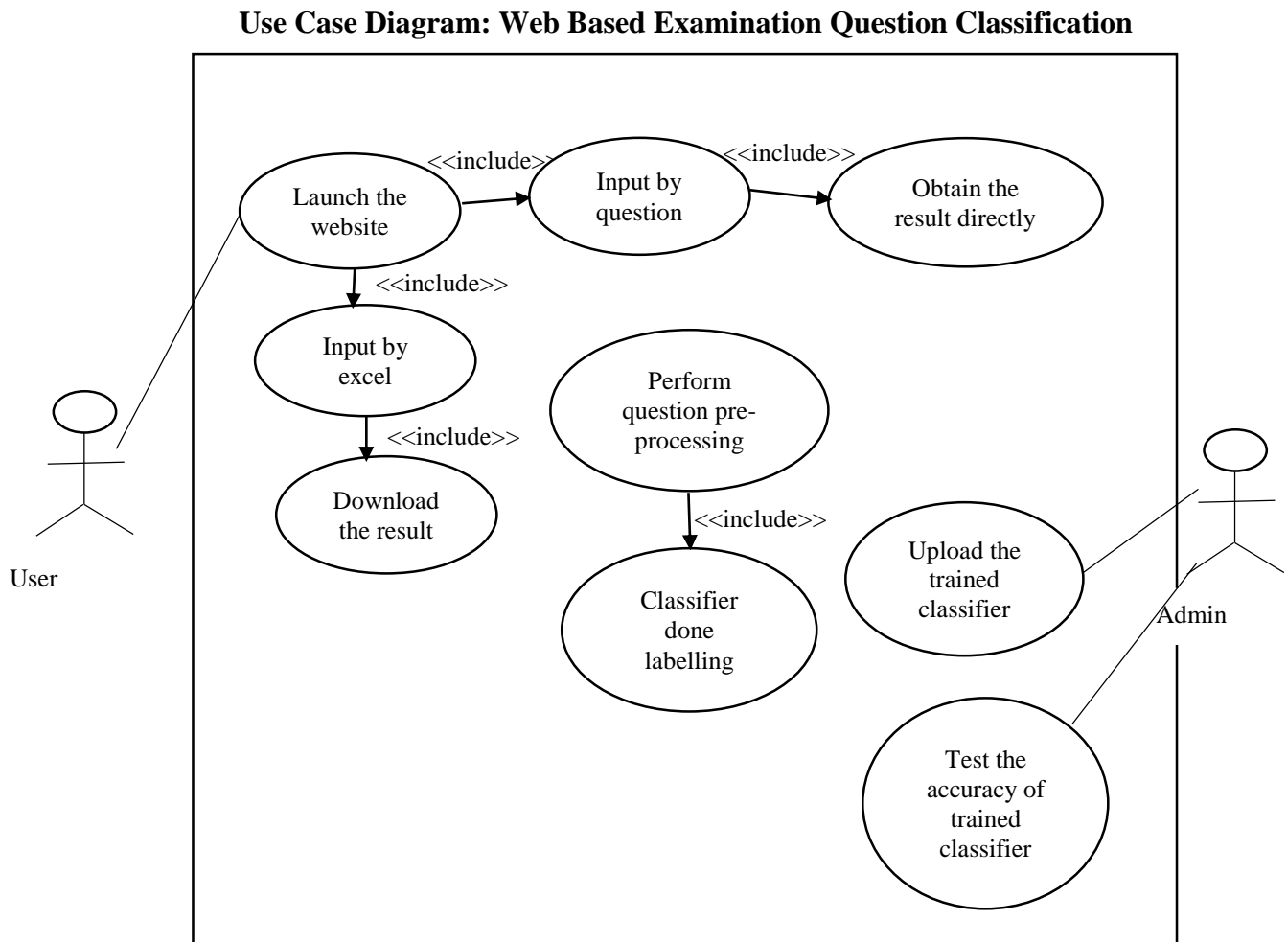


Figure 3.3.1 Use Case Diagram

The figure 3.3.1 show the use case diagram of web based examination question classification. When the user launches the website, he or she can choose to classify only one question one time or upload the excel file to be classified. User can enter the question into the textbox provided and the web application will process the question and show the result instantly. The result will be shown to user below the textbox at the web page. If the user upload the excel file which consists of a list of examination questions, the web application will process the questions in excel file and the result will be returned to the user in the form of attachments. Admin will upload the trained classifier to the web application to be used. If there is any modification to the classifier, admin will re-upload again the modified classifier with better accuracy to the web application. Admin will also test the accuracy of the trained classifier when deals with real examination questions.

Chapter 4 Evaluation

4.1 k-fold Cross Validation Method

Support Vector Machine

k	Score List	Average Score
3	[0.51785714, 0.5, 0.52830189]	0.515
4	[0.52272727, 0.53658537, 0.525, 0.60526316]	0.547
5	[0.57142857, 0.55882353, 0.63636364, 0.64516129, 0.6]	0.602

Final Mean Value = 0.555

Table 4.1.1 5-fold Cross Validation (SVM)

Decision Tree

k	Score List	Average Score
3	[0.51785714, 0.55555556, 0.39622642]	0.490
4	[0.56818182, 0.46341463, 0.625, 0.44736842]	0.526
5	[0.65714286, 0.47058824, 0.54545455, 0.64516129, 0.7]	0.603

Final Mean Value = 0.539

Table 4.1.2 5-fold Cross Validation (Decision Tree)

Naïve Bayes

k	Score List	Average Score
3	[0.375, 0.35185185, 0.43396226]	0.387
4	[0.52272727, 0.48780488, 0.45, 0.44736842]	0.477
5	[0.51428571, 0.44117647, 0.33333333,3 0.35483871, 0.46666667]	0.422

Final Mean Value = 0.429

Table 4.1.3 5-fold Cross Validation (Naïve Bayes)

In this study, three type of classifiers which were Support Vector Machine, Decision Tree and Na ïve Bayes had been selected to be evaluated in terms of score by using k-fold cross validation method. From these 3 table which are Table 4.1.1, 4.1.2 and 4.1.3, Na ïve Bayes has the lowest accuracy among these three classifiers which was only around 43%. While for the Decision Tree and Support Vector Machine, both of them have almost equal accuracy which were 54% and 56% respectively. Since Support Vector Machine has a higher accuracy compared to Decision Tree and also Na ïve Bayes, hence it will be chosen as the question classification tool in the web application.

