TEST Engineering & Management

March – April 2020
ISSN: 0193-4120 Page No. 6463 - 6470

Question Classification Based on Cognitive Levels using Linear SVC

Suliana Sulaiman*, Rohaizah Abdul Wahid, Asma Hanee Ariffin and Che Zalina Zulkifli
Computing Department,
Faculty of Art, Computing and Industry Creative,
Universiti Pendidikan Sultan Idris,
35900 Tanjong Malim, Perak, Malaysia.
suliana@fskik.upsi.edu.my*, rohaizah@fskik.upsi.edu.my, asma@fskik.upsi.edu.my, chezalina@fskik.upsi.edu.my

Article Info
Volume 83
Page Number: 6463 - 6470
Publication Issue:
March - April 2020

Article History
ArticleReceived*: 24 July 2019*
Revised*: 12 September 2019*
Accepted*: 15 February 2020*
Publication*: 04 April 2020*

Abstract:
A student's cognitive level can be determined through an assessment such as final examination. A person needs to have skills and knowledge with regard to educational assessments to make sure the questions are concurrent with the cognitive level. The aim of this paper is to find the best classifier to classify exam questions based on cognitive levels. The experiment is conducted in two phases. The first phase is to find the best mapping for SVM classifier (One-Versus-One and One-Versus-All). The classifier that produces the best result for mapping is used in the second phase for Naïve Bayes, KNN and Linear SVC. The result showsthat Linear SVC with OVO is the best classifier with 74.8% for f-measure and tf-idf as feature extraction which really benefits in increasing the classifier's result. In future, the classifier will be tested to classify questions in the Malay language

Keywords: *Exam question classification, text classification, question classification*.

I. Introduction

Examination questions should be based on some learning taxonomies to maintain the quality of the questions (Taqi, 2016). To develop the correct questions for certain taxonomy levels, a person needs to have skills and knowledge with regard to educational assessments. Those who do not come from this background will have difficulties in creating quality questions for each level (Walsh, Bower,& Sweller, 2017; Ullah et al., 2019).

Assessments are strongly related to teaching and learning (T&L), which are normally combined as one complete process. A student's cognitive level can be determined through an assessment, which measures the learning outcome of the T&L. Most studies used Bloom's Taxonomy to classify the cognitive levels (Bloom, 1956). In 2001, Krathwohl (Bloom, 1956) revised Bloom's Taxonomy into new levels, and named it the Revised Bloom's Taxonomy (Anderson & Krathwohl, 2001). The difference between these taxonomies can be seen in level 5 and level 6. Anderson and Krathwohl (2001)also changed the nouns for each level (Bloom, 1956) into verbs. Figure 1 shows the Revised Bloom's Taxonomy and Figure 2 shows the Bloom's Taxonomy.

Question classification is unique in comparison to document classification because it deals with short sentences with less information unlike text documents (Abduljabar &Omar, 2015; Li &Roth, 2006; Hui, Liu,& Ouyung, 2011). Many studies had tried several techniques to enhance the classifier result in order to ease the educator's task. These included statistical, machine learning, and optimization. There is a limitation for each of them, such as the statistical technique where it requires a huge amount of data

March – April 2020
ISSN: 0193-4120 Page No. 6463 - 6470

to maintain its accuracy (Abduljabar &Omar, 2015; Phan, Nguyen,&Horighuchi, 2008; Wang, Li,& Ren, 2010).

The aim of this study is to find the best classifier to classify the questions based on the cognitive levels in the Revised Bloom's Taxonomy. Section 2 explains previous work from other researcherswhile Section 3 describes the methodology used in this study. The results are described in Section 4 and the conclusion of the study is in Section 5.

