

Strengthening Software Make vs. Buy Decision: A Mixed-Method Approach

Farooq Javeed¹, Basit Shahzad², Asadullah Shaikh³, Mana Saleh Al Reshan^{3,*}, Saba Ahmad¹,
Hani Alshahrani³ and Khairan Rajab³

¹Department of Computer Science and Information Technology, University of Sargodha, Sargodha, 40100, Pakistan

²Department of Engineering, National University of Modern Languages, Islamabad, 44000, Pakistan

³College of Computer Science and Information System, Najran University, Najran, 61441, Saudi Arabia

*Corresponding Author: Mana Saleh Al Reshan. Email: manaalreshan@gmail.com

Received: 14 July 2021; Accepted: 18 August 2021

Abstract: Over the past few decades, multiple software development process models, tools, and techniques have been used by practitioners. Despite using these techniques, most software development organizations still fail to meet customer's needs within time and budget. Time overrun is one of the major reasons for project failure. There is a need to come up with a comprehensive solution that would increase the chances of project success. However, the "make vs. buy" decision can be helpful for "in time" software development. Social media have become a popular platform for discussion of all sorts of topics, so software development is no exception. Software developers discuss all the pros and cons of making vs. buy decisions on Twitter and other social media platforms. Twitter trending is a typical feature that evaluates the level of popularity of a specific event on online networking. A mixed-method approach comprising of interviews of software industry experts and Twitter data extraction is applied to scrutinize the effective decision of software build vs. buy decision. The findings of the analysis show that software makes vs. buy decisions depend on several factors including cost, development technology, software development team skills, and time. Based on the finding of the study a framework is proposed for the decision to build versus buy in Small and medium-sized enterprises (SMEs). Furthermore, the framework has been designed to statistically indicate make versus buy decisions of the organization and to suggest appropriate choices based on different parameters.

Keywords: Software make; software buy decisions; twitter classification; decision framework; SMEs

1 Introduction

Software development is one of the leading industries among other commercial industries. Although the software has been developed for the past many decades, however, the development process still needs improvements to predict its faultless development [1]. A big consideration in software development is the decision to "build versus buy". Particularly, commercial off the shelf is available for organizations of all types however all those organizations are facing difficulty in taking the right decision. Previously,



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

researchers have made arguments that support one approach over the other, while some are acknowledged as realities. Because of advancements in the software industry, few of the arguments have become unsubstantial [2]. Software leaders have questioned whether it is better to make software in-house or buy from the market. SMEs (Small and Medium Enterprises) play a very significant role in the development of any country [3] because of having a global impact on the GDP. Smaller-scale enterprises focus on maintaining viability, rise in profits, and increases in sales. A big consideration in development for the organization is the decision to build vs. buy. Build implies that custom subsystems will be built from its core essential parts, then integrates into a final product. Buy implies buying the subsystems and maintaining distance from customizing tasks [2]. Many organizations take a strategic approach to make vs. buy decisions [4]. Organizations that do not consider the build/buy option as an open door are more likely to employ sources of reference where social values based on expert peer acceptance are typically dominated [5]. When a business requires software, it has the option of building it or acquiring it. The decision to acquire more can be made in two ways: whether to buy the entire solution or just a few components. Modifiable off-the-shelf (MoTS) and Commercial off-the-shelf (COTS) solutions are available (CoTS). If the decision is made to construct software, the appropriate environment for software development must be provided. However, deciding whether to build or buy is not a simple decision that takes the careful evaluation of facts, understanding of the organization's strengths and shortcomings, economic situation, and other parameters [2].

In the development of software in SMEs, the build versus buy decision might take any form, as shown in Fig. 1. The pathways of software build vs. buy decision are depicted in Fig. 1. Build vs. buy is a difficult decision to make because it involves many factors such as capacity, organizational resources, sufficient time to be available, and the presence of required expertise in the organization; if all of these factors are present, the decision will be in favor of build; otherwise, an alternative approach will be considered. We gathered the elements from the literature and then updated them with expert opinions gleaned from social media.

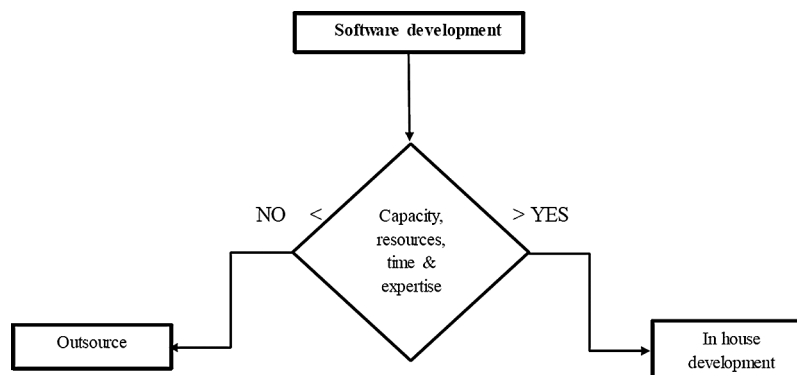


Figure 1: Software build versus buy decision

Social Media are a popular platform to discuss and experience things in a different way. Not only it brings change into someone's opinion but also becomes the cause of social change [6]. Twitter has risen as a broadly used microblogging service. It has a huge and rapidly rising user. They post status (messages) called tweets. It is a short message with a length of 140 characters [7] initially which has upgraded the character length now. Hence, Twitter data is also used as the base to classify Build vs. Buy. Classification is the categorization of data into one or more groups. Data is classified into two categories as labeled and unlabelled data. Unlabelled data is labeled into some classes by using machine learning algorithms. Commonly there are two types: classic binary and modern multiclass [8]. Multiple tools are

