❐   727

The disruptometer: an artificial intelligence algorithm for market insights

Mimi Aminah binti Wan Nordin, Dmitry Vedenyapin, Muhammad Fahreza Alghifari, Teddy Surya Gunawan
Department of Electrical and Computer Engineering, International Islamic University Malaysia, Malaysia

Article Info	ABSTRACT
	Social media data mining is rapidly developing to be a mainstream tool for marketing insights in today's world, due to the abundance of data and often freely accessed information. In this paper, we propose a framework for market research purposes called the Disruptometer. The algorithm uses keywords to provide different types of market insights from data crawling. The preliminary algorithm data-mines information from Twitter and outputs 2 parameters-Product-to-Market Fit and Disruption Quotient, which is obtained from a brand's customer value proposition, problem space, and incumbent space. The algorithm has been tested with a venture capitalist portfolio company and market research firm to show high correlated results. Out of 4 brand use cases, 3 obtained identical results with the analysts 'studies.

Corresponding Author:

Mimi Aminah binti Wan Nordin,
Department of Electrical and Computer Engineering,
International Islamic University Malaysia,
Jalan Gombak, 51300 Selangor, Malaysia.
Email: mimie@iium.edu.my

1. INTRODUCTION

Innovation, which is the heart of business, is especially crucial for ambitious start-ups in the early stage. Among the various types of innovation is the disruptive innovation-innovation that creates a new market by providing a different set of values, which ultimately (and unexpectedly) overtakes an existing market [1]. The theory of disruptive innovation describes how relatively simple, convenient and low-cost innovations can be useful to the growth of companies, even with the presence of strong competitors in the industry [2]. For a company to invest in a particular product, however, is never a straightforward process due to the uncertainty. Marketing research must be conducted to ensure that the product can return the investments, but for a marketing research to be successful, it requires professional expertise, putting extra time, or rely on technology.

Marketing science research which has been dated back from the last 50 years [3] has now revolutionalized with the advancement of technology, the methods has integrated even more new and upcoming technological fields. A recent trend is due to the rise of big data and the utilization of data mining techniques. Data mining is a technique of extracting hidden predictive information from large databases [4]. The data mining sources can vary, however in this paper we focus on data mining of social media. Certain social media such as Twitter became a popular place to data-mine as the information is in abundance and public. Market researchers can observe the market opportunities by tapping on daily conversations on social media, although the challenge comes from the massive volume of it.

An example of Twitter data mining applied in marketing was conducted by [5]. The researchers have attempted stock market forecasting using machine learning though Twitter sentiment classification.

Although their result left more to be desired (58% accuracy), their method of extracting information via Twitter Search API is a highly generalizable method, one which is also used in the proposed framework.

Another research based on sentiment is performed by [6]. In their study, they have attempted to approach textual sentiment analysis by 2 ways –lexicon based (occurrence certain keywords) and supervised classification based. Their sentiment analysis is then applied on TripAdvisor and Amazon reviews. The paper published by [7] is a massive scale field study which studies the marketability effect of Tweets by the company and influencers onto the marketability of a broadcast. The paper concluded that tweets indeed have a significant effect on the view count, particularly if retweeted by a strong influencer. Yet another application of data mining is applied in [8], where information of a disease outbreak can be closely monitored via social network data mining. The research highlights several text-based techniques to perform data mining such as K-Nearest Neighbour (KNN), support vector machine (SVM), and neural networks.

One more field that is fast expanding is machine learning and artificial intelligence. Computer performing manual tasks has already existed from the past decade, but now they are capable of performing more sophisticated tasks such as sentiment analysis. One advantage of machine learning in market research is the ability for the computer to perform automations, replacing the need of humans to do menial laborious tasks. As an example, [9] have successfully performed stock market prediction, achieving 77% accuracy using Multi-Layer Perceptron algorithm. The research conducted in [10] also applies machine learning which targets the stock market, however approaches the task using sentiment analysis. One interesting research was conducted by [11] which proposed a way of digital marketing ultizing a special machine learning algorithm, the deep learning from three social media platforms, Facebook, Twitter and Instagram. The authors also emphasised on the usage of hashtags as a way of interacting with the customer.

While concepts of data mining and artificial intelligence have seen application in the marketing field, a handful is still in the realm of sentiment analysis. Prototypes using Twitter data mining concepts such as by [12] have been published and tested, but is still in the realms of using hashtags to detect and visualizing trends. There is still a gap between the current research studies and real application for market purposes.

