

Smart Relationship Development in Community Based Marketing using Machine Learning

Handojo H. Triyanto^{a,1}, Akhmad Unggul P.^{b,2}, Arko Djajadi^{c,3}, Handri Santoso^{d,4}

^{a,d} Master's Degree of Information Technology, Pradita University, Tangerang

^b Institut Informatika dan Bisnis Darmajaya

^c Magister Teknologi Informasi, Universitas Raharja

e-mail: handojo.triyanto@gmail.com^{*1}, handri.santoso@pradita.ac.id⁴

ABSTRACT

Business world competition has compelled customer relationships to become a critical success factor. As new and advanced technologies drive digital transformation in many industries, there is an opportunity to leverage machine learning methods for customer relationship areas. The opportunity also covers the implementation for community-based marketing. The purpose of this research is to solve relationship problems between company and targeted segments or community. The quality and quantity of relationships are analyzed by using machine learning methods. The analysis is performed to customer interaction data collected cumulatively from event to event. The specific problems range from customer profiling to customer prediction, then to customer understanding categories. Finally, the research results in several optimal models for the identified problems such as: profiling model with agglomerative clustering, prediction model with artificial neural networks, and understanding model with simple RNN (Recurrent Neural Network). The novelty of this research includes the combination of 3 (three) things, namely: "relationship development", "community-based marketing", and "machine learning usage". The research also confirms technology's role in supporting businesses.

Keywords : artificial intelligence, customer relationship management, machine learning, marketing

Article Info :

Submitted: 16 November 2021

Reviewed: 20 Desember 2021

Accepted : 17 Januari 2022

Copyright © 2022 – CSRID Journal. All rights reserved.

1. INTRODUCTION

As competition in the business world is increasingly fierce, all companies strive to improve their marketing capability. The marketing capability includes new customer potentials and existing customers (relating to up-sell and cross-sell). Moreover, for existing customers, all companies strive to keep relationships with their customers for brand awareness and loyalty.

According to the literature [3], relationship concept refers to relationship between company and customers on an individual level, not in group or segment levels. This is different from the traditional marketing concept, which looks at the targeted market as a group or segment. In the current business world, typically relationship developments are performed only to existing customers.

When relationship development is implemented to a targeted market, there is a need to use community concepts. The concept is used to make identification and management easier. The intended community is a part of the targeted market where the community members can be identified and managed. The objective is to make relationship development at the individual customer level possible.

The previous research has focused on existing customers. On the other hand, this research focuses on the relationship development with community members, which includes existing customers and non-customers (customer potentials or leads or prospects). The novelty of this

research includes the combination of 3 (three) things, namely: “relationship development”, “community-based marketing”, and “machine learning usage”.

The purpose of this research is to solve relationship problems between company and targeted segments or community. Further, the research objectives can be defined as follow: (i) developing an analysis model for customer profiling capability; (ii) developing an analysis model for customer prediction capability; (iii) developing an analysis model for customer understanding capability.

Several important journals and relating to this research are as follow: (i) Customer relationship management and big data enabled: Personalization & customization of services [7], this journal covers Social CRM and the data usage in individual levels of customers by leveraging Big Data technology. (ii) An improved K-means clustering algorithm for fish image segmentation [10], the journal covers a combination of K-Means clustering segmentation algorithm and mathematical morphology. It includes a modification to improve the traditional K-Means model, specifically in determining optimal numbers of clusters and positions of clusters' centroids. (iii) Research on social data by means of cluster analysis [8], the journal covers an implementation of unsupervised learning (clustering analysis) to a social data set for differentiating profiles within. (iv) Deep belief networks and cortical algorithms: A comparative study for supervised classification [9], the journal covers supervised learning, namely Artificial Neural Networks with feedforward-related Deep Learning. The Deep Learning algorithms are Deep Belief Networks (DBN) and Cortical Algorithms (CA). (v) Semiautonomous methodology to validate and update customer needs database through text data analytics [6], the journal covers text analytics from survey results of customer needs and compares automatically with information of customer needs that the company already has. (vi) Text mining analysis of teaching evaluation questionnaires for the selection of outstanding teaching faculty members [5], the journal covers an implementation of textual sentiment analysis to process questionnaire results. Several kinds of classifiers are used, then their performances are compared.