used for the classification of data such as WEKA. WEKA stands for “Waikato Environment for Knowledge Analysis” [9]. It has many built-in features like feature generation, feature selection, classification, and evaluation of the selected model. Moreover, the result of the classifier depends on the nature of the data. Therefore, the objective of this research is to identify the factors that help organizations to take an appropriate decision either software should be in-house or buy from vendors.

The organization of the study is as follows. Section 2 elaborates the related work from literature, section 3 discusses the methodology adopted. Section 4 elaborates on the results of the study whereas, the conclusion and future work is presented in section 5.

2 Related Work

Previously, software tasks are considered as sequences of codes. Single person interpreted and solved all problems of obligatory software where no team management is needed [10]. IT can be utilized to recognize and find people with specific expertise nevertheless of a person’s physical location [11]. To write the code people started hiring others as “developers” particularly in modern programming languages. Developers levels are different and hence the code may contain different types of vulnerabilities. The level of developer from given source code is detected by [12]. Moreover, new concepts of some different roles in software development are introduced by the industry i.e., users and developers. Many organizations take a strategic approach to make *vs.* buy decisions. Organizations that did not see the build/buy choice as an open door tend to use points of reference [13]. Commercial-off-the-shelf (COTS) may not fully fit all requirements [14]. According to Xu [15], outsourcing is entirely over an in-house generation if the yield of seller’s production is adequately low or its economies of the extension are remarkably appealing. Cortellessa [16] divided the requirements into two parts; functional and non-functional for components. Some companies take the decision of make, buy and make or buy separately on four different theories that are used to elaborate the decision of make or buy for companies (transaction cost economics composite, neoclassical economics/rational choice decision, the resource-based view of strategy and institutional theory explanations) [17]. According to Daneshgar et al. [18], discoveries make the required decision more time-consuming. Simply they acknowledged that if dealer support is insufficient, the decision to buy is mostly among SMEs. Cost is always a big consideration; its risk remains the same in both cases [19,20]. Availability of multiple solutions to requirement fit leads to buying in many cases. Instead of the typical software development approach build *vs.* buy decision needs the contribution of people (networks, data centers, etc.) that may gather data from different techniques by Torrecilla-Salinas [20] to explore that the agile techniques enrich the build *vs.* buy decision. The organization mostly goes for a third option except for build *vs.* buy but to cope with the risk factor [21]. In the process of literature review, multiple factors have been identified that affect the organization’s development method. Factors that affect the build *vs.* buy decision of SEMs are summarized in [Tab. 1](#). Previous research [2,18] has backed up these factors, which have been used to create a generic set of criteria.

Machine learning (ML) is a branch of artificial intelligence. It works on the two approaches supervised and unsupervised learning. In supervised learning provided output is used for the training of the input data. For the algorithm, a training labeled data set has been used under supervised learning. Different supervised techniques are support vector machine (SVM), naïve Bayes (NB), Logistic regression, Decision tree, and Maximum Entropy. Clustering is an absolute unsupervised learning methodology. Some unsupervised algorithms are HMM, K-Mean, and Neural Network. Classification of data is necessary to acquire desirable results. Commonly there are two types, classic binary, and modern multi-class. Social media, nowadays, has numerous impacts on society. Many social sites are famous including Facebook, Instagram, Viber, WhatsApp, Twitter, Linked In, and many more. Twitter as a social networking site, started on 21 March 2006. Through tweets, people can express their views and emotions precisely due to

limited character space, which is 140 in numbers now its 280-character space [22]. With some limit's tweets can be searched by API, a search facility provided by Twitter [23]. Per day production of tweets is half of a billion by its users [24]. It is a faster way to gain knowledge. Tweet data is collected and data is divided into two classes; yes, and no.

Table 1: Factors that affect the build vs. buy decision of SEMs

F#	Factor	Reference
1	The factor of strategy and competitive advantage	[18]
2	Scale and complexity factor	[18], [25],[26–29]
3	Cost factor	[18], [21],[29],[30]
4	The factor of requirement fit	[18],[30]
5	Time factor	[18], [26],[27], [28],[29]
6	In-house expertise factor	[17],[18]
7	The factor of risk involved	[17],[18], [25], [30], [31]
8	The factor of the Support structure	[17],[18],
9	Operational factor	[18], [30]
10	The factor of commoditization, flexibility, and change.	[18], [21], [29], [31]
11	Economic factor	[21], [25], [28], [29]
12	The factor of Standards involved	[26], [27], [29],[32]

The literature gives rise to different questions like how to build or buy decisions should be taken for different organizations? What are the factors that affect the organization's development method? and what are the reasons behind the build vs. buy decision? Therefore, the objective of this research is to identify the factors that help organizations to take an appropriate decision either software should be in-house built or buy from vendors. Another objective is to suggest an appropriate decision to build vs. buy in SMEs with the help of a framework. Machine learning classification is used in this study to classify data into two categories i.e., build & buy.