Hence in this paper, we introduce a preliminary framework for utilization of artificial intelligence and data mining in market researching which is called the Disruptometer. The framework is an algorithm that crawls the web, searching for relevant information related to the business attributes. Unlike conventional marketing research methods using surveys or interviews, the Disruptometer adopts a 'fly-on-the-wall' principle and listens to social conversations. Using this method, the market research can be considered more organic when compared to asking directly. The algorithm has been tested with Kalaari capital, a venture capitalist (VC) company based in India, and verified that out of 4 test cases, 3 obtained using the Disruptometer had same result as a professional market analyser. The preliminary framework has used Twitter to obtain the information.

The rest of this paper is organized as follows. Section 2 focuses on the research methodology of the Disruptometer, firstly outlining the parameters and then the validation methodology. Section 3 outlines the results obtained from the Disruptometer by focusing on one brand, then repeating the steps for the other three, which is followed by a benchmark with results given by venture capital analysts and discussion. Finally, the last section is the conclusion which summarizes the paper, then followed by future work recommendations.

2. RESEARCH METHODOLOGY

The Disruptometer algorithm predicts whether a business proposition is feasible based on 2 categories-the Product-to-Market Fit (P-M Fit) and Disruption Quotient (D-Quotient). These two categories are determined by 3 parameters-the customer value proposition (CVP), problem space (P), and incumbent space (I) which are obtained through social media data mining. A vernacular study is performed to obtain [keywords] which expresses the three parameters. Twitter handlers or 'influencers' which expresses the [keywords] are then included to be in the particular parameter's set. Finally, a 3 tier scoring system from A to C is used, based on the number of CVP covered by either P for P-M Fit or I for D-Quotient. The general flow of the Disruptometer algorithm is shown in Figure 1. In this section, each parameter and how to obtain them is explained in more detail, followed by the validation method.
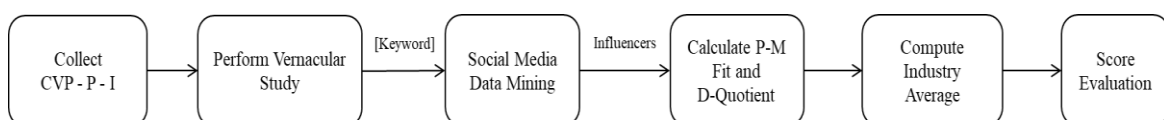

Figure 1. Disruptometer flow

2.1. Parameters
2.1.1. Customer value proposition (*CVP*)

The customer value proposition space is a set which encapsulate the users which emulates [keywords] similar to that which the business CVP is promoting. The expression searched is generally positive-oriented, either by praising the [keyword] or by proposing to solve P by offering a [keyword] related service/product.

2.1.2. Problem space (*P*)

The Problem Space are the users that express a dissatisfaction or a need for a particular product/service. The nature can be disapproval (ie the status quo solution is not satisfying the need) or a need of an entirely new product which does not exist in the market yet. Comparatively to CVP space, the expressions searched are negative-oriented, often in the form of complaints. In short, the Problem Space is a set of users who indicates "problems" around a particular service/product.

2.1.3. Incumbent space (*I*)

The Incumbent Space refers to the set of users who are the competitors to the CVP. They offer a potential solution to the Problem Space. The solution can overlap with the CVP proposed or it can be entirely different. The solution, however should be an already existing product in the market and not simply be still in an idea stage.

2.1.4. Product-to market fit (*P-M Fit*)

The Product-to-Market Fit (P-M Fit) is the forecasted value which a product fits a market demand. A high score indicates a potential strong interest in the product proposed. The P-M Fit score is obtained by calculating the Jaccard index from P and CVP.

$$PM\ Fit(CVP, P) = \frac{|CVP \cap P|}{|CVP \cup P|} \times 10^4 \qquad (1)$$

where CVP is a set that contains the users and the followers up to second degree that indicates a positive expression towards [keyword] and P is a set that contains users and followers up to second degree that express the need for [keyword].

Essentially the Product-to-Market Fit aims to answer these following questions:
a. Does the market express feelings of need/desire for the product characteristic [keyword]?
b. Has an influencer expressed a positive attribute towards a similar [keyword]?

2.1.5. Disruption quotient

The Disruption Quotient measures the potential that the CVP has to disrupt the current market space occupants. Although a product may show potential with a strong market demand, if the market has already been saturated with competitors the product may not be as successful. The Disruption Quotient is obtained by calculating the Jaccard index from CVP and I.

$$Disruption\ Quotient(CVP, I) = \frac{|CVP \cap I|}{|CVP \cup I|} \times 10^4 \qquad (2)$$

where I is a set that contain the user and the followers up to second degree that offer a solution to the problem (P).