2. METHOD

The research methodology used in this research basically adopts the data science pipeline [11]. The flow diagram of this research is shown in Figure 1.

This research involves implementations of concept and technique from several knowledge fields or disciplines. Those fields are as follow:

A. Customer Relationship Management (CRM)

According to the literature [3], the relationship between customer and enterprise provides a framework for all other things related to the customer value business model, even when the customer is not aware that that relationship exists. Customer Relationship Management (CRM is a relationship management with customers performed by an enterprise. Customer experience is all things experienced by customers as a result. There are 4 (four) processes for CRM, namely: identify, differentiate, interact, and customize. The relationship between this research and those processes are: (i) Identify: this research involves the data collection about community member information (the information can be categorized into 2 (two) groups, namely: customer profile (name, email, handphone number, company, department, etc.) and customer interaction (response, attendance, etc.)); (ii) Differentiate: this research covers the analysis process to each community member. Examples: customer profiling, customer prediction, etc; (iii) Interact: this research only relates to the interaction-related data collection, but not covering the interaction design and execution; (iv) Customize: this research only relates to the analysis to understand the community members, but not covering the design of customized products/services.

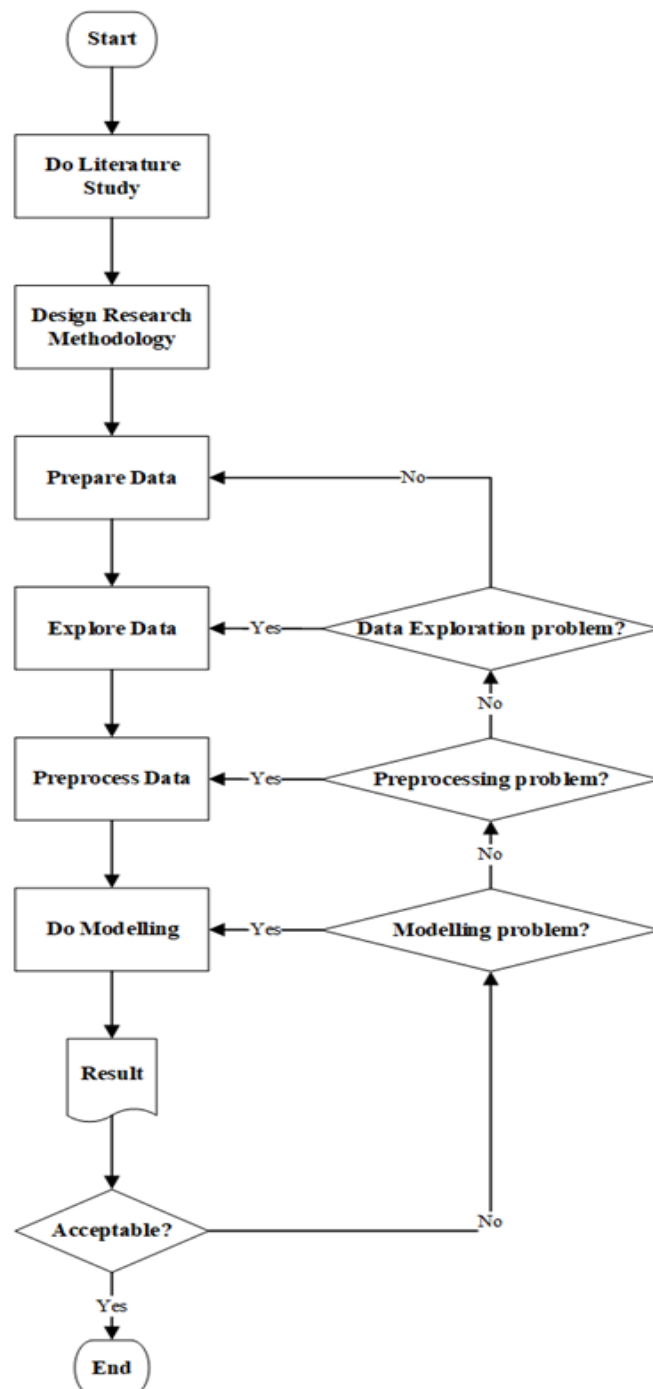


Figure 1. Flow Diagram of Research Methodology

B. Data Science

According to the literature [11], data science is the interdisciplinary field that deals with the creation of insights or data products from a given set of data files (usually in unstructured form), using analytics methodologies. The data it handles is often what is commonly known as “big data,” although it is often applied to conventional data streams, such as the ones usually encountered in the databases, the spreadsheets, and the text documents of a business. The whole process of turning data into insights and data products is not at all straight-forward. In fact, it’s more of an iterative process, with impromptu loops and unexpected situations causing delays and re-evaluations of assumptions. That’s why it often says as the data science pipeline, a complex process comprised of several inter-dependent steps, each bringing power users closer to the end result, be it a set of insights to hand off to decision makers, or a data product for end-users. This research adopts the

data science process, which is organized in three general stages: data engineering, data modelling, and information distillation.

C. Machine Learning

According to the literature [1], machine learning is the science (and art) of programming computers so they can learn from data. This research uses unsupervised and supervised learnings. For unsupervised learning it includes several algorithms such as: K-Means, DBSCAN, Gaussian Mixtures Model (GMM), Agglomerative Clustering (Hierarchical Clustering),