In the literature study, the most significant factors studied by many researchers are “Time” and “Risk,” followed by economics, cost, complexity, and changeability. The variables “strategy and competitive advantage,” “support structure,” and “operational factor” are the ones that have received the least attention.

3 Methodology

To make the study more complete, a mixed-method approach based on qualitative and quantitative methodologies is used. This research tries to determine the best decision for businesses to build vs. buy based on the characteristics listed in Tab. 1 and gleaned through a literature evaluation.

The methodology adopted in this study is based on four steps.

- i. First, factors influencing the decision to make vs. buy are found in the literature, with a total of twelve factors indicated in Tab. 1.
- ii. Second, a survey questionnaire is constructed to collect data from IT specialists, owners, developers, and senior management from various small to medium enterprises based on these factors. All of the information acquired in the orderly survey is done so by asking the appropriate questions and following a well-defined process [32]. To organize the information, different prioritization strategies are used. Pilot research is conducted by sending questioners to ten industry experts, after which the questionnaire is changed and given to other organizations

based on the input received from the experts. The relevant information is extracted from the dataset using the SPSS program [33].

- iii. Then, to cross-validate, the survey findings, interviews with IT specialists from five SMEs are undertaken.
- iv. Finally, using Twitter, compare the results of the survey and interview with overall software-industry practices in different countries. Machine learning techniques are used to gather and classify Twitter Tweets.

Data collection is the key part of these types of studies. Numerous ways are used for data pool such as interviews, surveys, meetings, and an online survey. The collection of data is a crucial requirement to establish the cogency of this approach. For the survey, data is collected by digital means. It is conceivable to compute the reliability and the capability of summing up results to a bigger population. Furthermore, the profundity of a quantitative report is restricted [28]. The online questionnaire is sent to 65 organizations in Pakistan among which 58 had responded. 10 responses are not filled while 2 are not filled by relevant persons. Thus, valid data of 46 organizations are utilized in this study. The survey consists of a five-point Likert scale i.e., from 0-strongly disagree, to 4-strongly agree is applied to get responses from experts.

For the experts' interview phase, face to face interview session of 30 min for each interviewee is conducted. Interviewees are asked five open-ended questions regarding build vs. buy decisions in their organizations. The findings of the interviews are summarized and analyzed for common themes.

For this study, 569 Tweeter tweets are collated are classified. All tweets of IT experts (known) from 2009 to the present are collected during the two months. Tweets collection is an essential and basic phase of this approach therefore by using API search, tweets are collected from Twitter. The pre-processing technique for withdrawal of inappropriate characters and elimination of other frequent words is applied. All data is pre-processed before giving it to classifiers. Furthermore, Tweets Data, Outlier Removal, Keyword Stemming, Spelling Correction, and Category Ranking steps have been performed. After the execution of a classifier, the next stage is the validation of the model by a cross-validation method. There are multiple techniques used for this purpose. Precision, accuracy level, recall, and F-measure are frequently utilized [22]. Recall and precision are inversely proportional to each other; when recall increases precision decreases and when recall decreases precision increases [23].

4 Data Analysis and Results

The outcomes of the survey's data analytics with SPSS, interview sessions, and Tweeter data analysis using machine learning algorithms are discussed in this section.

4.1 Survey Results

The data used in this study is largely dependent on the response from the software houses. The data analysis is performed on survey data with the help of the SPSS tool. The demographic analysis is shown in Fig. 2a indicates that 10.9% of the organization's owners gave responses, while another 10.9% of responses are given by senior management and 15.2% of responses were given by middle-level management. 45.7% of responses came from IT experts and 8.7% of responses were given by entry-level employees or others. Similarly, Fig. 2b indicates that 69% of the interview respondents are IT experts, while 18% are executives and 13% of the respondents belong to senior management.