4.2 Evaluate with dataset from same subject

Support Vector Machine and Dataset from different subjects

k	Score List	Average Score
3	[0.51785714, 0.5, 0.52830189]	0.515
4	[0.52272727, 0.53658537, 0.525, 0.60526316]	0.547
5	[0.57142857, 0.55882353, 0.63636364, 0.64516129, 0.6]	0.602

Final Mean Value = 0.555

Table 4.2.1 5-fold Cross Validation (SVM)

Support Vector Machine and Dataset from same subject

k	Score List	Average Score
3	[0.60606061, 0.68965517, 0.48148148]	0.592
4	[0.6, 0.59090909, 0.77272727, 0.7]	0.666
5	[0.65, 0.68421053, 0.66666667, 0.76470588, 0.66666667]	0.686

Final Mean Value = 0.648

Table 4.2.2 5-fold Cross Validation (SVM)

As mentioned before, two dataset had been used in this study. The first dataset questions were collected from various subjects in business domain while the second dataset questions were collected from same subject in business domain. From the results in Table 4.2.1 and Table 4.2.2, the accuracy of second dataset is much higher than accuracy of first dataset. The difference in accuracy among these two datasets is approximately 10%. The reason behind why the first dataset will have lower accuracy is due to the first dataset consists of questions from different subjects and those questions are distinct and unique. There are no much similar or repeated questions existed in first dataset, hence the SVM classifier do not have satisfied performance when tested with k-fold cross validation method. However, when came to second dataset, SVM had achieved a satisfied accuracy which was around 65%.

4.3 Evaluate with Real Data Type

In order to evaluate the real performance of model built with real data, a total of 60 questions from different Bloom Taxonomy level had been collected from the past year examination papers of different business subjects through UTAR Portal. Those questions were collected from different business subjects such as Business Analytics, Organization Behaviour, Electronic Commerce, Technopreneurship, Business Information Management and Operation Management. Those questions collected will be divided into six categories and 10 questions for each category according to Bloom Taxonomy's level.

Knowledge Level

No	Questions	Correct Label	Predicted Label
1	List FIVE situational factors in the MARS model.	KNOWLEDGE	KNOWLEDGE
2	Define supplier power by detailing the power it has over consumers.	KNOWLEDGE	KNOWLEDGE
3	Identify Groupon's business model.	KNOWLEDGE	KNOWLEDGE
4	Define substitution cipher.	KNOWLEDGE	KNOWLEDGE
5	Define "social engineering".	KNOWLEDGE	KNOWLEDGE
6	Identify one of the successful video games produced by Atari.	KNOWLEDGE	KNOWLEDGE
7	Define buyer bargaining power.	KNOWLEDGE	KNOWLEDGE
8	List THREE examples of business transactions that need to be captured and stored in a Sales transaction processing system (TPS).	KNOWLEDGE	KNOWLEDGE
9	List ANY ONE escalating trend of information technology in tomorrow's world, besides fluid collaboration.	KNOWLEDGE	KNOWLEDGE
10	Define entrepreneurial stress.	KNOWLEDGE	KNOWLEDGE
	Total Accuracy	10/10	100%

Table 4.3.1 Knowledge Level Result

Comprehension Level

No	Questions	Correct Label	Predicted Label
1	Explain the FOUR types of intellectual property.	COMPREHENSION	COMPREHENSION
2	Provide any FIVE reasons for unethical behavior.	COMPREHENSION	COMPREHENSION
3	Explain the FOUR quadrants of ethical decisions.	COMPREHENSION	COMPREHENSION
4	Identify and briefly explain one social issue in Malaysia.	COMPREHENSION	COMPREHENSION
5	Explain the possibilities of entrepreneurial growth in the proposed solution.	COMPREHENSION	COMPREHENSION
6	Briefly describe FIVE characteristic of a object that influences perception.	COMPREHENSION	COMPREHENSION
7	Provide THREE reasons why an information technology project might be abandoned.	COMPREHENSION	COMPREHENSION
8	Briefly explain the difference between supervised learning and unsupervised learning.	COMPREHENSION	COMPREHENSION
9	Identify and explain FOUR methods used to alter another individual's attitude.	COMPREHENSION	COMPREHENSION
10	Define driving forces and explain the THREE stages in Lewin's change model.	COMPREHENSION	KNOWLEDGE
	Total Accuracy	9/10	90%

Table 4.3.2 Comprehension Level Result

Application Level

No	Questions	Correct Label	Predicted Label
1	Use a diagram to explain cloud based business to business platform.	APPLY	COMPREHENSION
2	Apply reverse cipher concept and convert the word "commerce" into cipher text.	APPLY	COMPREHENSION
3	Apply the aggressive, conservative and opportunity loss strategies.	APPLY	ANALYSIS
4	Apply Porter's five competitive forces analysis to examine the winter job industry for Albert.	APPLY	APPLY
5	Discuss the extent to which Supply Chain Analysis can be applied in the logistics sector.	APPLY	APPLY
6	Use Porter's five competitive forces to analyze Apple Inc.	APPLY	APPLY
7	Using an appropriate example, discuss how Django can lead to a good design of a particular web application.	APPLY	APPLY
8	Using appropriate examples, discuss a situation in which Bloom Taxonomy would be suitable to use.	APPLY	ANALYSIS
9	Explain how SST approach can be deployed in the context of Accounting Information System.	APPLY	APPLY
10	Define OB and discuss how the OB can be implemented in the business environment.	APPLY	APPLY
	Total Accuracy	6/10	60%

Table 4.3.3 Application Level Result

Analysis Level

No	Questions	Correct Label	Predicted Label
1	Differentiate data mining and text mining.	ANALYSIS	ANALYSIS
2	Contrast between "First Mile" and "Last Mile". Briefly discuss the important roles of these two distinctions.	ANALYSIS	ANALYSIS
3	Define the open source and distinguish between open source and closed source.	ANALYSIS	ANALYSIS
4	Differentiate between the concept of "in-house" and outsourcing to build an e-Commerce site.	ANALYSIS	ANALYSIS
5	Differentiate the TWO types of knowledge that can co-exist in KM systems, with ONE suitable example.	ANALYSIS	ANALYSIS
6	Differentiate Web 1.0 and Web 2.0 in terms of its page design, content controller, and type of data source.	ANALYSIS	ANALYSIS
7	Analyze Apple Inc by applying Porter's model and determine the strength of these factors for Apple Inc.	ANALYSIS	APPLY
8	Define focus forecasting and differentiate TWO principles in focus forecasting.	ANALYSIS	ANALYSIS
9	Compare and contrast the characteristics of good and service.	ANALYSIS	ANALYSIS
10	Examine FIVE service industry and support each industry with an example.	ANALYSIS	ANALYSIS
	Total Accuracy	9/10	90%