For supervised learning it includes several algorithms such as: k-Nearest Neighbors (kNN), Decision Tree (Tree), Support Vector Machines (SVMs), Stochastic Gradient Descent (SGD), Random Forest, Naïve Bayes, Logistic Regression, Neural networks.

Part of Natural Language Processing (NLP) or text analytics that is also under supervised learning (Deep Learning) category, this research includes several algorithms such as: Artificial Neural Networks (ANN), SimpleRNN (Recurrent Neural Network), Long-short Term Memory (LSTM), Gated Receiver Unit (GRU).

D. Cloud Computing

According to the literature [2], cloud is a collection of services that helps developers focus on their project rather than on the infrastructure that powers it. The advantage of cloud hosting is flexibility. It is very suitable for a situation when the need or requirement of computing size is not yet identified or unknown. Likewise, cloud hosting covers maintenance services so that minimizing works of providing infrastructure for a company's application software. This research also looks at the design possibilities of cloud and other applications usages.

The research has type “Applied”, which has purpose to implement existing concepts, tools, techniques, and technology in optimal ways within solution design of predetermined problems.

3. RESULTS AND DISCUSSION

This research uses a case study from a company called PT. ABC (not the real name). The company has business in the technology market research services area (the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets). PT. ABC has products/services such as subscription, consulting, Go to Market Services (GMS), and events. This research only relates to the events product/service with the focus on the analysis capability using Machine Learning.

The events product has a unique revenue model since PT. ABC gets revenue from third parties, namely IT vendors or providers as sponsor of the event conducted by PT. ABC. The conducted event is free to IT and business professionals (seen as community members), but only by invitation. PT. ABC has a role as a relationship development firm on behalf of all IT providers in the market or industry.

To identify the relevant and specific problems from the company, the previous methodology discussed in the previous section is implemented. The result of specific problems identification is as follow: (i) to know the list of community members for each category (based on profile data value and interaction value); (ii) to predict the possibility of event attendance for each invited community member; (iii) to categorize event topics expected by each community member into the existing categories.

The first problem above is a customer profiling problem and is called capability 1. The second problem above is a customer prediction problem called capability 2. The third problem is a

customer understanding problem called capability 3. Figure 2 shows 3 (three) models as output of the modelling process.