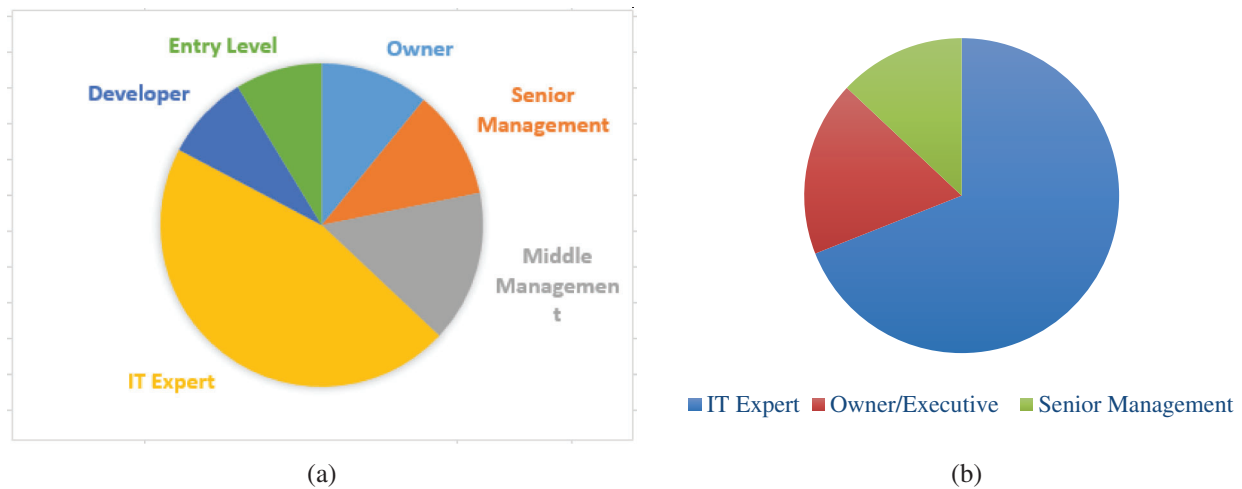


Figure 2: Demographic information of respondents (a) Responses of different organizational representatives (b) Expert types for the interview

According to the survey data correlation analysis in SPSS, there is an inverse relationship between organization expertise and the choice to buy software; this indicates that the greater the expertise, the lower the possibilities of buying, and the lower the expertise, the higher the possibilities of buying, as shown in [Tab. 2](#). Pearson correlation tells about the direction and strength of the relationship between organization expertise and the decision of software buying.

Table 2: Correlations of organization’s expertise

Correlation Factors		Organization Expertise	What does your organization do when the project is out of organizational expertise
Organization Expertise	Pearson	1	-.682**
	Correlation	46	000
	Sig. (2-tailed) N		46
What does your organization do when the project is out of organizational expertise	Pearson	-.682**	1
	Correlation	000	46
	Sig. (2-tailed) N	46	

The findings confirmed that there is a high association between organization expertise and the decision of software buying. The values of Pearson $-.682^{**}$ confirmed the inverse proportion between the variables. The value $-.682$ is near 1 which means that there is a high association between the variables. Conclusively, one variable is negatively associated with the other variable. The findings of the study are consistent with [\[15,16,34–36\]](#). Also, there is a direct proportion between organization capacity and the decision of software build means that higher will be the capacity of the organization, higher will be chances of build and lower will be the capacity of organization lower will be the chances of the build as shown in [Tab. 3](#).

Correlation is significant at the 0.01 level. The findings confirmed that there is a high association between organization capacity and the decision of software building. The values of Pearson $.716^{**}$ confirmed the direct proportion between the variables. The value $.716$ is near 1 which means that there is a high association between the variables. It also means that one variable is positively associated with the other variable. So our hypothesis “higher will be the capacity higher will be the chances of build in

house” is proved. The finding of this hypothesis is in line with [15,16,37,38]. Similarly, there is a direct proportion between Situational factors (Time, Cost, and Budget) and software building decisions, shown in Tab. 4.

Table 3: Correlations of organization’s capacity

Correlation Factors		Organization Capacity	What does your organization do when the project is out of organizational capacity
Organization Capacity	Pearson	1	.716**
	Correlation	46	000
	Sig. (2-tailed) N		46
What does your organization do when the project is out of organizational capacity	Pearson	.716**	1
	Correlation	000	46
	Sig. (2-tailed) N	46	

Table 4: Correlations of organization’s situational factor

Correlation Factors		Situational Factors (Time, Cost, and Budget)	Considering the risk factors, you will go for in house build
Situational Factors (Time, Cost, and Budget)	Pearson	1	.638**
	Correlation	46	000
	Sig. (2-tailed) N		46
Considering the risk factors, you will go for in house build	Pearson	.638**	1
	Correlation	000	46
	Sig. (2-tailed) N	46	

As stated, correlation is significant at the 0.01 level. The findings confirmed that there is highly association between situational factors (time, cost, and budget) and the decision of software building. The values of Pearson .638** confirmed the direct proportion between the variables. The value .638 is near 1 which means that there is a high association between the variables. It also means that one variable is positively associated with the other variable. Hence, our hypothesis “There is a direct proportion between Situational factors (time, cost and budget) and software building decision” is correct. The results of the hypotheses are consistent with previous studies [17], [18,39,40].

In Fig. 3, the learning curve shows that when a new project comes in, the organization checks its expertise. If the required expertise is higher in the organization, then more chances of build and if the expertise is low fewer chances of build and go for the buy. If the situational factors (time, cost, etc.) are higher than the decision made by the organization is build and vice versa. Also, the budget and capacity of the organization are checked, and if it is out of budget go for buy else build.