Table 4.3.4 Analysis Level Result

Evaluation Level

No	Questions	Correct Label	Predicted Label
1	Recommend with justification whether to install smart CCTV by using risk management analysis.	EVALUATE	EVALUATE
2	Recommend and describe the new business model Amazon can offer.	EVALUATE	KNOWLEDGE
3	Recommend TWO business revenue models to increase Atari's business revenue. Justify your answer.	EVALUATE	EVALUATE
4	Recommend FIVE generic business strategies for him to achieve a long run business. Justify your answer.	EVALUATE	EVALUATE
5	Provide TWO characteristic of cash and briefly justify cash is preferred than online payment.	EVALUATE	COMPREHENSION
6	Justify on why the threat of bargaining power of customers is considered the reverse of the bargaining power of suppliers.	EVALUATE	COMPREHENSION
7	Justify to what extent 5G technology is better than the 4G technology.	EVALUATE	EVALUATE
8	Evaluate priority rules for sequencing jobs. Elaborate FOUR most popular priority rules.	EVALUATE	EVALUATE
9	Justify answer by highlighting the risks and the concerns you would anticipate from the top management.	EVALUATE	EVALUATE
10	Recommend any FIVE ways to make money through mobile applications.	EVALUATE	EVALUATE
	Total Accuracy	7/10	70%

Table 4.3.5 Evaluation Level Result

Synthesis Level

No	Questions	Correct Label	Predicted Label
1	Suggest any TWO limitations of mass customization.	SYNTHESIS	SYNTHESIS
2	Suggest any information-system-based CRM that can be used to enrich travellers' experience.	SYNTHESIS	SYNTHESIS
3	Suggest and explain ONE entrepreneurial idea to solve the problem by IT.	SYNTHESIS	COMPREHENSION
4	Suggest FIVE rules which you must follow in order to bring innovation to organization. Briefly explain how these rules lead to innovation.	SYNTHESIS	SYNTHESIS
5	Suggest and justify THREE tools you have in mind to propose her in your next meeting.	SYNTHESIS	SYNTHESIS
6	Develop a business solution plan to help her to track customer movements on the Internet?	SYNTHESIS	SYNTHESIS
7	Formulate the estimated regression model that predict the annual cost given the age of the bus.	SYNTHESIS	SYNTHESIS
8	Develop a linear programming formulation to maximize total audience contact.	SYNTHESIS	EVALUATE
9	Create an activity diagram for the process of calculating player's points.	SYNTHESIS	COMPREHENSION
10	Suggest any THREE types of alpha testing for Danish Software Firm.	SYNTHESIS	SYNTHESIS
	Total Accuracy	7/10	70%

Table 4.3.6 Synthesis Level Result

After the testing with real data, the results were recorded down in 6 Table, Table 4.3.1, 4.3.2, 4.3.3, 4.3.4, 4.3.5 and 4.3.6. Table 4.3.1 shows the results of labelling the questions in Knowledge Level. All the questions in this table were labelled correctly by the model and an accuracy of 100% were obtained. Hence, the results indicate that the model constructed has a very well performance on labelling those Knowledge Level questions.

Table 4.3.2 shows the results of labelling the questions in Comprehension Level. 9 out of 10 questions were labelled correctly and an accuracy of 90% were obtained. The last question of this category was mislabelled with Knowledge Level because the question consists of two keywords, the “define” keyword that belongs to the Knowledge Level and the “explain” keyword that belongs to the Comprehension Level. However, this question should be labelled with Comprehension Level because the Comprehension Level has a higher precedence to Knowledge Level due to its level is higher. But the question 4 and 9 with multiple BT keywords were labelled correctly. Even though the model built has a very well performance on labelling those Comprehension Level question, but it still has a very small chance for the model to misclassify the question.

Table 4.3.3 shows the result of labelling the questions in Application Level. 6 out of 10 questions were labelled correctly and an accuracy of 60% were obtained. The model has an average performance on labelling the Application Level questions. Based on the results obtained, some questions with “Use” and “Apply” keywords were labelled wrongly to other BT levels. This may due to the insufficient amount of training data with Application Level. Hence, increasing of amount of data with Application Level when trained the model might improve the model performance in labelling those questions in Application Level.

Table 4.3.4 shows the result of labelling the questions in Analysis Level. 9 out of 10 questions were labelled correctly and an accuracy of 90% were obtained. Question 7 has 3 keywords which are “analyse”, “apply” and “determine”. “Analyse” keyword should has the highest precedence among the keywords but the question was labelled with Application Level. This problem can be overcame by training the model with more training data similar to the question that consists of Analysis Level and Application Level keywords. Indeed, the model did very well in labelling the questions in Analysis Level.

Table 4.3.5 shows the result of labelling the questions in Evaluation Level. 7 out of 10 questions were labelled correctly and an accuracy of 70% were obtained. Question 2 with “recommend” and “describe” keywords shows that when the “recommend” keyword combined with other level keywords, it will mislabelled to other BT level. Question 5 and 6 also have some issues with “justify” keywords because of the misclassification of the BT level. Hence, the model is still considered having a satisfied performance in labelling those Evaluation Level questions.

Table 4.3.6 shows the result of labelling the questions in Synthesis Level. 7 out of 10 questions were labelled correctly and an accuracy of 70% were obtained. Question 3 consists of “suggest” and “explain” keywords. The “suggest” keyword should be higher than “explain” keyword in terms of precedence. However, it was mislabelled with Comprehension Level which is the level of “explain” keyword. The question 8 and 9 with “develop” and “create” keywords were misclassified to other BT levels and this might due to insufficient amount of training data with those keywords. Thus, the model has a quite satisfied performance in labelling the Synthesis Level questions.

BT Level	Total Correct Answer	Accuracy
Knowledge	10	100%
Comprehension	9	90%
Application	6	60%
Analysis	9	90%
Evaluation	7	70%
Synthesis	7	70%
Total	48	80%

Table 4.3.7 Comparison and Total Accuracy

Based on the Table 4.3.7, a total accuracy of the model obtained was 80%. The performance of constructed model is quite good since the accuracy obtained already exceeds the accuracy that stated in the objective.