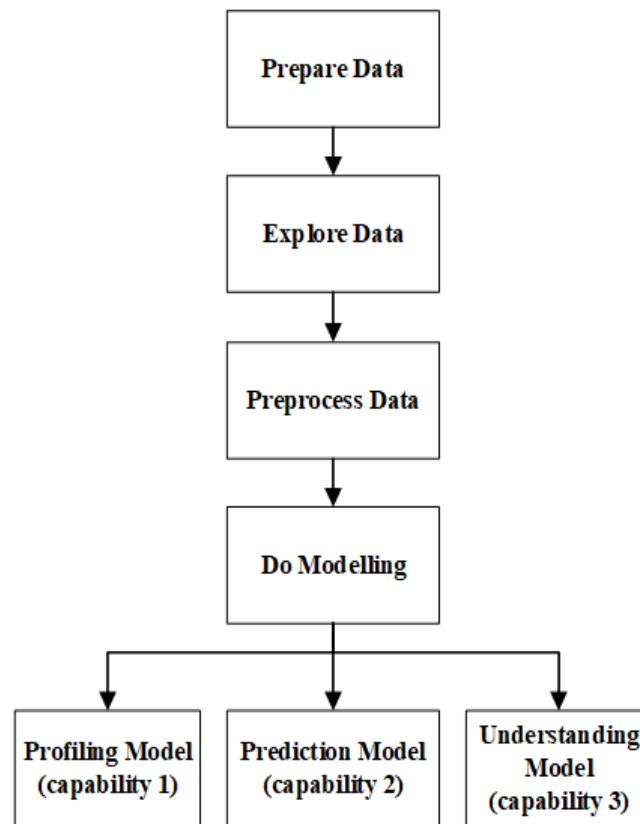


Figure 2. Flow Diagram Resulting in The Models

The relevant data from PT. ABC are collected by doing some cleansing and formatting process. The data size (the number of records) for capability 1 and 2 is 178 community members. The data size for capability 3 is 67 topics and 5 categories. Table 1 and 2 shows the data variables or features/attributes.

Table 1. Data for Capability 1 and 2

No.	Variable	Description
1	ID	Number of community member for recording purpose.
2	Name	Name of community member.
3	Email	Email address of community member.
4	Handphone	Handphone number of community member.
5	LinkedIn	LinkedIn address of community member.
6	Client	Whether company of community member is PT. ABC's client or not. Value = 1 for client; 0 for non-client.
7	Company	Company name of community member.
8	Subindustry	Company subindustry of community member. Example: bank, insurance, etc.
9	Tier	Company size category of community member. Example: tier 1, tier 2, tier 3.
10	Department	Department of community member in his/her company.
11	Level	Job level of community member. Example: C-level, division head, manager, etc.
12	Last Update	Duration in months since the data last update until now.

13	Event1	Response level (value) of first event interaction.
14	Event2	Response level (value) of second event interaction.
15	Event3	Response level (value) of third event interaction.
16	Event4	Response level (value) of fourth event interaction.
17	Event5	Response level (value) of fifth event interaction.
18	Event6	Response level (value) of sixth event interaction.
19	Award	Response level (value) of award program participation conducted by PT. ABC.
20	Research	Response level (value) of research program participation conducted by PT. ABC.
21	Sharing	Response level (value) of sharing session program acceptance conducted by PT. ABC.
22	Attended	Attendance/participation of member community in last interactions conducted by PT. ABC. Value = 1 for attending/participating; 0 for not attending/participating.

Table 2. Data for Capability 3

No.	Variable	Description
1	Topic	Expected topic by event participant/audience for next events.
2	Category	Category according to PT. ABC for the expected topics from the event participants.

Several visual analysis processes to explore the data are performed to understand the data and their correlations. The objective is to define features for each problem or capability. Afterwards, data pre-processing is conducted to make the data ready for modelling. The process includes new derivative features creation and formatting. Table 3 shows the data features or attributes after pre-processing.

Table 3. Additional Data After Pre-Processing

No.	Variable	Description
1	Handphone1	Conversion of handphone data. Value = 0 if no data; 1 if data available.
2	LinkedIn1	Conversion of linkedin data. Value = 0 if no data; 1 if data available.
3	Client1	Conversion of client data. Value = 0 if no data; 1 if data available.
4	Company1	Conversion of company data. Value = 0 if no data; 1 if data available.
5	Subindustry1	Conversion of subindustry data. Value = 0 if no data; 1 if data available.
6	Tier1	Conversion of tier data. Value = 0 if no data; 1 if data available.
7	Department1	Conversion of department data. Value = 0 if no data; 1 if data available.
8	Level1	Conversion of level data. Value = 0 if no data; 1 if data available.
9	N_data	Addition result from all data values of handphone1, linkedin1, client1, company1, subindustry1, tier1, department1, and level1.
10	N_interaction	Addition result from all data values of event1, event2, event3, event5, event6, award, research, and sharing.