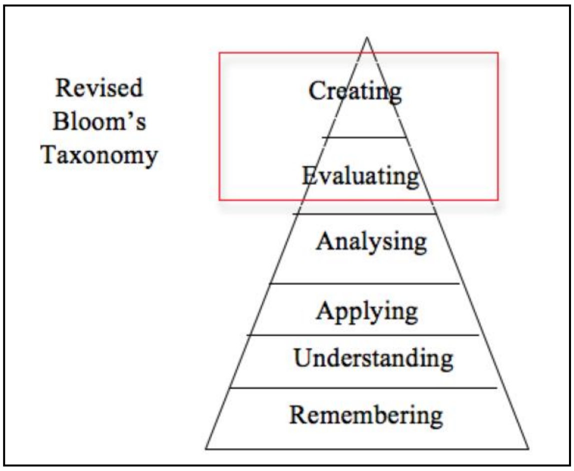

Figure 1: Revised Bloom's Taxonomy

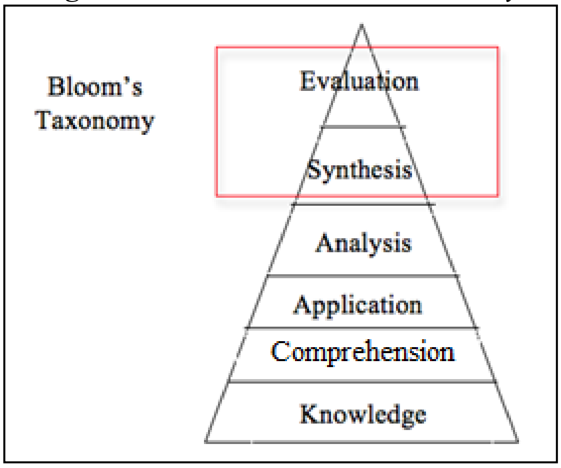

Figure 2: Bloom's Taxonomy

II. Previous Works

The papers collected from IEEE, ScienceDirect and Scopus for the systematic literature review can be divided into three clusters. These papers focused on classifying the questions based on cognitive level. Thirteen of the papers used Bloom's Taxonomy (Bloom, 1956) and only one of them used Revised Bloom's Taxonomy (Anderson & Krathwohl, 2001). The first cluster is concentrated on a review of question classification, while the second cluster is more on proposing the method or technique to increase the performance (accuracy, precision recall and F-measure) for question classification, and the last cluster is to propose a framework to develop the whole question classification system. Analysis from all the papers in the second cluster helped this study to find the best classifier to classify the questions.Detailed of the cluster are discussed below.

2.1 Natural Language Processing (NLP) and Rule-based

The rule-based approach always involved a domain expert in order to generate the correct rule for certain problems. Haris and Omar (2012) developed a rule-based question classification for Bloom's Taxonomy using POS tagging and Regex. The pattern and rules are separated based on supporting statement, symbol, method, class or function's name and also special word. Jayakodi and Perera (2015) used WordNet Similarity and rule-based to increase the accuracy of question classification. More rules will boost the accuracy of the result but itis only suitable for specific languages and domains.

2.2 Machine Learning

The machine learning approach was used in many studies in order to classify questions in multiple languages and domains. Six out of seven papers in Table 1 used Support Vector Machine (SVM) while Supriyanto (2013) used Naïve Bayes as the classifier. The other study tried to compare the accuracy of SVM, Naïve Bayes and K-Nearest Neighbour using different feature selections and extractions. Anekbon (2019) used unigram as the feature selection while Abduljabar (Abduljabar & Omar, 2015) used chi-Square, Odd Ratio and

Published by: The Mattingly Publishing Co., Inc

Mutual Information. The use of tf-idf as a feature extraction can also be meaningful to increase the accuracy of classifier as stated in Mohammed (Mohammed & Omar, 2018). Besides KNN and Naïve Bayes, Anekbon (2019) also used decision trees and multilayers perceptron to compare the resultswhichshowed that SVM produceda higher accuracy compared to the others. When the results from all the classifiers were analyzed, SVM was more inclined to show the highest accuracy for F-measures compared to the others. Table 1 summarizesthe classifiers that had been used in previous work.

Table 1: Summary of classifiers that had been used in previous work

Author	**Paper Type**	**ML**
Supriyanto et al. (2013)	Journal	Naïve Bayes
Abduljabbar &Omar, N (2015)	Journal	KNN, Naïve Bayes and SVM
Kusuma, Siahaan &Yuhana(2016)	Conference	SVM
Sangodiah, Ahmad &Ahmad (2017)	Conference	SVM
Mohammed &Omar (2018)	Journal	KNN, Naïve Bayes and SVM
Anekboon (2019)	Journal	Decision Tree, multilayer perceptron, Naïve Bayes and SVM
Yahya et al. (2013)	Journal	KNN, Naïve Bayes and SVM

Based on Table 1, most of the studies stated that SVM gave accurate results when compared to others. SVM can be used together with binary classifications such as One Versus All (OVA) and One Versus One (OVO).