Factors	A	B	C	D
Expertise	X		X	
		X		X
Time	X		X	
Cost	X		X	
Budget	X		X	
		X		X
Capacity	X		X	
		X		X

Where A= High, B= Low, C= Build, D=Buy

Figure 3: The learning curve for the build vs. buy decision

4.2 Interview Findings

According to the first interviewee, organizational scale, capacity, and resources affect the build vs. buy decision. And it's better to build in-house if expertise is present. If the technology changes help can be taken from the vendors, then make sure to build in-house. Similarly, the second respondent agreed that the build versus buy decision depends on the time and cost. And it's better to build in-house if there is no issue of budget to complete the project on time. New technology can be adapted to compete in the market and get help from vendors. In contrast to previous comments, the third responded view is that the organizational requirements, risk factors, capacity, and resources affect the build vs. buy decision. If solutions are available, then go for them. Build something that the market wants. The point of view of the 4th respondent is organizational requirements, risk factors, resources, and capacity affect the build vs. buy decision. If solutions are available to go for it. If time is limited, will buy, and on the other hand if have a better solution outsource take it. Finally, the last respondent argues that build vs. buy depends on multiple factors and the size of the project. Try to build a system on your own if the technology change hires a new skilled person and builds in-house. However, if there is no possible solution then go for the buy option.

4.3 Result of Machine Learning Classifiers

Machine learning classifiers are used for the training, testing, and classification of data. Tweets of software industry experts are collected and are used for further verification of evaluated results of previous sections. Results are in line with the previous study [39]. Tab. 5 presents the results of all classifiers including logistic regression, naïve Bayes, simple logistic, Bayes net, voted perceptron, and zeroR on the scale of precision, recall, f-measure, and accuracy.

Table 5: Results of all classifiers

Classifier	Precision	Recall	F-measure	Accuracy (%)
Logistic regression	0.885	0.885	0.892	91.03
Naive bayes	0.877	0.877	0.859	84.3
Simple logistic	0.851	0.851	0.874	90.6
Bayes net	0.838	0.838	0.875	91.5
Voted preception	0.875	0.875	0.881	91.3
ZeroR	0.838	0.838	0.875	91.5

Using the Logistic Regression classification technique, we detected 518 correctly identified tweets and 51 incorrectly categorized tweets, resulting in a 91 percent accuracy rate. The weighted average for Precision is 0.885, Recall 0.910, and F-Measure is 0.892 as shown in Fig. 4.


```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      518          91.0369 %
Incorrectly Classified Instances    51           8.9631 %
Kappa statistic                    0.2006
Mean absolute error                 0.0905
Root mean squared error            0.298
Relative absolute error            58.096 %
Root relative squared error        107.231 %
Total Number of Instances         569

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.979   0.833   0.927     0.979   0.952     0.225   0.738    0.957    yes
                0.167   0.021   0.421     0.167   0.239     0.225   0.738    0.315    no
Weighted Avg.   0.910   0.765   0.885     0.910   0.892     0.225   0.738    0.903

=== Confusion Matrix ===

  a  b  <-- classified as
510 11 |  a = yes
 40  8 |  b = no

```

Figure 4: Logistic regression results on build vs. buy

By applying the Naïve Bayes classification algorithm, 480 correctly classified tweets are found, incorrectly classified tweets are 89, and the accuracy rate is 84 percent. The weighted average for Precision is 0.877, Recall 0.844, and F-Measure is 0.859. Similarly, by applying the Simple Logistic classification algorithm, we have founded 516 correctly classified, incorrectly classified tweets are 53, and the accuracy rate is 90 percent. The weighted average for Precision is 0.851, Recall 0.907, and F-Measure is 0.874. By applying Bayes Net classification algorithm, we have founded 521 correctly classified tweets, incorrectly classified tweets are 48, and the accuracy rate is 91 percent.

The weighted average for Precision is 0.838, Recall 0.916, and F-Measure is 0.875. It is also analyzed that by applying the Voted Perceptron classification algorithm, we have founded 520 correctly classified tweets, incorrectly classified tweets are 49, and the accuracy rate is 91 percent. The weighted average for Precision is 0.875, Recall 0.914, and F-Measure is 0.881. Furthermore, by applying the ZeroR classification algorithm, we have founded 521 correctly classified tweets, incorrectly classified tweets are 48, and the accuracy rate is 91 percent. The weighted average for Precision is 0.838, Recall 0.916, and F-Measure is 0.875. The classification characteristics values of various methods are represented as a chart in [Fig. 5](#).

4.4 Proposed Framework

Based on the findings of the survey analysis and machine learning classification a framework for the build vs. buys decision is prosed as shown in [Fig. 6](#) Expertise is drawn from a variety of fields in the proposed framework, including programming, database management, networking, and Android. A specific amount of time is allotted to accomplish a project or meet deadlines; finance includes resources (skilled labor costs, equipment costs, etc.); and capacity includes infrastructure, personnel, and machinery.