Customer Profiling or capability 1 problem involves 2 (two) features as shown in Table 4. The result from applying several clustering algorithms to the above features are as below:

- i. K-Means: the algorithm depends on the initial positions of centroids so that sometimes it does not result in the expected clustering results.
- ii. DBSCAN: the algorithm depends on the distance of data points so that the clustering cannot be fully controlled by the user.

- iii. Gaussian Mixture Model (GMM): the algorithm depends on the choice of covariance_type so that the clustering cannot be fully controlled by the user.
- iv. Hierarchical Agglomerative Clustering: the algorithm depends on the distance of data points and the hierarchy of clusters built from bottom up. This algorithm results in the expected clustering results.

Table 4. Data for Capability 1

No.	Feature/Variable	Format
1	N_data	Integer
2	N_interaction	Integer

From the above modelling results the best clustering algorithm for the problem is Hierarchical Agglomerative Clustering. The reasons why the algorithm is the best are:

- i. The resulted clusters fulfill the user's need of analysis. The model results in 6 (six) clusters that are useful for user to continue with the next best actions. Those clusters are as following: Cluster 1: the community members with their identification and interaction data values are highest; Cluster 2: the community members with identification data value is high, but interactions data value is relatively low (compared to cluster 1); Cluster 3: the community members with identification data value is medium and interaction data value is high; Cluster 4: the community members with identification data value is medium and interaction data value is low; Cluster 5: the community members with identification data value is low and interaction data value is high; Cluster 6: the community members with identification and interaction data values are low.
- ii. The configuration is simpler or easier. To setup the model using the algorithm is relatively easier than the other algorithms.

Customer Prediction or capability 2 problem involves 14 variables (13 features + 1 target) as shown in Table 5. The problem is categorized as binary classification problem.

Table 5. Data for Capability 2

No.	Feature/Variable	Format
1	Handphone1	Boolean
2	LinkedIn1	Boolean
3	Last Update	Integer
4	Event1	Integer
5	Event2	Integer
6	Event3	Integer
7	Event4	Integer
8	Event5	Integer
9	Event6	Integer
10	Award	Integer
11	Research	Integer
12	Sharing	Integer
13	N_interaction	Integer
14	Attended	Boolean

The training and validation are performed by implementing cross-validation 80:20 for each of 8 (eight) pre-determined algorithms. There are 4 (four) performance metrics calculated. AUC is the main metric as the data is relatively balanced for two classes. The result from applying each classification problem algorithm to the above variables are shown in the Table 6.

Tabel 6. Performance Metrics of Capability 2 Modelling

Model	AUC	FI	Precision	Recall
kNN	0.711	0.662	0.662	0.663
Tree	0.698	0.688	0.688	0.691
SVM	0.713	0.733	0.782	0.753
SGD	0.697	0.719	0.749	0.736
Random Forest	0.703	0.741	0.746	0.747
Neural Network	0.743	0.733	0.759	0.747
Naïve Bayes	0.718	0.744	0.756	0.753
Logistic Regression	0.687	0.694	0.709	0.708

From above results the temporary or initial best classification algorithm for the problem is Neural Network because the AUC value of that algorithm is highest than other algorithms.

The next action is to do further model development of those ANN (Artificial Neural Network). The further model development includes steps to determine optimal factors or parameters. Several model parameters are determined based on literature study [1] and [4] as follow:

- Hidden Layer Activation Function: ReLU (Rectified Linear Unit) [1] and [4].
- Output Layer Activation Function: Sigmoid [1] and [4].
- Optimizer: RMSprop (Root Mean Square Propagation) [4].
- Loss: binary_crossentropy [4].

The other factors and parameters are determined by performing steps as follow:

- Features extraction. The objective of this step is to get features that have more impact or contributions to prediction capability. Meanwhile, it also has an objective to remove or reduce features that have a little impact or contributions to prediction capability. The final objective is to get better prediction results. One of the ways is using PCA (Principal Component Analysis) method. The method does not remove a little impact features at all, but only reducing the impact. That condition is not a problem because the features selection has been done in the stage before the modelling. Likewise, certain features would probably have more impact/contributions in the future when the data increased or more complete. The advantages of PCA method are: (i) the reduction is performed on the number of components, not directly on original features. Those removed components have automatically had the less information from the original features; (ii) the computation duration is faster when the data becomes larger in the future. The determination of the number of components is performed by referring to the best validation AUC from 3 (three) possibilities:
 - Original features information is 100% (the number of components 13)
 - Original features information is 99% (the number of components 11)
 - Original features information is 95% (the number of components 9)

Referring to the literature [4], the training and validation processes are performed with the initial parameters as follow: Hidden layer = 2; Neuron per hidden layer = 16; Epoch = 200. The result shows that the number of components 11 have the highest validation AUC value, namely 83.97%. Even though the reduction with the number of components lower than 9 has possibilities of higher validation AUC values., the information reduction in modelling would reduce the prediction capability in the future.