- **One Versus All (OVA)**

  OVA tends to classify N classes into N binary problems. Each problem distinguishes a given class from the other N-1 classes. For example,a binary classifier is represented as N and it trains with N class (positive example) and N-1 class (negative example). The results for the prediction of a new object will take the maximum output and the corresponding class label will be assigned to the object (Ryan & Aldebano, 2004).

- **One Versus One (OVO)**

  OVOis also known as pairwise classification. In the learning phase, each dataset in certain classeswill be compared to other classes. The binary classifier N(N-1)/2 will differentiate between each pair of class. If the N is 5, then the total of learned model is 10. At the end of the classification, each class with the maximum value will be given 1 vote. The highest vote will be determined as the class. For the biggest number of class, this method will result in an imbalance classification (Gualtieri & Cromp, 1999).

III. Methodology

The dataset from Anwar (2011) was used. Overall,it consists of 600 questions (100 questions respectively for Remembering, Understanding, Applying, Analysis, Evaluating, and Creating). The questionswere structured and in essay form without any figures or tables.The experiments were divided into two phases. The first phase was to find the best mapping result between OVO and OVA for the SVM classifier. The dataset was cleaned using regex. All the upper case were converted into small capital letters and exclamationswere removed from the questions.

The stop word was used to remove the common words. Vectorization and tf-idf were used as feature extraction for the questions. During this phase, two classifiers (Linear SVC and SVC) from the Scikit-learn library were used. Classifiers that produced a higher result for F-measure were used in the second phase.

The second phase was to find the best classifier that produced the best result. The same dataset was used for this phase. The data was cleaned using regex and stop words were removed. For this phase,the snowball stemmer was applied. The same feature extraction from the first phase was used. The results for these two phases are reported in Section 4.

IV. RESULTS

Based on the literature review, SVM produced the best results compared to other classifiers. For the first phase, Scikit-learn library for the SVMwas used (Alex J. & Scholkopf, 2004). Linear SVC and SVC classifiers weretested with the dataset. Table 2 shows the results for OVO mapping using Linear SVCclassifiers and Table 3 shows the results for OVA mapping using SVCclassifiers

Table 2: Resultsfor OVO mapping using Linear SVCclassifiers

Classifier	k-fold	Vectorization andTFIDF				Vectorization			
		OVO				OVO			
		Precision	Recall	F-measure	Accuracy	Precision	Recall	F-measure	Accuracy
Linear SVC	n1	0.880	0.880	0.880	0.883	0.910	0.900	0.900	0.900
	n2	0.670	0.650	0.650	0.650	0.690	0.670	0.670	0.667
	n3	0.870	0.850	0.850	0.850	0.830	0.820	0.820	0.817
	n4	0.800	0.800	0.790	0.800	0.730	0.730	0.730	0.730
	n5	0.710	0.700	0.700	0.700	0.750	0.720	0.720	0.717
	n6	0.750	0.730	0.730	0.733	0.750	0.730	0.730	0.733
	n7	0.790	0.780	0.780	0.783	0.740	0.730	0.730	0.733
	n8	0.690	0.670	0.670	0.667	0.660	0.630	0.640	0.633
	n9	0.700	0.700	0.690	0.700	0.680	0.670	0.660	0.667
	n10	0.780	0.750	0.740	0.750	0.730	0.720	0.710	0.717
Average		**0.764**	**0.751**	**0.748**	**0.752**	**0.747**	**0.732**	**0.731**	**0.731**

Table 3: Resultsfor OVO mapping using SVC

Classifier	k-fold	Vectorization andTFIDF				Vectorization			
		OVO				OVO			
		Precision	Recall	F-measure	Accuracy	Precision	Recall	F-measure	Accuracy
SVC	n1	0.880	0.870	0.860	0.867	0.910	0.900	0.900	0.900
	n2	0.660	0.630	0.640	0.633	0.710	0.670	0.670	0.667
	n3	0.830	0.800	0.800	0.800	0.830	0.820	0.810	0.817
	n4	0.780	0.780	0.780	0.783	0.720	0.720	0.710	0.717