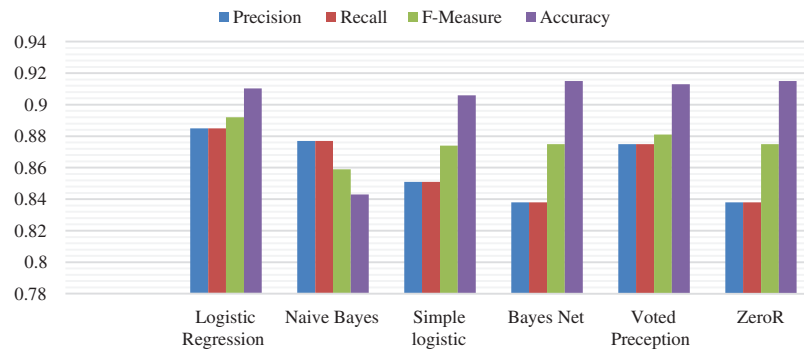


Figure 5: Results of Classifiers

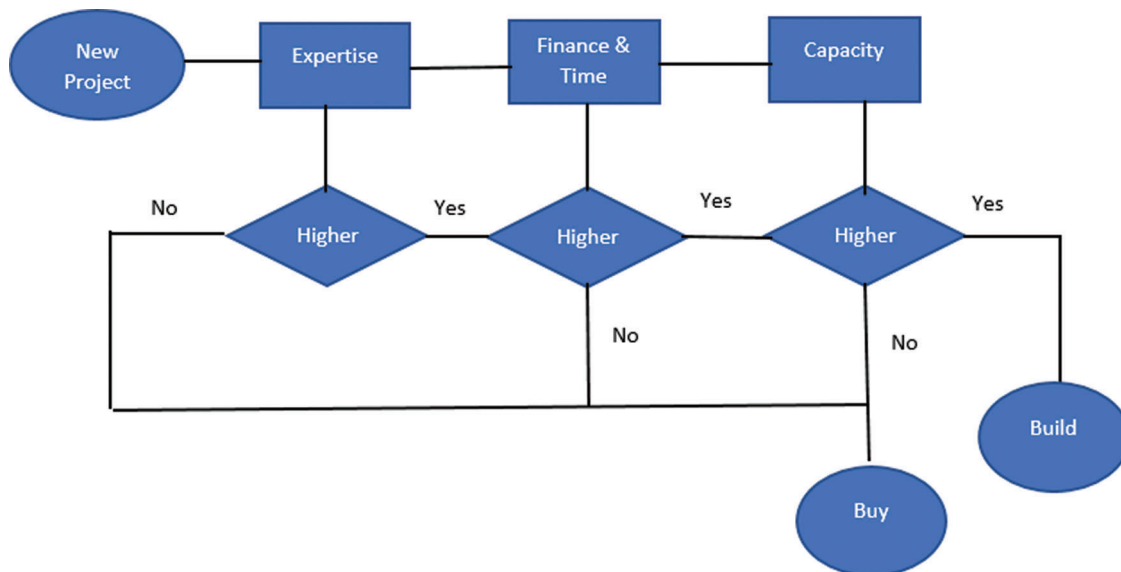


Figure 6: A proposed framework for the build vs. buy decision

The greater the expertise within an organization, the more the chances of developing, and the lower the knowledge within an organization, the higher the possibilities of buying. More situational elements (time, cost budget) increase the likelihood of building, while fewer elements increase the likelihood of buying. If an organization's resources and capacity are limited, the chances of a build are low; conversely, if the organization's capacity is high, the probabilities of a build are higher. When a new project comes in, the framework shows that the organization checks to see if it has enough competence. The project is denied if the requisite expertise is not present inside the organization. Also, the organization's budget and capacity are examined, and if all of the categories are greater, the organization should build rather than buy.

5 Conclusions and Future Work

For a long time, commercial of the shelf is available for organizations of all types, however, the problem is the right choice of 'build versus buy'. All those organizations are facing difficulty in taking the right decision due to which a deep understanding of the system is required to find out core factors on which the decision to build or buy should be taken. In this research, we tried to help organizations in taking accurate decisions for software development. We used both qualitative and quantitative approaches to get

the most appropriate results. Firstly, core factors are identified on which the decision depends. It shows that the capacity and situational factors (time, cost, and budget) are directly proportional, higher these factors best to build in-house and vice versa. Furthermore, organizational expertise is indirectly proportional to buy decisions, higher than the expertise fewer chances of buy and lower the expertise more chances of buy. Secondly, the context-based analysis of Twitter data is performed. Different machine learning algorithms have been used to get the finest results from Twitter data. Data are gathered about the Make vs. Buy decision. Dataset is passed through two stages (training and testing) to attain the best result from tweet data. A close examination of different machine learning classifiers has been done. Bayes Net and ZeroR have shown the best results inaccuracy.

This study is conducted on a limited number of SME organizations and the proposed framework is based on identified factors. The researchers can apply the identified elements in their research. In the future, the approach could aid major organizations in deciding whether to build or buy. The biasness can be handled by using data collected from various social media platforms. Furthermore, the tweets covering images can also be taken under observation.