4.4 Graphical User Interface

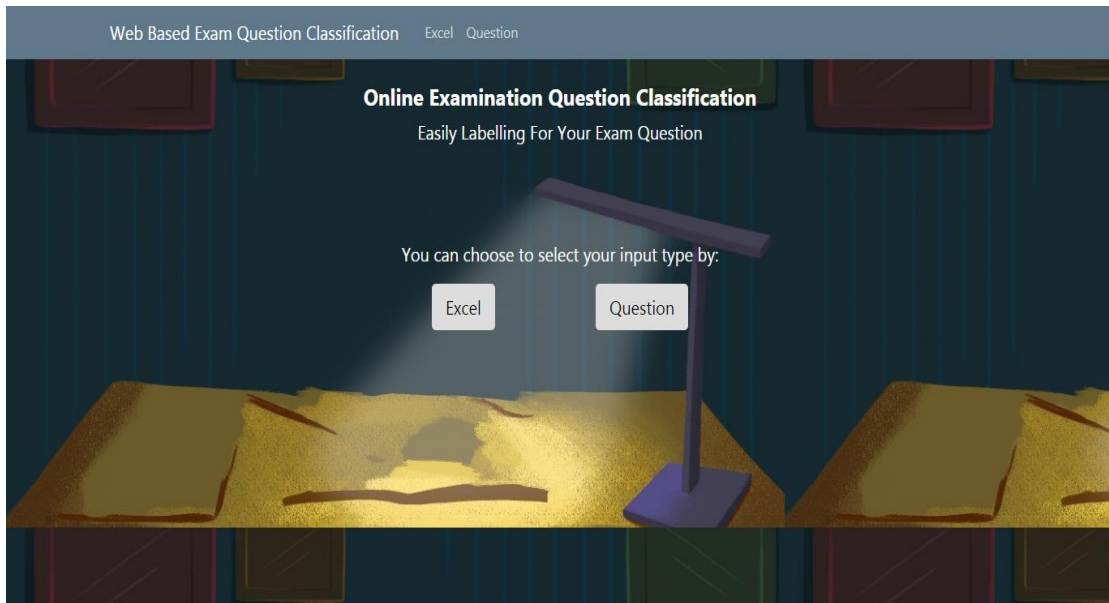


Figure 4.4.1 Home Page

Basically, the web application consists of 3 web pages such as Home page, Excel page and Question page. The web's URL is "gktan016.pythonanywhere.com", Figure 4.4.1 that shows the home page of the web application. In this page, the web application allows the user to choose their input type by clicking on the "Excel" button or "Question" button. Those buttons will bring the user to the respective pages such as "Excel" button will bring the user to Excel page while the "Question" button will bring the user to Question page.

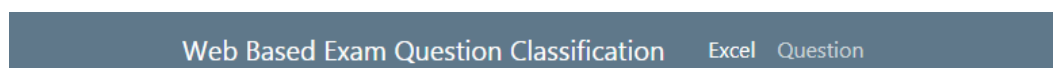


Figure 4.4.2 Navigation Bar

Figure 4.4.2 shows the navigation bar of the web page. And this navigation bar will always appear at the top of every web pages in this web application. The navigation bar consists of 3 items which are "Web Based Exam Question Classification" that will link to the Home page, "Excel" that will link to the Excel page and "Question" that will link to the Question page.

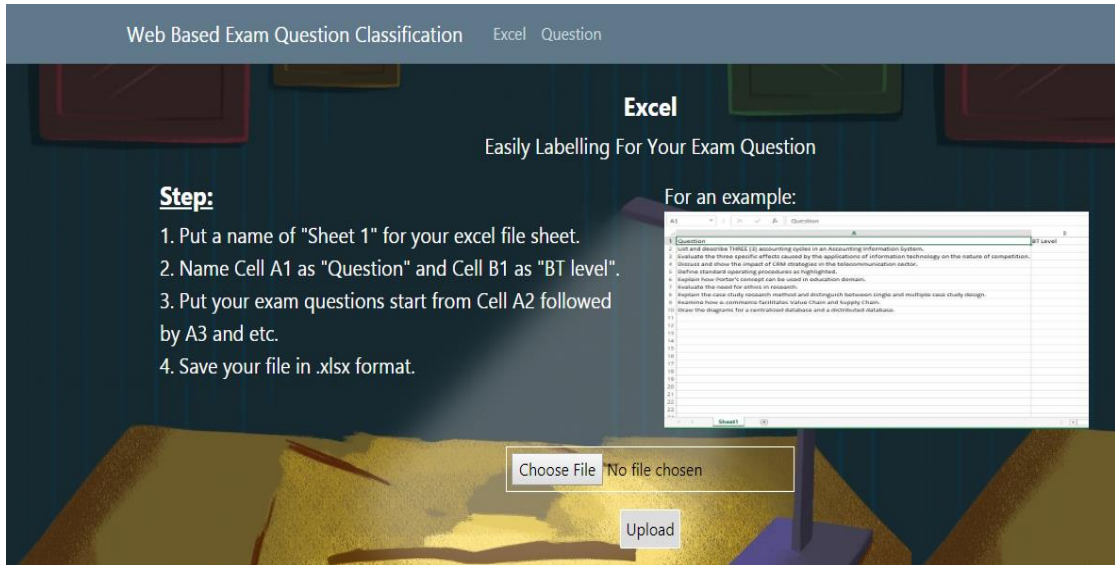


Figure 4.4.3 Excel Page

Figure 4.4.3 shows the interface of Excel page after user clicked on the “Excel” button or the “Excel” link in the navigation bar. In this page, there will be an instruction that guide the user to prepare their excel file properly in order to avoid errors occur. Then, there will also be a sample image beside that allow user to visualize the correct input.

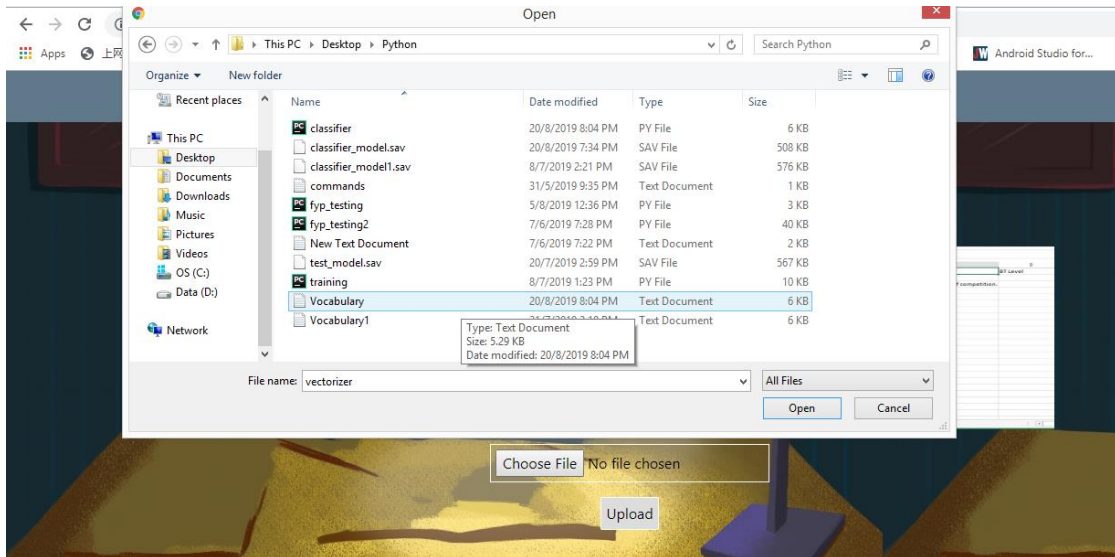


Figure 4.4.4 Choose File

Figure 4.4.4 shows the interface of the Excel page after user clicked on the “Choose File” button. A document window will pop out and enable the user to select one file to be uploaded.