March – April 2020
ISSN: 0193-4120 Page No. 6463 - 6470

	n5	0.700	0.680	0.680	0.683	0.720	0.680	0.690	0.683
	n6	0.780	0.770	0.770	0.767	0.770	0.730	0.740	0.733
	n7	0.760	0.750	0.750	0.750	0.730	0.730	0.730	0.733
	n8	0.710	0.700	0.700	0.700	0.700	0.670	0.670	0.667
	n9	0.730	0.720	0.710	0.717	0.660	0.650	0.640	0.650
	n10	0.780	0.750	0.750	0.750	0.720	0.700	0.690	0.700
Average		**0.761**	**0.745**	**0.744**	**0.745**	**0.747**	**0.727**	**0.725**	**0.727**

The same dataset wasused for OVA mapping between Linear SVC and SVC. The full results are reported in Table 4 and Table 5.

Table 4: Resultsfor OVA mapping using Linear SVC classifiers

Classifier	k-fold	Vectorization andTFIDF				Vectorization			
		OVA				OVA			
		Precision	Recall	F-measure	Accuracy	Precision	Recall	F-measure	Accuracy
Linear SVC	n1	0.900	0.880	0.880	0.883	0.900	0.880	0.880	0.883
	n2	0.720	0.670	0.670	0.667	0.700	0.650	0.660	0.650
	n3	0.830	0.820	0.810	0.817	0.830	0.820	0.810	0.817
	n4	0.800	0.800	0.790	0.800	0.750	0.750	0.750	0.750
	n5	0.730	0.720	0.720	0.717	0.700	0.680	0.690	0.683
	n6	0.730	0.720	0.720	0.717	0.700	0.680	0.690	0.683
	n7	0.800	0.780	0.790	0.783	0.810	0.780	0.780	0.783
	n8	0.680	0.670	0.670	0.667	0.730	0.700	0.710	0.700
	n9	0.670	0.670	0.660	0.667	0.650	0.630	0.630	0.633
	n10	0.770	0.750	0.750	0.750	0.740	0.720	0.700	0.717
Average		**0.763**	**0.748**	**0.746**	**0.747**	**0.751**	**0.729**	**0.730**	**0.730**

Table 5: Resultsfor OVA mapping using SVC

Classifier	k-fold	Vectorization andTFIDF				Vectorization			
		OVA				OVA			
		Precision	Recall	F-measure	Accuracy	Precision	Recall	F-measure	Accuracy
SVC	n1	0.820	0.820	0.810	0.817	0.817	0.850	0.850	0.850
	n2	0.720	0.680	0.690	0.683	0.690	0.650	0.650	0.650
	n3	0.840	0.820	0.810	0.817	0.810	0.800	0.790	0.800
	n4	0.760	0.770	0.760	0.767	0.750	0.750	0.750	0.750
	n5	0.710	0.700	0.700	0.700	0.740	0.730	0.730	0.733
	n6	0.800	0.780	0.790	0.783	0.770	0.770	0.760	0.767
	n7	0.790	0.770	0.770	0.767	0.780	0.770	0.760	0.767
	n8	0.650	0.650	0.650	0.650	0.710	0.680	0.690	0.683

	n9	0.670	0.670	0.660	0.667	0.690	0.680	0.680	0.683
	n10	0.740	0.720	0.710	0.717	0.750	0.730	0.720	0.733
Average		**0.750**	**0.738**	**0.735**	**0.737**	**0.751**	**0.741**	**0.738**	**0.742**

The use of Tf-idf as a feature extraction really benefited the classifier in classifying the class correctly. The Linear SVC using OVO produced the higher F-measure and accuracy if compared to other classifiers. OVO mapping and Linear SVC were used for the experiment in the second phase.

Table 6: Results for Naïve Bayes, KNN and Linear SVC using Vectorization and Tf-idf

Classifier	**Precision**	**Recall**	**F-measure**	**Accuracy**
Naïve Bayes	0.7660	0.7510	0.6811	0.6856
KNN	0.7100	0.6822	0.5940	0.6217
Linear SVC	0.6130	0.6010	0.7480	0.7370

Naïve Bayes, KNN and Linear SVC from Scikit-learn library were tested and the results are listed in Table 6. Naïve Bayes produceda higher precision and recall but in terms of F1-measure and accuracy, Linear SVC with OVO mapping showed the best result compared to the others.