Funding Statement: The authors would like to acknowledge the support of the Deputy for Research and Innovation, Ministry of Education, Kingdom of Saudi Arabia for this research through a grant (NU/IFC/INT/01/008) under the institutional Funding Committee at Najran University, Kingdom of Saudi Arabia.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] J. C. Lee, Y. C. Shiue and C. Y. Chen, "Examining the impacts of organizational culture and top management support of knowledge sharing on the success of software process improvement," *Computers in Human Behavior*, vol. 54, pp. 462–474, 2016.
- [2] B. Shahzad, A. M. Abdullatif, N. Ikram and A. Mashkoor, "Build software or buy a study on developing large scale software," *IEEE Access*, vol. 5, pp. 24262–24274, 2017.
- [3] I. Ahmed, A. Shahzad, M. Umar and B. A. Khilji, "Information technology and SMEs in Pakistan," *International Business Research*, vol. 3, no. 4, pp. 237–240, 2010.
- [4] I. Shin, "Adoption of enterprise application software and firm performance," *Small Business Economics*, vol. 26, no. 3, pp. 241–256, 2006.
- [5] B. S. Buchowicz, "A process model of make-vs.-buy decision-making," *The Case of Manufacturing Software IEEE Transactions on Engineering Management*, vol. 38, no.1, pp. 24–32, 1991.
- [6] T. Dong, C. Liang and X. He, "Social media and internet public events," *Telematics and Informatics*, vol. 34, no. 3, pp. 726–739, 2016. [Online] Available: <http://dx.doi.org/10.1016/j.tele.2016.05.024>.
- [7] S. E. Shukri, R. I. Yaghi, I. Aljarah and H. Alsawalqah, "Twitter sentiment analysis: a case study in the automotive industry," in *Applied Electrical Engineering and Computing Technologies (AEECT), IEEE Jordan Conference*, Amman, Jordan, pp. 1–5, 2015.
- [8] S. H. Peled, D. Roth and D. Zimak, "Constraint classification for multiclass classification and ranking," in *NIPS'02: Proceedings of the 15th International Conference on Neural Information Processing Systems*, MIT Press, 55 Hayward St., Cambridge, MA, United States, pp. 809–816, 2002.
- [9] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann *et al.*, "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
- [10] E. W. Dijkstra, "Complexity controlled by hierarchical ordering of function and variability," in *Software Engineering: Report on a Conference Sponsored by the NATO Science Committee*, Garimisch, Germany, pp. 181–185, 1968.
- [11] N. Mehta, D. Hall and T. Byrd, "Information technology and knowledge in software development teams: The role of project uncertainty," *Information and Management*, vol. 51, no. 4, pp. 417–429, 2014.