- ii. Setup of the number of optimal epochs. The objective of this step is to determine the minimum number of epochs where the loss value and accuracy value are relatively not changed. Typically, the higher number of epochs, the lower training loss value and the higher training accuracy value. The training process is performed with initial parameter as follow: Hidden layer = 2; Neuron per hidden layer = 16; Component = 11 (following the determination of previous step). The resulting graphic for epoch vs training loss shows that the minimal number of epochs are 125 epochs. The resulting graphic for epoch vs training accuracy shows that the minimum number of epochs are 150 epochs. Therefore, the higher number, namely 150 epochs are taken as the optimal number of epochs.
- iii. Setup of regularization or dropout. The objective of this step is to prevent overfitting. From the result of training accuracy, it is 85.5%, which is still lower than 90%. Then, the model is still underfitting, not overfitting. Therefore, it does not need to do regularization or dropout.
- iv. Configuration of the number of hidden layers and the number of neurons per hidden layer. The objective of this step is to determine the number of hidden layers and the number of neurons per hidden layer where the value of validation AUC is maximum. The result shows that the number of hidden layers 2 (two) and the number of neurons per hidden layer 128 have the highest validation AUC value. Table 7 shows the result of what-if analysis.

Table 7. What-if Analysis Result

No.	Hidden Layer	Neuron per Hidden Layer	Validation
1	2	16	84.9%
2	2	32	85.9%
3	2	64	85.1%
4	2	128	86.5%
5	2	256	86.3%
6	2	512	85.8%
7	2	1024	85.0%
8	3	16	84.9%
9	3	32	85.7%
10	4	16	84.6%

The best AUC value from the Table 7 is 86.5%. That result is still acceptable with some reasons as below: (i) already performed cross-validation and fine-tuning to all parameters; (ii) several factors affecting the prediction result have high complexity, namely: human factor or relationship factor is unstable; other affecting factors (those factors can be developed further as the data are increased in the future); level of reliability from event/interaction operation affects the relationship level with community members (PT. ABC can improve its process in the future to increase the reliability of event/interaction operation). Figure 3 shows the system diagram of training process for capability 2.