V. CONCLUSION

Even though OVO and OVA are binary classifiers, they can also be used to classify multi class problems. The use of OVO for SVM classifierscan help to improve the classifier results. When compared to well-known classifiers (Naïve Bayes, KNN) for question classification, Linear SVC gives the best result among others. In future, this classifier will be used to classify questions in the Malay language. The constraint for this study is that the classifier only supports structured and essay questions, and does not support multiple choice of questions.

Acknowledgements

This work was funded by Universiti Pendidikan Sultan Idris (Malaysia) and supported by Grant. No.2017-0315-107-01

References

[1] Taqi, M. K., & Ali, R. (2016). Automatic question classification models for computer programming examination: A systematic literature review. Journal of Theoretical and Applied Information Technology, 93(2), 360–374. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85001977454&partnerID=40&md5=aca22582c7f2b1b0ce94317836425b7c

[2] Walsh, R., Bowes, J., & Sweller, N. (2017). Why would you say goodnight to the moon? Response of young intellectually gifted children to lower and higher order questions during storybook reading. Journal for the Education of the Gifted, 40, 220-246.

[3] Ullah, Z., Lajis, A., Jamjoom, M., Altalhi, A. H., Shah, J., & Saleem, F. (2019). A Rule-Based Method for Cognitive Competency Assessment in Computer Programming Using Bloom's Taxonomy. IEEE Access, 7, 64663–64675. doi:10.1109/access.2019.2916979

[4] B. S. Bloom, M. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, "Taxonomy of educational objectives: Handbook I: Cognitive domain". New York: David McKay, 1956, 19, 56.

[5] Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

[6] Abduljabbar, D. A., & Omar, N. (2015). Exam questions classification based on Bloom's taxonomy cognitive level using classifiers combination. Journal of Theoretical and Applied Information Technology, 78(3), 447–455. Retrieved from https://www.scopus.com/inward/record.uri?e

id=2-s2.0-84940667147&partnerID=40&md5=0b7f8c78cb54652c54bc18e8b9ae2e01

[7] X. Li, and D. Roth, "Learning question classifiers: the role of semantic information". Natural Language Engineering, 12(03), 2006, 229-249.

[8] Z. Hui, J. Liu and L. Ouyang, "Question Classification Based on an Extended Class Sequential RuleModel". The 5th International Joint Conference on Natural Language Processing,

[9] X. H. Phan, L. M. Nguyen, and S. Horiguchi, "Learning to classify short and sparse text web with hidden topics from large-scale data collections". In Proceedings of the 17th international conference on World Wide Web, 2008, pp. 91-100. ACM.

[10] J. Wang, L. Li and F. Ren, "An Improved Method of Keywords Extraction Based on Short Technology Text". International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE). 2010, pp. 1-6

[11] Yahya, A. A., Osman, A., Taleb, A., & Alattab, A. A. (2013). Analysing the Cognitive Level of Classroom Questions Using Machine Learning Techniques. Procedia - Social and Behavioural Sciences, 97, 587–595. https://doi.org/https://doi.org/10.1016/j.sbspro.2013.10.277

[12] Supriyanto, C., Yusof, N., Nurhadiono, B., & Sukardi. (2013). Two-level feature selection for Naive Bayes with kernel density estimation in question classification based on Bloom's cognitive levels. 2013 International Conference on Information Technology and Electrical Engineering (ICITEE), 237–241. https://doi.org/10.1109/ICITEED.2013.6676245

[13] Abduljabbar, D. A., & Omar, N. (2015). Exam questions classification based on Bloom's taxonomy cognitive level using classifiers combination. Journal of Theoretical and Applied Information Technology, 78(3), 447–455. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84940667147&partnerID=40&md5=0b7f8c78cb54652c54bc18e8b9ae2e01

[14] Kusuma, S. F., Siahaan, D., & Yuhana, U. L. (2016). Automatic Indonesia's questions classification based on bloom's taxonomy using Natural Language Processing a preliminary study. 2015 International Conference on Information Technology Systems and Innovation, ICITSI 2015 - Proceedings. https://doi.org/10.1109/ICITSI.2015.7437696

[15] Sangodiah, A., Ahmad, R., & Ahmad, W. F. W. (2017). Taxonomy based features in question classification using support vector machine. Journal of Theoretical and Applied Information Technology, 95(12), 2814–2823. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85021700847&partnerID=40&md5=8ddbf6efb7b0f9a8cb4a6f7839848194

[16] Mohammed, M., & Omar, N. (2018). Question classification based on Bloom's Taxonomy using enhanced TF-IDF. International Journal on Advanced Science, Engineering and Information Technology, 8(4–2), 1679–1685. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85055311927&partnerID=40&md5=785f86b9e1de767aefd6015f6ded80bd

[17] Anekboon, K. (2019). Feature selection for bloom's question classification in Thai Language. Advances in Intelligent Systems and Computing, 858, 152–162. https://doi.org/10.1007/978-3-030-01174-1_12

[18] Yahya, A. A., Osman, A., & El-Bashir, M. S. (2017). Rocchio algorithm-based particle initialization mechanism for effective PSO classification of high dimensional data. Swarm and Evolutionary Computation, 34, 18–32. https://doi.org/https://doi.org/10.1016/j.swevo.2016.11.005

[19] Haris, S. S., & Omar, N. (2012). A rule-based approach in Bloom's Taxonomy question classification through natural language processing. 2012 7th International Conference on Computing and Convergence Technology (ICCCT), 410–414.

[20] Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's*

taxonomy of educational objectives. New York: Longman.
[21] Jayakodi, K., Bandara, M., Perera, I., & Meedeniya, D. (2016). WordNet and cosine similarity-based classifier of exam questions using bloom's taxonomy. International Journal of Emerging Technologies in Learning, 11(4), 142–149. https://doi.org/10.3991/ijet.v11i04.5654
[22] Yahya, A. A. (2019). Swarm intelligence-based approach for educational data classification. Journal of King Saud University - Computer and Information Sciences, 31(1), 35–51. https://doi.org/https://doi.org/10.1016/j.jksuci.2017.08.002
[23] Ryan Rifkin, Aldebano Klautau. *In Defense of One-Vs-All Classification.* Journal of Machine Learning Research 5 (2004) 101-141.
[24] Yahya, Anwar. (2011). Bloom's Taxonomy Cognitive Levels Data Set. 10.13140/RG.2.1.4932.3123.
[25] Alex J. & Scholkopf S. B. Statistics and computing archive Volume 14 Issue 3, August 2004, 199-222.

Biographies

Suliana Sulaiman received her degree in computer science (artificial intelligence) from the University of Malaya, Malaysia in 2003. She received her M.Sc. degree from Universiti Kebangsaan Malaysia in 2008. She obtained her PhD degree from Universiti Kebangsaan Malaysia, Malaysia in 2013. Currently, she is a senior lecturer at Universiti Pendidikan Sultan Idris, Malaysia.

Rohaidah Abdul Wahid received her degree in information technologies (artificial intelligence) from Universiti Utara Malaysia, Malaysia in 2000. She received her M.Sc degree from Universiti Teknologi Malaysia, Malaysia in 2005. Currently, she is a lecturer at Universiti Pendidikan Sultan Idris, Malaysia.

Asma Hanee Ariffin received her degree in information technologies (artificial intelligence) from Universiti Utara Malaysia, Malaysia in 2000. She received her M.Sc degree from Universiti Utara Malaysia, in 2004. Currently, she is a lecturer at Universiti Pendidikan Sultan Idris, Malaysia.

Che Zalina Zulkifli is an Associate Professor in Computing Department, Faculty of Arts, Computing and Creative Industry at Sultan Idris Education University, Malaysia. She had over 20 years professional teaching experience as an academia and her research projects have been collaborated with multinational company and Seberang Perai City Council which contributes to a network that lead to new ideas and concrete research project. The developed automation projects that focused on Sensor Monitoring, Embedded System, Software, IoT and Wireless Communication fields have been successfully adopted by the industry to date. A total of more than a million Ringgit has been generated as an income to the University mainly from the Research grant, Commercialization of research innovative products and also the services as a principle consultant. Expertise in the agriculture sector with new invention to improve the crop production adopted high technology. She sincerely dedicated to the very wise in the green project about composting and reusage of waste. To date, she has won several international awards and national award based on her patented innovation.