- [12] F. Javeed, A. Siddique, A. Munir, B. Shehzad and M. I. Lali, "Discovering software developer's coding expertise through deep learning," *IET Software*, vol. 14, no. 2, pp. 213–220, 2020.
- [13] H. Dar, M. I. Lali, H. Ashraf, M. Ramzan, T. Amjad *et al.*, "A systematic study on software requirements elicitation techniques and its challenges in mobile application development," *IEEE Access*, vol. 6, pp. 63859–63867, 2018.
- [14] P. C. Jha, V. Bali, S. Narula and M. Kalra, "Optimal component selection based on cohesion & coupling for component based software system under build-or-buy scheme," *Journal of Computational Science*, vol. 5, no. 2, pp. 233–242, 2014.
- [15] W. Xiuting and D. Haibo, "Analysis of buying decision making process on Chinese clothing for Ethiopian young consumers," in *2011 Int. Conf. on Product Innovation Management (ICPIM 2011)*, Hubei, China, pp. 754–757, IEEE, 2011.
- [16] V. Cortellessa, F. Marinelli and P. Potena, "An optimization framework for build-or-buy decisions in software architecture," *Computers & Operations Research*, vol. 35, no. 10, pp. 3090–3106, 2008.
- [17] H. J. Rosenberg, N. P. Mols and A. R. Villadsen, "Make and buy—an alternative to make or buy? An investigation of four theoretical explanations in danish municipalities," *International Journal of Public Administration*, vol. 34, no. 8, pp. 539–552, 2011.
- [18] F. Daneshgar, G. C. Low and L. Worasinchai, "An investigation of 'build vs. buy' decision for software acquisition by small to medium enterprises," *Information and Software Technology*, vol. 55, no. 10, pp. 1741–1750, 2013.
- [19] S. K. Bhatti, M. I. Lali, B. Shahzad, F. Javid, F. U. Mangla *et al.*, "Everaging the big data produced by the network to take intelligent decisions on flow management," *IEEE Access*, vol. 6, pp. 12197–12205, 2018.
- [20] C. J. Torrecilla-Salinas, J. Sedeño, M. J. Escalona and M. Mejías, "Estimating, planning and managing agile Web development projects under a value-based perspective," *Information and Software Technology*, vol. 61, pp. 124–144, 2015.
- [21] B. Shahzad, K. M. Awan, M. I. Lali and W. A. Aslam, "Identification of patterns in failure of software projects," *Journal of Information Science Engineering*, vol. 33, no. 6, pp. 1465–1479, 2017.
- [22] M. S. Tiwana, F. Javeed, M. I. Lali, H. Dar and M. Bilal, "Comparative analysis of context based classification of twitter," in *2018 Fourth Int. Conf. on Advances in Computing, Communication & Automation (ICACCA)*, Subang Jaya, Malaysia, pp. 1–5, IEEE, 2018.
- [23] A. M. Abdullatif, B. Shahzad and A. Hussain, "Evolution of social media in scientific research: A case of technology and healthcare professionals in Saudi universities," *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 6, pp. 1461–1468, 2017.
- [24] B. Shahzad, K. M. Awan, A. M. Abdullatif, M. I. Lali, M. S. Nawaz *et al.*, "Quantification of productivity of the brands on social media with respect to their responsiveness," *IEEE Access*, vol. 7, pp. 9531–9539, 2019.
- [25] S. T. Bakhsh, B. Shahzad and S. Tahir, "Risk management approaches for large scale software development," *Journal of Information Science and Engineering*, vol. 33, no. 6, pp. 1547–1560, 2017.
- [26] K. Saleem, B. Shahzad, M. A. Orgun, J. Al-Muhtadi, J. J. Rodrigues *et al.*, "Design and deployment challenges in immersive and wearable technologies," *Behaviour and Information Technology*, vol. 36, no. 7, pp. 687–698, 2017.
- [27] V. Cortellessa, F. Marinelli and P. Potena, "Automated selection of software components based on cost/reliability tradeoff," in *European Workshop on Software Architecture, Nantes, France*, pp. 66–81, 2006.
- [28] M. Choshin and A. Ghaffari, "An investigation of the impact of effective factors on the success of e-commerce in small-and medium-sized companies," *Computers in Human Behavior*, vol. 66, pp. 67–74, 2017.
- [29] J. M. Müller, "Business model innovation in small-and medium-sized enterprises: Strategies for industry 4.0 providers and users," *Journal of Manufacturing Technology Management*, vol. 30, no. 8, pp. 1127–1142, 2019. Available: 10.1108/jmtm-01-2018-0008.
- [30] D. A. Pacheco, C. S. Caten, C. Jung, C. F. Sassanelli and S. Terzi, "Overcoming barriers towards sustainable product-service systems in small and medium-sized enterprises: State of the art and a novel decision matrix," *Journal of Cleaner Production*, vol. 222, pp. 903–921, 2019.

- [31] K. R. Fowler and S. A. Dyer, "A model for the build-versus-buy decision in developing embedded systems," *IEEE Systems Journal*, vol. 14, no. 1, pp. 1317–1324, 2020.
- [32] O. Gomez, P. Wriedt and F. Zhao, "Build or Buy: A case study for ERP system selection in SMEs," in *Int. Conf. on Human-Computer Interaction*, Toronto, Canada, pp. 23–33, 2016.
- [33] SPSS Software, "International Business Machines Corporation," (Accessed 6 May 2021), 2021. Available: <https://www.ibm.com/analytics/spss-statistics-software>.
- [34] P. Verschuren, "Case study as a research strategy: Some ambiguities and opportunities," *International Journal of Social Research Methodology*, vol. 6, no. 2, pp. 121–139, 2003.
- [35] E. Lloret and M. Palomar, "Text summarisation in progress: A literature review," *Artificial Intelligence Review*, vol. 37, no. 1, pp. 1–41, 2012.
- [36] Q. N. Naveed, M. R. N. M. Qureshi, A. Shaikh, A. O. Alsayed, S. Sanober *et al.*, "Evaluating and ranking cloud-based e-learning critical success factors (CSFs) using combinatorial approach," *IEEE Access*, vol. 7, pp. 157145–157157, 2019.
- [37] M. A. Memon, A. Shaikh, A. Sulaiman, A. Alghamdi, M. Alrizq *et al.*, "Time and quantity based hybrid consolidation algorithms for reduced cost products delivery," *Computers, Materials and Continua*, vol. 69, no. 1, pp. 409–432, 2021.
- [38] N. A. Mahoto, R. Iftikhar, A. Shaikh, Y. Asiri and A. Alghamdi, "An intelligent business model for product price prediction using machine learning approach," *Intelligent Automation & Soft Computing*, vol. 30, no.1, pp. 147–159, 2021.
- [39] A. Sahaym, A. A. Datta and S. Brooks, "Crowdfunding success through social media: Going beyond entrepreneurial orientation in the context of small and medium-sized enterprises," *Journal of Business Research*, vol. 125, pp. 483–494, 2019.
- [40] P. Maroufkhani, M. L. Tseng, M. Iranmanesh, W. K. Ismail and H. Khalid, "Big data analytics adoption: Determinants and performances among small to medium-sized enterprises," *International Journal of Information Management*, vol. 54, no. 102190, pp. 1–15, 2020.