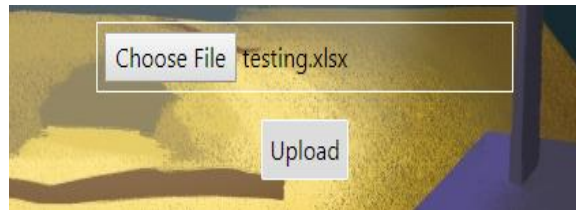


Figure 4.4.5 Upload

Figure 4.4.5 shows the interface of Excel page after user select a file to be uploaded. The name of the file will appear next to the “Choose File” button and the format of the file. Afterwards, user can click on the “Upload” button to get the result.

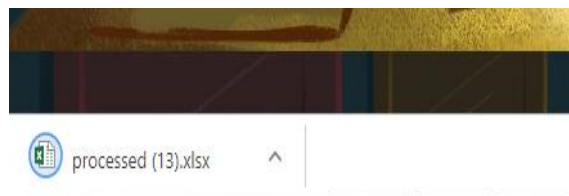


Figure 4.4.6 Attachments

Figure 4.4.6 shows the attachments downloaded when the user clicked on “Upload” button. An output file named “preprocessed.xlsx” will be downloaded by the user as attachments.

1	Question	BT Level
2	List FIVE situational factors in the MARS model.	
3	Define supplier power by detailing the power it has over consumers.	
4	Explain the FOUR types of intellectual property.	
5	Provide any FIVE reasons for unethical behavior.	
6	Define electronic commerce.	
7	Briefly explain any TWO (2) observations of information technology trend using Moore's Law.	
8	Information systems value is determined by the strong relationships among THREE (3) major components. Name them.	
9	Discuss the extent to which Value Chain Analysis can be applied in the logistics sector.	
10	Apply Porter's five competitive forces analysis to examine the summer job industry for your uncle.	
11	Suggest ONE way the outsourcing model mentioned in Q3.(a)(i) will work to address the issue presented in the case.	

Figure 4.4.7 Sample Input

1	Question	BT Level
2	List FIVE situational factors in the MARS model.	KNOWLEDGE
3	Define supplier power by detailing the power it has over consumers.	KNOWLEDGE
4	Explain the FOUR types of intellectual property.	COMPREHENSION
5	Provide any FIVE reasons for unethical behavior.	COMPREHENSION
6	Define electronic commerce.	KNOWLEDGE
7	Briefly explain any TWO (2) observations of information technology trend using Moore's Law.	COMPREHENSION
8	Information systems value is determined by the strong relationships among THREE (3) major components. Name them.	KNOWLEDGE
9	Discuss the extent to which Value Chain Analysis can be applied in the logistics sector.	APPLY
10	Apply Porter's five competitive forces analysis to examine the summer job industry for your uncle.	APPLY
11	Suggest ONE way the outsourcing model mentioned in Q3.(a)(i) will work to address the issue presented in the case.	COMPREHENSION
12		
13	Knowledge Question	4
14	Comprehension Question	4
15	Application Question	2
16	Analysis Question	0
17	Evaluation Question	0
18	Synthesis Question	0
19		

Figure 4.4.8 Sample Output

The content of sample input and output are shown in Figure 4.4.7 and Figure 4.4.8 respectively. Initially, the column of BT level were empty in Figure 4.4.7. After processing by the web application, the questions will be labelled with the predicted BT level by the model. Furthermore, the number of the labelled questions for each BT level will be shown at bottom of the excel file.

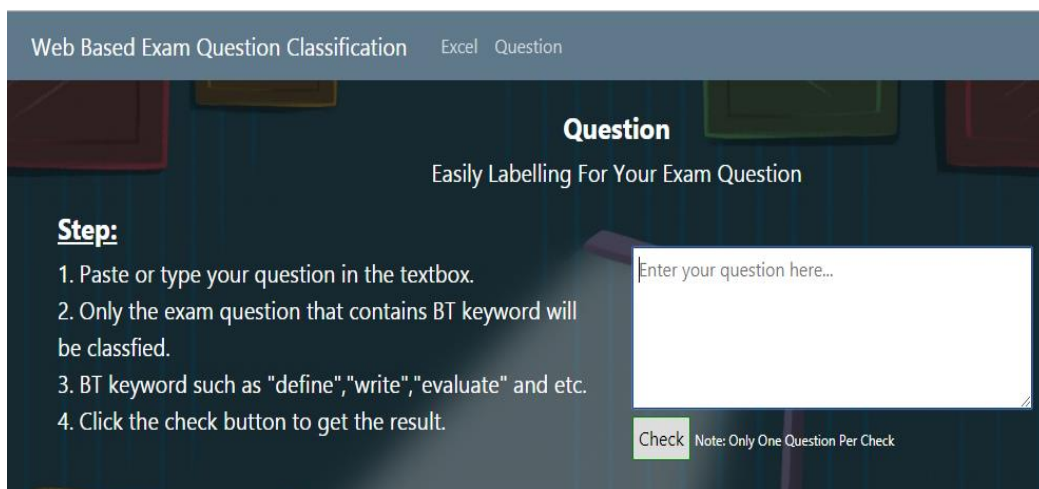


Figure 4.4.9 Question Page

Figure 4.4.9 shows the interface of Question page after user clicked on the “Question” button or the “Question” link in the navigation bar. In this page, there will be an instruction that guide the user to use the features in this web page. There will be a textbox beside the instruction that allow the user to input their question. However, only one question will be classified each time.

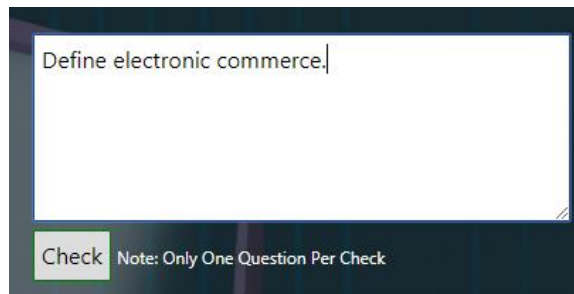


Figure 4.4.10 Sample Input of Question

Figure 4.4.10 shows the sample input of question that typed by the user. After user paste or type the question, the user can click on the “Check” button to get result.

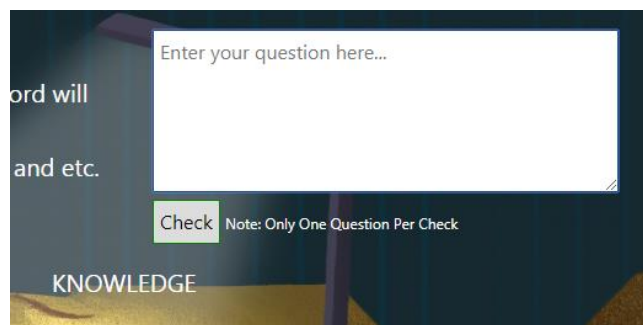


Figure 4.4.11 Sample Output of Question

Figure 4.4.11 shows the result get when user clicked on the “Check” button. The result will be displayed below the “Check” button. The textbox will be clear and user can retype or paste another question to perform classification for the other question.