Essentially the Disruption Quotient aims answer these following questions:
a. Has the market expressed the lack of the product characteristic [keyword] in the current competition?
b. Has the competition mentioned the product characteristic [keyword]?

The Disruptometer parameters' relationship can be summarized in Figure 2. The resulting P-M Fit and D-Quotient are then inputted to a matrix, one for each category. A sample P-M Fit Matrix with CVP (A B C), P (D E F) is shown in Table 1. For users and followers of CVP(A) which overlaps with the users and followes of P(D), their calculated P-M Fit would be inserted in cell (ad). Similiarly, for a sample D-Quotient Matrix with with CVP (A B C) and I (G H I) as shown in Table 2, the Disruption quotient of CVP(A) which overlaps with I(G) is inserted in cell (ag).
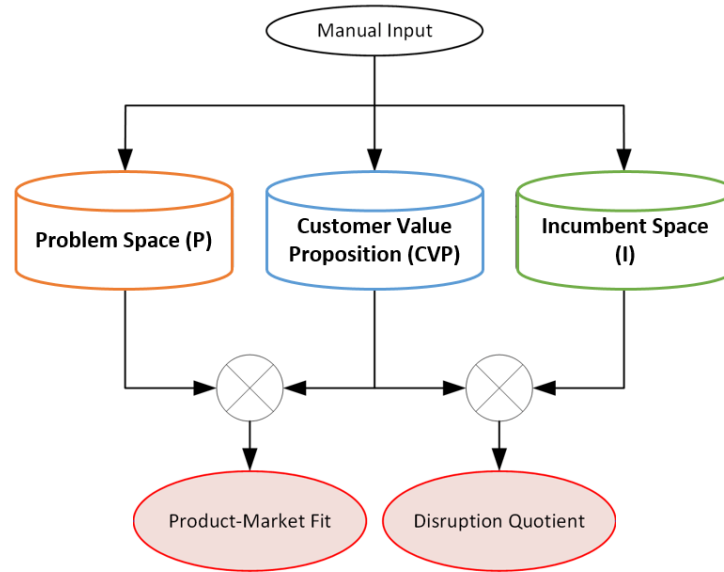
Figure 2. Disruptometer parameter overview

CVP	Problem Space		
	D	E	F
A	ad	ae	af
B	bd	be	bf
C	cd	ce	cf

Table 1. P-m fit matrix

CVP	Incumbent Space		
	G	H	I
A	ag	ah	ai
B	bg	bh	bi
C	cg	ch	ci

Table 2. D-quotient matrix

2.1.6. Industry average

To determine the industry average of the P-M Fit, we compute the summation of the P-M Fit divided by the multiplication of the number of Ps and the number of CVPs. Similarly, to compute the industry average for D-Quotient, we compute the summation of the D-Quotient divided by the multiplication of number of Is and CVPs.

$$PM\ Fit\ Industry\ Average = \frac{\sum(PM\ Fit)}{n(P) \times n(CVP)} \tag{3}$$

$$D\ Quotient\ Industry\ Average = \frac{\sum(D\ Quotient)}{n(I) \times n(CVP)} \tag{4}$$

2.1.7. Evaluation score

For each CVP in P-M Fit Matrix, values obtained above the industry average is marked as successful. A high industry average is deemed to be potentially suitable for the market. The score **A** is awarded if all CVP P-M Fit values are above industry average, **B** is given if at least half of CVP covers the P space, and C if the less than a third of the CVP covers the P space.

The same procedure is repeated for the D-Quotient Matrix, however the values obtained *below* the industry average is desirable, as it indicates low overlapping between that offered by competitors and business CVP.

2.2. Validation method

To validate the results, we compare the P-M Fit and D-Quotient with grades scored by Kalaari's analysts/fellows. For each brand, the analysts are prompted to give a score between A to C based on their experience. For the Disruptometer algorithm, each value proposition is given a grade between A to C, where A shows a high correlation for Product-to-Market Fit, and A in Disruption Quotient means low influence of competitors on potential clients, hence a high disruption potential.