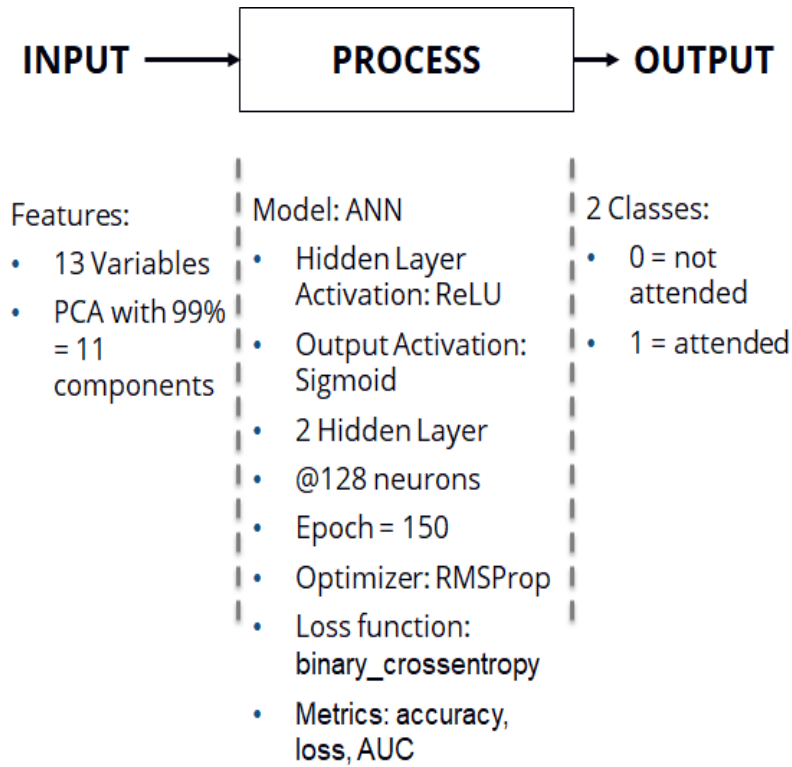


Figure 3. System Diagram of Training Process for Capability 2

Customer Understanding or capability 3 problem involves 2 (two) variables (1 feature + 1 target) as shown in Table 8. The result from applying each algorithm to the above variables are shown in the Table 9. With epoch = 45 the result of validation loss and accuracy for each model are shown in Table 10. With accuracy = 100% (except for LSTM) the result of validation loss and epoch for each model are shown in Table 11.

Table 8. Data for Capability 3

No.	Feature/Variable	Format
1	Topic	String
2	Category	String

Table 9. Performance Metrics of Capability 3 Modelling (epoch = 100)

Model	Loss	Accuracy	Epoch
ANN	0.73	100%	100
SimpleRNN 1 layer	0.19	100%	100
SimpleRNN 2 layer	0.05	100%	100
GRU	0.24	100%	100
LSTM	0.44	98.5%	100

Table 10. Performance Metrics of Capability 3 Modelling (epoch = 45)

Model	Loss	Accuracy	Epoch
ANN	1.21	86.6%	45
SimpleRNN 1 layer	0.74	100%	45
SimpleRNN 2 layer	0.46	100%	45
GRU	1.09	79.1%	45
LSTM	1.22	68.7%	45

Table 11. Performance Metrics of Capability 3 Modelling (accuracy = 100%)

Model	Loss	Accuracy	Epoch
ANN	0.76	100%	96
SimpleRNN 1 layer	0.71	100%	47
SimpleRNN 2 layer	0.46	100%	45
GRU	0.31	100%	93
LSTM	0.44	98.5%	100

The modelling result has shown that SimpleRNN has the lowest validation loss with validation accuracy value 100% on epoch = 100. For GRU it has the second lowest loss with accuracy 100% on epoch = 100. For LSTM, having the highest loss value on epoch = 100.

From the above results, LSTM algorithm is not suitable for the existing problem. On the other side, SimpleRNN algorithm is the most suitable for the problem where the model has the lowest loss value with epoch 100 (compared to GRU model on the same epoch).

The problem in this research requires memory for the previous words in a phrase, but it does not need long memory like a sentence. Therefore, LSTM model, which is suitable for a long sentence, needs a longer process (more epoch) to get 100% accuracy. On the other side, ANN model (without RNN) is too simple to remember previous words. So, the most suitable model or algorithm for a phrase is SimpleRNN. Figure 4 shows the system diagram of training process for capability 3.