Chapter 5 Conclusion

The project has presented a web based examination question classification application with 80% of accuracy achieved and will benefit the lecturers and students in the education environment. The web application will classify the input question into six different level of Bloom Taxonomy based on Bloom Taxonomy cognitive domain automatically. The web application is important because it will save the lecturer's time and improve the student's understanding about the particular subject.

The study of Bloom Taxonomy (BT), Natural Language Processing (NLP), bag-of-words model, question classification tools, Python programming language and web programming languages are required and had been done in order to develop the web application. Furthermore, NLP and machine learning classification techniques such as Support Vector Machine also play a very important role in order to succeed the application. In order to train the model, a total number of 163 questions along with expert's label had been fit to the model in order to let the model learn and familiar with the pattern of the questions.

This question classification tool developed might help the lecturer to overcome the problems arise when deal with huge amount of questions. However, there are some limitations for this tool to classify the questions especially with Application Level question because only 60% of accuracy was obtained. The problem might be overcome by increasing the number of training data in the dataset.

5.1 Implementation Issue and Challenge

In this project, a web based examination question classification tool have been developed in order to classify the text-based examination questions into six different of Bloom Taxonomy level and a satisfied accuracy of 80% was obtained. However, the tool still will misclassify the question wrongly when deal with the question with multiple BT keywords. For instance, “Identify and explain issue of cheating”. There will be some chances for the tool to mislabel the question to the wrong BT level. Besides, there are some BT keywords will be fall into different categories BT level under certain circumstances. For an example, “describe” belongs to Knowledge level in normal situation but when it joined with the “briefly” and became “briefly describe”, it is no longer a Knowledge Level keyword but a Comprehension Level keyword. Furthermore, the training data set is still insufficient in order to achieve a very high accuracy and a larger training data set is required. Moreover, the tool will not able to classify those examination questions that consists of figures, tables, images or calculation. Finally, the tool is only developed for classify the examination question in English language.

5.2 Future Work

There are some aspects or features of this web based question classification tool can be improved. First, the tool’s accuracy can be improved. Although the tool achieved accuracy had exceeded the accuracy stated in the objective, but the accuracy still has the chances to be improved by increasing the number of training data in the dataset. Next, the tool is only able to classify the text-based questions. Thus, the tool can be further improved by adding the ability of the tool to classify the questions in figure or image form and also the questions which include mathematical operations and calculation. Moreover, the tool is only developed to classify the questions in English language. A huge improvement can be made if the tool is able to classify the examination questions not only in English language but other languages too.

Bibliography

Abdulhadi, D., 2015. Exam Question Classification Based On Bloom Taxonomy Cognitive Level Using Classifiers Combination. *Journal Of Theoretical And Applied Information Technology*, 78(3), pp. 447-454.

Chen, CX & Han, HJ & Liu, Z, 2014. KNN question classification method based on Apriori algorithm. *Journal of COMPUTER MODELLING & NEW TECHNOLOGIES* , 18(9), pp. 371-379.

Chris, B., 1985. Implementation Of SMART Information Retrieval System. *Journal Of Computer Science Engineering*, pp. 1-38.

Fattah, S & Tanalol, ST & Mamat, M, n.d. Classification of Examination Questions Difficulty Level. pp. 204-207.

Fredick, W., 2018. *STORYBOARD GRAPHIC: BLOOM'S TAXONOMY* , s.l.: WordPress and Origin.

Gupta, P & Gupta, V, 2012. A Survey of Text Question Answering Techniques. *Journal Of Computing Applications*, 53(4), pp. 1-8.

Han, J., 2006. *Data Mining: Concepts and Training*. *Journal Of Computer Science*.

Hearst, M., 2003. What Is Text Mining?. *Journal Of Text Data Mining*, pp. 1-3.

Heiner, C & Joseph, LZ, 2009. Improving Student Question Classification. *Journal Of Education Data Mining*, pp. 259-268.

Kim, SB et al., 2006. Some Effective Techniques for Naive Bayes. *Journal Of Knowledge And Data Engineering*, 18(11), p. 1.

Marban, O & Segovia, J & Menasalvas, E & Covadonga, F, 2009. Toward data mining engineering: A software engineering approach. *Journal of Information Systems*, 34(1), pp. 87-107.

Marko, B., 1995. Learning Information Retrieval Agents: Experiment With Automated Web Browsing. *Journal Of Computer Science Engineering* , pp. 13-18.

Melissa, K., 2019. Bloom's Taxonomy - Application Category. Available from: <https://www.thoughtco.com/blooms-taxonomy-application-category-8445> [22 Aug 2019]

Micheal, A., 2013. THE JIKITOU BIOMEDICAL QUESTION ANSWERING SYSTEM. *Journal Of Mining and Technology*.

Micheal, J., 1998. A Language Modeling Approach To Information Retrieval. *Journal Of Computer Science*, pp. 1-12.

Omar, N & Haris, SS & Hassan, R & Rahmat, M & Zainal, NFA & Zulkifli, R, 2011. Automated analysis of exam questions according to bloom's taxonomy. *Journal of Social and Behaviour Science*, pp. 297-303.

Prasad, J et al., n.d. Web-Based Question Answering System. *Journal Of Computer Science*.

Sangodiah, A & Ahmad, R & Ahmad, WFW, 2017. Taxonomy based features in question classification using support vector machine. *Journal Of Theoretical and Applied Information Technology*, 95(12), pp. 2814-2816.

Shrihari, C & Desai, A, 2015. A Review of Knowledge Discovery using Text Classification Techniques In Text Mining. *Journal Of Computer Applications*, 111(6), pp. 12-14.

Sriram, B., 2010. SHORT TEXT CLASSIFICATION IN TWITTER TO. *Journal Of Computer Science*, pp. 1-12.

Sundblad, H., 2007. Question Classification in Question Answering Systems. *Journal in Science and Technology*, pp. 1-9.

Yadav, R & Mishra, M, 2013. Question Classification Using Naïve Bayes. *International Journal of Engineering and Innovative Technology (IJEIT)*, 2(8), pp. 291-294.

Yuan, Q et al., 2012. Enhancing Naives Bayes with Various Smoothing Mehods for Short Text Classification. p. 645.