3.   RESULTS AND ANALYSIS

The Disruptometer has tested its algorithm for 4 brands. The results for brand A flow is showcased below. The problem that Brand A is trying to address the current problem of shopping – human-manned shop tellers which are inefficient, long shopping queues, and a simulatious shopping at certain days, such as Saturday and Sunday. To solve this, Brand A offers a solution that avoids queuing normally, a fast checkout for items, as well as conveniency for their customers. Investigating the already existing solutions which occupies the space are self-checkout kiosks and Point-of-Sale (POS) systems. The summary of P, CVP and I of Brand A can be viewed in Table 3.

Table 3. P, CVP, and I for brand A

P	CVP	I
Inefficient tellers	Avoid queue	Self-checkout kiosks
Long shopping queue	Fast checkout	POS systems
Simultaneous shopping	Convenient shopping	

Figure 3, 4 and 5 are the tweets collected for brand A's problem space, CVP, and incumbent space respectively.
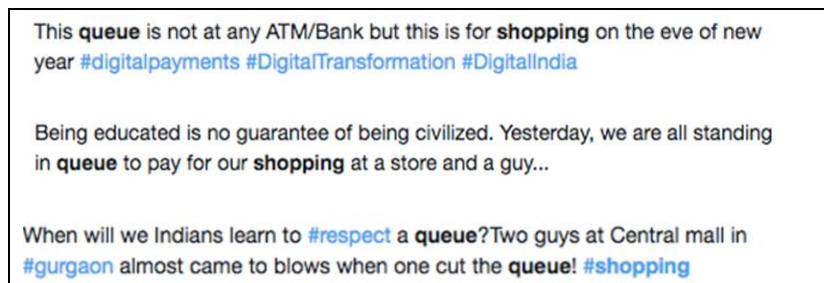

Figure 3. Sample tweets which express P for brand A
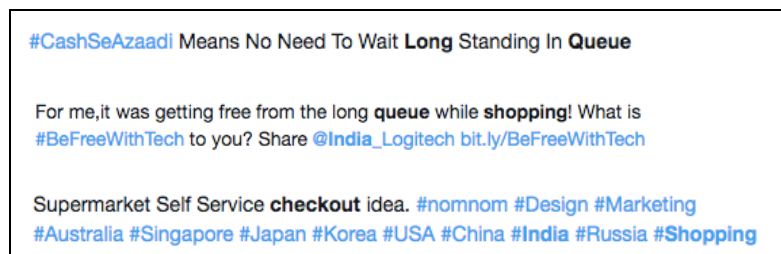

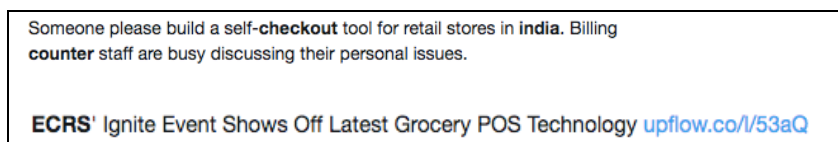Figure 4. Sample tweets which express CVP for brand A

Figure 5. Sample tweets which express I for brand A

From the tweets, we calculate the PM Fit and D-Quotient based on the influencer's Twitter handles up to 2nd degree followers. The computed values are shown in Table 4 and 5. From the values obtained, we found the industry average to be roughly 76 for D-Quotient and 200 for P-M Fit. Values that are below the average for D-Quotient means that the competitor has not sufficiently covered the CVP offered. Similarly, those CVP values for a particular problem space obtained that are above the industry average for P-M Fit is deemed to have fulfilled the needs. These values are indicated as bold.

From Table 4, we observe that 3/3 of Brand A CVP is below the industrial average value, indicating a potential disruptive space. From Table 5, 2/3 of Brand A's CVP scores are above the industrial average

value, indicating a fit between their CVP and the problem space. The steps are repeated for Brand B, C and D. The results are displayed in Table 6.

Table 4. D-quotient for brand A

CVP	Incumbent space	
	Self-check-out kiosks	Existing POS
Checkout time/Checkout counter	**7**	275
Shopping queue	103	**6**
Convenient shopping	**43**	**22**

Table 5. P-M fit for brand A

CVP	Problem space		
	Long checkout time (Tellers are inefficient)	Long shopping queue	Simultaneous shopping (saturday sunday shopping)
Checkout time/Checkout counter	118	200	75
Shopping queue	191	**628**	83
Convenient shopping	171	**216**	170

Table 6. Comparison of disruptometer results and analysts grades

Brand		Disruptometer Results	Analysts/Fellows' Grades
A	P-M Fit	B	B+
	D-Quotient	A	A
B	P-M Fit	B	B
	D-Quotient	B	B
C	P-M Fit	C	C
	D-Quotient	A	A
D	P-M Fit	B	A
	D-Quotient	A	B

From the results obtained in Table 6, we can observe a high correlation between the Disruptometer results and those obtained by professional VC analysts. Brand A, B and C has obtained identical grades, which indicates that the algorithm has potential to be used for market research. Although brand D obtained opposite results, further evaluation with a senior VC determined that he would also grade brand D with results mirroring that obtained by the disruptometer.

Social media can be a attractive source for market research, as highlighted by one study conducted by [13]. This paper which could be considered the greatest influence to our algorithm proposes an approach which computes a similarity function between the brand's Twitter followers and specific attribute exemplars (Eg, environmental friendly). A high similarity index indicates that the brand has that specified attribute perception. The methodology proposed which was highly generalizable and can be used for marketing purposes such as to create perceptual maps, monitoring market structures, and informing research models, was further adopted into the Disruptometer. Their methodology however pulls exemplars which are internationally renowned for the similarity index, therefore requiring manual tuning for local influencers. The Disruptometer has tackled this by taking the tweet locations into consideration, as well as the language and context used. Another weakness that was highlighted in future works section mentioned that the methodology assumes that the brand is using Twitter, which is impended if the brand does not use Twitter to communicate in the first place. The algorithm proposed in this paper is vernacular-based, therefore it can be extended to other social media platforms with social links.

The impact of P-M Fit parameter in this study is highlighted by the in-depth survey that was conducted by [14] which highlighted venture capitalists trends and practices when choosing to invest on a business. Their survey which had 889 respondents representing 681 different VC firms indicated that the product and market is the 3rd and 4th most important factor in consideration whether to invest in a firm or not. The 1st factor, which is team strength was also taken into consideration for the earlier version Disruptometer, however it was discontinued to focus on vernacular dimension of market research.

It should be noted that the proposed algorithm has no direct relation with the 'Disrupt-o-Meter' concept introduced in [15] and further evaluated by [16] and [2]. Although both model aim to forecast the feasibility of a business idea, the approach differs in the methodology. The 'Disrupt-o-meter' model measures the disruptive index by observing whether the business follows the pattern of a disruptive business.

The Disruptometer framework on the other hand forecasts the feasibility of a product based on numbers obtained from social media data mining. Furthermore, a deep learning method [15] could be used to enhance the classification algorithm, as conducted by [11].

4. CONCLUSION

The disruptometer framework is presented in this paper is a marketing insight artificial intelligence algorithm. The algorithm forecasts the viability of a business idea based on its product-to-market fit and disruption quotient, which are determined by the idea's customer value proposition, problem space, and incumbent space. The research has tested the algorithm with 4 brand case studies, 3 which obtained indentical results to that obtained by professional investors.

We acknowledge the following Disruptometer limitations, so that it can be investigated or improved in future research: First, the current algorithm captures the audience for a particular market solely-based on Twitter. In a region where Twitter is not a commonly used social media, obtaining tweets may prove to be challenging. More social platforms can be crawled for data scraping for future iterations, such as Facebook. Second, the algorithm captures the Jaccard index up to the 2nd circle of followers. Most researches such [13] only capture up to first followers, however we found that 2nd circle resulted in higher Jaccard Index correlation. Finally, the current algorithm lacks a concrete step determine the validity of tweets obtained from the user. The current iteration performs manual check of the user history and logical sense, however a future version could incorporate the ability to detect bots/non-human users.

The Disruptometer is planned to be improved in the near future by the combining the concepts of data mining and deep learning. Automation is one of the priority, as most of the Diruptometer task is performed manually and consumed a considerable amount of time.

ACKNOWLEDGEMENTS

REFERENCES
[1]   C. Christensen, *The innovator's dilemma: when new technologies cause great firms to fail*. Harvard Business Review Press, 2013.
[2]   L. O. Gavião, F. T. Ferraz, G. B. A. Lima, and A. P. Sant'Anna, "Assessment of the "Disrupt-O-Meter" model by ordinal multicriteria methods," *RAI Revista de Administração e Inovação,* vol. 13, no. 4, pp. 305-314, 2016.
[3]   P. Chintagunta, D. M. Hanssens, and J. R. Hauser, "Marketing science and big data," ed: INFORMS, 2016.
[4]   P. B. Dastanwala and V. Patel, "A review on social audience identification on twitter using text mining methods," in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2016, pp. 1917-1920.
[5]   M. Qasem, R. Thulasiram, and P. Thulasiram, "Twitter sentiment classification using machine learning techniques for stock markets," in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2015, pp. 834-840.
[6]   S. Das and A. Das, "Fusion with sentiment scores for market research," in *2016 19th International Conference on Information Fusion (FUSION)*, 2016, pp. 1003-1010.
[7]   S. Gong, J. Zhang, P. Zhao, and X. Jiang, "Tweeting as a marketing tool: A field experiment in the TV industry," *Journal of Marketing Research,* vol. 54, no. 6, pp. 833-850, 2017.
[8]   S. Saini and S. Kohli, "Machine learning techniques for effective text analysis of social network E-health data," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 3783-3788.
[9]   M. Usmani, S. H. Adil, K. Raza, and S. S. A. Ali, "Stock market prediction using machine learning techniques," in *2016 3rd International Conference on Computer and Information Sciences (ICCOINS)*, 2016, pp. 322-327.
[10]   V. S. Rajput and S. M. Dubey, "Stock market sentiment analysis based on machine learning," in *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, 2016, pp. 506-510.
[11]   H. N. Bhor, T. Koul, R. Malviya, and K. Mundra, "Digital media marketing using trend analysis on social media," in *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, 2018, pp. 1398-1400.
[12]   Z. Doshi, S. Nadkarni, K. Ajmera, and N. Shah, "TweerAnalyzer: Twitter Trend Detection and Visualization," in *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, 2017, pp. 1-6.
[13]   A. Culotta and J. Cutler, "Mining brand perceptions from twitter social networks," *Marketing science,* vol. 35, no. 3, pp. 343-362, 2016.
[14]   P. Gompers, W. Gornall, S. N. Kaplan, and I. A. Strebulaev, "How do venture capitalists make decisions?," National Bureau of Economic Research2016.

[15] S. D. Anthony, M. W. Johnson, E. J. Altman, and J. V. Sinfield, *The innovator's guide to growth: Putting disruptive innovation to work*. Harvard Business Press, 2008.
[16] F. Hahn, S. Jensen, and S. Tanev, "Disruptive innovation vs disruptive technology: The disruptive potential of the value propositions of 3D printing technology startups," *Technology Innovation Management Review,* vol. 4, no. 12, 2014.

BIOGRAPHIES OF AUTHORS

Mimi Aminah binti Wan Nordin received her B.Eng and Masters at International Islamic University Malaysia and completed her PhD at National University of Malaysia in 2014. She also obtained her MBA at Asia School of Business in 2018. Mimi has started several startups and has experience working on sensor-based solutions. Currently, she is the deputy director for innovation and commercialization unit at research management centre, IIUM, since 2019.

Dmitry Vedenyapin obtained his degree, masters, and PhD in computer science at Volgograd State University. He has also recenly completed his MBA Masters at Asia School of Business. Dmitry started several companies and has experience working on mathematical algorithms for business challenges. Dmitry has spent the past 2 years in Asia, working on live business cases in Asia.

Muhammad Fahreza Alghifari has completed his B.Eng. (Hons) degree in Electronics: Computer Information Engineering from International Islamic University Malaysia (IIUM) in 2018 and is currently pursuing his Masters in Computer Engineering. His research interests are in signal processing, artificial intelligence and affective computing. He received a best FYP award from IEEE Signal Processing – Malaysia chapter and achieved recognition in several national level competitions such as Alliance Bank EcoBiz Competition and IMDC2018.

Teddy Surya Gunawan received his BEng degree in Electrical Engineering with cum laude award from Institut Teknologi Bandung (ITB), Indonesia in 1998. He obtained his M.Eng degree in 2001 from the School of Computer Engineering at Nanyang Technological University, Singapore, and PhD degree in 2007 from the School of Electrical Engineering and Telecommunications, The University of New South Wales, Australia. His research interests are in speech and audio processing, biomedical signal processing and instrumentation, image and video processing, and parallel computing. He is currently an IEEE Senior Member (since 2012), was chairman of IEEE Instrumentation and Measurement Society – Malaysia Section (2013 and 2014), Associate Professor (since 2012), Head of Department (2015-2016) at Department of Electrical and Computer Engineering, and Head of Programme Accreditation and Quality Assurance for Faculty of Engineering (2017-2018), International Islamic University Malaysia. He is Chartered Engineer (IET, UK) and Insinyur Profesional Madya (PII, Indonesia) since 2016, and registered ASEAN engineer since 2018.