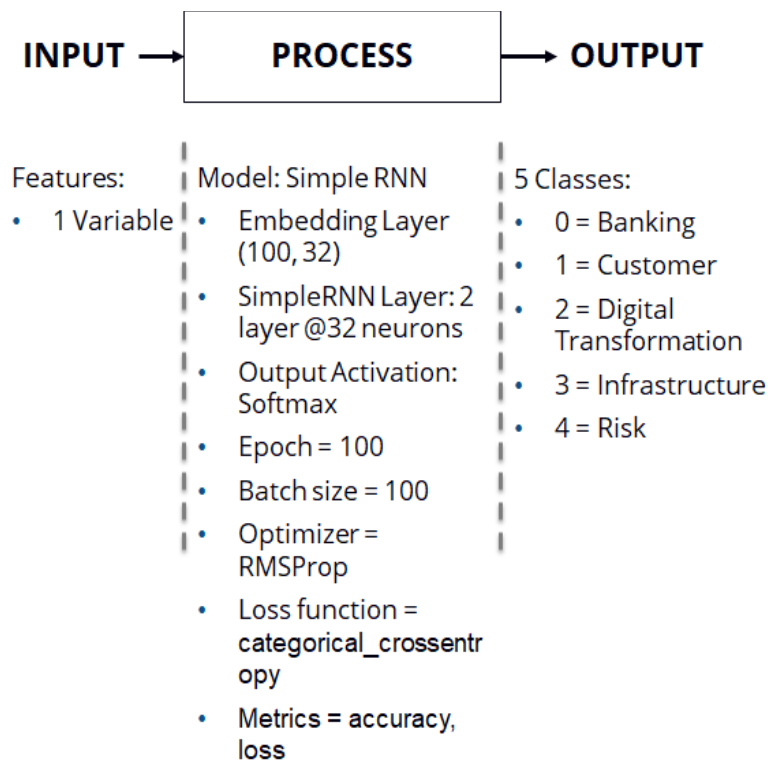


Figure 4. System Diagram of Training Process for Capability 3

4. CONCLUSION

According to the research objectives, the conclusions are as the following:

- Customer Profiling is best fulfilled by Hierarchical Clustering model, namely Agglomerative Clustering.
- Customer Prediction is best fulfilled by Artificial Neural Network model with Keras API.
- Customer Understanding is best fulfilled by Simple RNN model with 2 (two) layers.

Machine Learning is able to help the company to run community-based marketing by analyzing customers or community members in individual level (one to one relationship) at scale.

Future research can be conducted for things as the following:

- i. Detailed or technical capability of a company in performing interaction and customization for products, services, and customer experiences;
- ii. Data quality techniques of a company, e.g. error in data collection process, bad design of feedback questionnaire, low motivation from event/interaction participants in filling feedback questionnaire;
- iii. Event/interaction design evaluation of a company, e.g. event format, product/service pricing (for third party sponsors), event/interaction promotion;
- iv. Effect of competitors (other event organizers) to event conducted by a company;
- v. Usage of NLP (Natural Language Processing) for analysing texts coming from feedback that has essay format as well as it can include the NLP usage for research survey purpose.

REFERENCES

- [1] Géron, Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, Second Ed. O'Reilly Media, Inc., 2019.
- [2] J. Geewax, Google Cloud Platform in Action. Manning Publications Co., 2018.
- [3] D. Peppers and M. Rogers, Managing Customer Experience and Relationships: A Strategic Framework, Third Ed. John Wiley & Sons, Inc., 2017.
- [4] F. Chollet, Deep Learning with Python. Manning Publications Co., 2018.
- [5] C. W. Tseng, J. J. Chou, and Y. C. Tsai, "Text mining analysis of teaching evaluation questionnaires for the selection of outstanding teaching faculty members," IEEE Access, vol. 6, pp. 72870–72879, 2018, doi: 10.1109/ACCESS.2018.2878478.
- [6] A. Martí Bigorra, O. Isaksson, and M. Karlberg, "Semi-autonomous methodology to validate and update customer needs database through text data analytics," Int. J. Inf. Manage., vol. 52, p. 102073, Jun. 2020, doi: 10.1016/j.ijinfomgt.2020.102073.
- [7] M. Anshari, M. N. Almunawar, S. A. Lim, and A. Al-Mudimigh, "Customer relationship management and big data enabled: Personalization & customization of services," Appl. Comput. Informatics, vol. 15, no. 2, pp. 94–101, 2019, doi: 10.1016/j.aci.2018.05.004.
- [8] C. Maione, D. R. Nelson, and R. M. Barbosa, "Research on social data by means of cluster analysis," Appl. Comput. Informatics, vol. 15, no. 2, pp. 153–162, 2019, doi: 10.1016/j.aci.2018.02.003.

- [9] Y. Rizk, N. Hajj, N. Mitri, and M. Awad, "Deep belief networks and cortical algorithms: A comparative study for supervised classification," *Appl. Comput. Informatics*, vol. 15, no. 2, pp. 81–93, 2019, doi: 10.1016/j.aci.2018.01.004.
- [10] H. Yao, Q. Duan, D. Li, and J. Wang, "An improved K-means clustering algorithm for fish image segmentation," *Math. Comput. Model.*, vol. 58, no. 3–4, pp. 790–798, 2013, doi: 10.1016/j.mcm.2012.12.025.
- [11] Z. Voulgaris, *Data Science Mindset, Methodologies, and Misconceptions*, First Ed. Technics Publications, 2017.