Appendix A: Training Dataset

NO	QUESTION	LEVEL	BT LEVEL
1	Suggest any TWO (2) efforts that organization may perform in order to discourage unethical behavior.	HOTS	CREATE
2	Discuss how Resource Based View can be implemented in organizations to reinforce financial standing.	MOTS	APPLY
3	Define factors of production.	LOTS	REMEMBER
4	Briefly describe FIVE (5) general types of factors of production.	LOTS	UNDERSTAND
5	In your opinion, provide any TWO (2) reasons why the actions of one firm can significantly affect the sales of the other firms in an oligopoly market.	LOTS	UNDERSTAND
6	Define electronic commerce.	LOTS	REMEMBER
7	Explain THREE (3) broad categories of e-commerce with examples.	LOTS	UNDERSTAND
8	Discuss any THREE (3) ways by which an organization can benefit from e-commerce.	LOTS	UNDERSTAND
9	Advise Smith on any FIVE (5) types of information which should be included in the website.	HOTS	CREATE
10	Suggest any THREE (3) specific ways with elaborations on how this new business may utilize the Internet technology to achieve cost efficiency.	HOTS	CREATE
11	Differentiate between data and information using examples.	MOTS	ANALYZE
12	State any FOUR (4) characteristics of valuable information.	LOTS	REMEMBER
13	Explain the differences between volatile and non-volatile storage.	LOTS	UNDERSTAND
14	Suggest THREE (3) specific types of application software that may help to support James's business activities.	HOTS	CREATE
15	List and describe THREE (3) accounting cycles in an	LOTS	REMEMBER

Appendix

	Accounting Information System.		
16	Explain how SWOT and Porter's techniques is applied in the manufacturing industries.	MOTS	APPLY
17	State FOUR (4) basic business activities that are performed in the revenue cycle.	LOTS	REMEMBER
18	List any THREE (3) practical approaches for payment collection in revenue cycle.	LOTS	REMEMBER
19	Compare and contrast Manufacturing Resource Planning (MRP-II) and Lean Manufacturing.	MOTS	ANALYZE
20	As a manager in an Information Technology firm, you are required to suggest and elaborate any THREE (3) actions that organizations can take as part of the responsibility towards the society.	HOTS	CREATE
21	Illustrate each of these situations with TWO (2) real-life examples.	LOTS	UNDERSTAND
22	Explain the differences between ethical dilemma and ethical lapse.	LOTS	UNDERSTAND
23	Suggest and justify ONE (1) strategy that your company may consider for marketing its products internationally.	HOTS	EVALUATE
24	Differentiate between Market Economy and Planned Economy.	MOTS	ANALYZE
25	Provide each of Market Economy and Planned Economy with an example.	LOTS	UNDERSTAND
26	Differentiate between a wholesaler and a retailer.	MOTS	ANALYZE
27	In your opinion, identify the relationship between factors of and a country's economic system.	LOTS	UNDERSTAND
28	Briefly describe THREE (3) types of business process that traditionally organized around functional areas of business. Provide an example of each.	LOTS	REMEMBER
29	Propose and justify ONE (1) international market-entry strategy that you may consider to market your products internationally.	HOTS	EVALUATE
30	Construct a basic diagram or figure to demonstrate your understandings on what Information System is.	LOTS	UNDERSTAND

Appendix

31	State a main difference between a customer and a contact.	LOTS	UNDERSTAND
32	Identify ONE (1) most likely error of the following quotation.	LOTS	REMEMBER
33	Determine which sub-stage that the deal is currently in for the following cases.	MOTS	APPLY
34	Three sales sub-stages: qualification, proposition and negotiation are included in both OpenERP and SugarCRM, before a deal is won. Give a brief description of these sub-stages.	LOTS	UNDERSTAND
35	State your recommended solution with justification.	HOTS	EVALUATE
36	Explain TWO (2) ways on how monopoly harms consumers and hurts the economy. Justify and illustrate your answers.	HOTS	EVALUATE
37	Suggest and elaborate any TWO (2) efforts or methods that your company may use in order to gain competitive advantage over your competitors.	HOTS	CREATE
38	Justify your answers with elaborations by providing any TWO (2) reasons.	HOTS	EVALUATE
39	Differentiate between data and information using a real life example.	MOTS	ANALYZE
40	Briefly discuss on the difference between volatile and non-volatile storage.	LOTS	UNDERSTAND
41	Briefly discuss any THREE (3) specific internet technologies or tools that are commonly found in providing support for business activities.	LOTS	UNDERSTAND
42	Provide a real life example on conflict of interest situation.	LOTS	UNDERSTAND
43	Illustrate your understanding for oligopoly competition and monopolistic competition using ONE (1) real life example for each.	LOTS	UNDERSTAND
44	Briefly discuss any TWO (2) consequences for whistle-blowing activity.	LOTS	UNDERSTAND
45	List any TWO (2) real life examples of the organization that implement the clicks-and-bricks B2C E-Commerce.	LOTS	REMEMBER


Appendix B Poster

A WEB BASED APPLICATION THAT CLASSIFY EXAM QUESTION DIFFICULTY BASED ON BLOOM TAXONOMY COGNITIVE LEVEL

Introduction

- TO ANALYSE THE RESEARCH WORK OF EXISTING QUESTION CLASSIFICATION BASED ON BT
- TO DEVELOP A WEB BASED APPLICATION THAT APPLY THE BT'S CONCEPT AND SUPPORT VECTOR MACHINE CLASSIFIER.
- TO EVALUATE THE WEB BASED APPLICATION IN TERMS OF METHODOLOGY.

Objective



UTAR PRESENTS

WEB BASED APPLICATION OF EXAM QUESTION CLASSIFICATION



Presented by Tan Guan Kheng

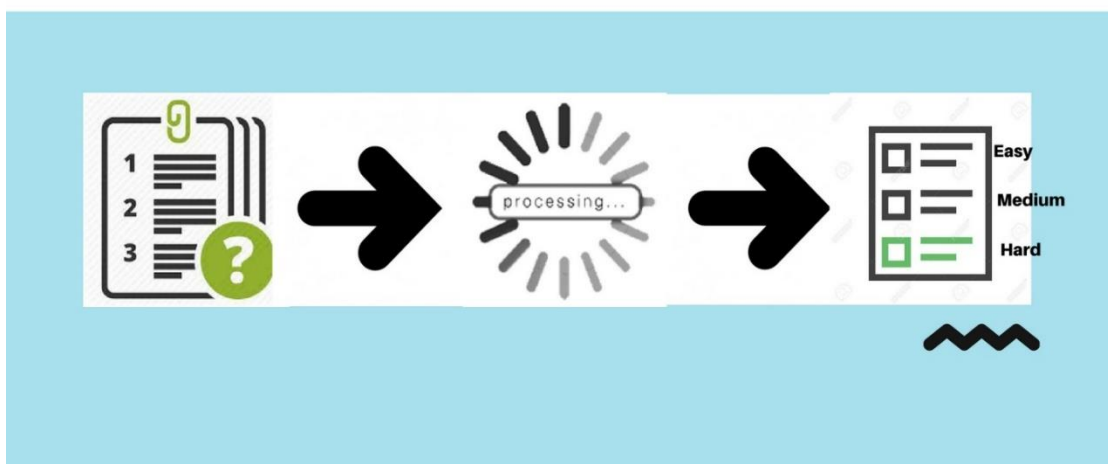
Supervisor: Mr Anbuselvan a/l Sangodiah

- CRISP DMLC MODEL
- SUPPORT VECTOR MACHINE LEARNING
- PYTHON, PYCHARM, DJANGO

Methodology

THIS WEB APPLICATION WILL BENEFIT BOTH LECTURERS AND STUDENTS BY SAVING LECTURER'S TIME TO CLASSIFY MANUALLY AND IMPROVE UNDERSTANDING OF STUDENT ABOUT THE SUBJECT

Conclusion



Plagiarism Check Result



1.2 Background Information

In order to understand how this system works, there are few terms that need to be understood first. There are text mining, text classification, supervised learning, question classification and Bloom Taxonomy. According to Hearst (2003, p1), 'Text Mining is the discovery by computer of new, previously unknown information by automatically extracting information from different written resources. A key element is the linking together of the extracted information together to form new facts or new hypotheses to be explored further by more conventional means of experimentation.' Determine the unknown information that are not found by others and not yet written will be the goal of the text mining. Text mining actually is a subset of data mining that will work to discover the pattern or trend based on historical data. (Hearst, 2003). For instance, Econsave maybe found that when a customer buy a beer, they will also have high chance to buy a diaper. This is the interesting pattern that the text mining or data mining want to look for from the databases.

Text mining normally have three processes such as text clustering and classification,

Match Overview
✕

6%

#	Source	Similarity
1	www.emeraldinsight.c... <small>Internet Source</small>	1%
2	www.cs.bham.ac.uk <small>Internet Source</small>	1%
3	cmnt.lv <small>Internet Source</small>	1%
4	Anbuselvan Sangodiah, ... <small>Publication</small>	<1%
5	www.educationaldata... <small>Internet Source</small>	<1%
6	Maria Grazia Violante, ... <small>Publication</small>	<1%
7	Subashini, M. Monica, ... <small>Publication</small>	<1%

Count: 9110 Text-only Report | High Resolution On

Turnitin - Google Chrome

turnitin.com/newreport.asp?r=13.424887939005&svr=57&lang=en_us&oid=1162217094&pbid=2&ft=1&ro=

Processed on: 22-Aug-2019 11:25 +08
ID: 1162217094
Word Count: 9110
Submitted: 1

FYP2 Report
By Tan GUAN KHENG

Similarity Index	Similarity by Source
6%	Internet Sources: 4% Publications: 3% Student Papers: N/A

[include quoted](#) | [include bibliography](#) | [exclude small matches](#)
mode: show highest matches together | [Change mode](#)

Chapter 1 Introduction 1.1 Problem Statement Examination is defined as a formal test that we need to take in order to test our knowledge, skill and understanding on certain subjects. Examination can be considered as the most effective way to evaluate the student's performance on certain subjects. Normally, a standard examination paper nowadays will be designed according to Bloom Taxonomy's concept which consist of six categories of questions that will be knowledge, comprehension, application, analysis, evaluation and synthesis. The level of difficulty will increase start from knowledge to synthesis. However, the problem arises when the student and lecturer want to classify the difficulty of questions manually. According to Yusof and Chai cited in Omar et al. (2011), normally all the lecturers have the knowledge and skills to classify the question difficulty manually, but not all of them can identify the question difficulty correctly. Even the lecturers have the problems in classifying the question, not to mention the students because mostly all the students don't have adequate skills and knowledge about the question difficulty. In addition, even some lecturers are able to classify the questions manually, but it could be inefficient and very time consuming when there are huge amount of questions that need to be classified. It is good for both student and lecturer if there is a system exists that will help the lecturer and student to classify the examination questions automatically just by inputting a list of examination questions. According to Sangodiah et al. (2017), there are some question classifier system that exists at the past but they have poor performance for the questions that come from different areas or domains. Furthermore, there are few researchers (Sangodiah et al. 2017; Fattah et al. 2015; Omar et al. 2011) who had done research on how to use the question classifiers to

classify the question based on cognitive level of Bloom Taxonomy.
29

However, all of the research done are only limited to the theory level. There are no such complete system exists yet to help the student and lecturer to classify the exam question from different areas based on difficulty. Hence, a web based application that can classify the examination questions based on Bloom Taxonomy concept accurately will be developed in this paper. 1.2 Background Information In order to understand how this system works, there are few terms that need to be understood first. There are text mining, text classification, supervised learning, question classification and Bloom Taxonomy. According to

- 1 1% match (Internet from 23-Oct-2018)
<https://www.emeraldinsight.com/doi/full/10.1>
- 2 1% match (Internet from 27-Nov-2018)
<http://www.cs.bham.ac.uk>
- 3 1% match (Internet from 02-Jul-2018)
<http://cmnt.lv>
- 4 < 1% match (publications)
[Anbuselvan Sangodiah, Rohiza Ahmad, Wan Fatimah Wan Ahmad. "A review in feature extraction approach in question classification using Support Vector Machine", 2014 IEEE International Conference on Control System, Computing and Engineering \(ICCCSE 2014\), 2014](#)
- 5 < 1% match (Internet from 09-Dec-2018)
<http://www.educationaldatamining.org>
- 6 < 1% match (publications)
[Maria Grazia Violante, Sandro Moos, Enrico Vezzetti. "A methodology for supporting the design of a learning outcomes-based formative assessment: the engineering](#)

Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1 of 1



FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	
ID Number(s)	
Programme / Course	
Title of Final Year Project	

Similarity	Supervisor's Comments (Compulsory if parameters of originality exceeds the limits approved by UTAR)
Overall similarity index: _____ % Similarity by source Internet Sources: _____ % Publications: _____ % Student Papers: _____ %	
Number of individual sources listed of more than 3% similarity: _____	
Parameters of originality required and limits approved by UTAR are as Follows: (i) Overall similarity index is 20% and below, and (ii) Matching of individual sources listed must be less than 3% each, and (iii) Matching texts in continuous block must not exceed 8 words <i>Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.</i>	

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Signature of Supervisor

Name: _____

Date: _____

Signature of Co-Supervisor

Name: _____

Date: _____



UNIVERSITI TUNKU ABDUL RAHMAN

**FACULTY OF INFORMATION & COMMUNICATION
TECHNOLOGY (KAMPAR CAMPUS)**

CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	
Student Name	
Supervisor Name	

TICK (√)	DOCUMENT ITEMS
	Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
	Front Cover
	Signed Report Status Declaration Form
	Title Page
	Signed form of the Declaration of Originality
	Acknowledgement
	Abstract
	Table of Contents
	List of Figures (if applicable)
	List of Tables (if applicable)
	List of Symbols (if applicable)
	List of Abbreviations (if applicable)
	Chapters / Content
	Bibliography (or References)
	All references in bibliography are cited in the thesis, especially in the chapter of literature review
	Appendices (if applicable)
	Poster
	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)

*Include this form (checklist) in the thesis (Bind together as the last page)

I, the author, have checked and confirmed all the items listed in the table are included in my report.	Supervisor verification. Report with incorrect format can get 5 mark (1 grade) reduction.
_____ (Signature of Student)	_____ (Signature of Supervisor)
Date:	Date: