Int. J. Production Economics 226 (2020) 107621

Contents lists available at ScienceDirect

International Journal of Production Economics

journal homepage: http://www.elsevier.com/locate/ijpe

The impact of entrepreneurship orientation on project performance: A machine learning approach

Sima Sabahi [a], Mahour Mellat Parast [b,*]

[a] North Carolina A&T State University, 1601 E Market Street Greensboro, NC, 27411, USA
[b] Arizona State University, 660 S College Ave Tempe, AZ, 85281, USA

ARTICLE INFO

Keywords:
Project performance
Entrepreneurship orientation
Machine learning
Supervised learning
Predictive analytics

ABSTRACT

Recent studies in project management have shown the important role of entrepreneurship orientation of the individuals in project performance. Although identifying the role of entrepreneurship orientation as a critical success factor in project performance has been considered as an important issue, it is also important to develop a measurement system for predicting performance based on the degree of an individual's entrepreneurial orientation. In this study, we use predictive analytics by proposing a machine learning approach to predict individuals' project performance based on measures of several aspects of entrepreneurial orientation and entrepreneurial attitude of the individuals. We investigated this relationship using a sample of 185 observations and a range of machine learning algorithms including lasso, ridge, support vector machines, neural networks, and random forest. Our results showed that the best method for predicting project performance is lasso. After identifying the best predictive model, we then used the Bayesian Information Criterion and the Akaike Information Criterion to identify the most significant factors. Our results identify all three aspects of entrepreneurial attitude (social self-efficacy, appearance self-efficacy, and comparativeness) and one aspect of entrepreneurial orientation (proactiveness) as the most important factors. This study contributes to the relationship between entrepreneurship skills and project performance and provides insights into the application of emerging tools in data science and machine learning in operations management and project management research.

1. Introduction

Projects are unique, short-term activities that lead to new products, services, or outcomes (APM, 2012; PMI, 2017). Successful project performance is regarded as some or all of project management success, project success, and product success, which are aligned to the individual stakeholders' performance criteria (Collins and Baccarini, 2004; Barclay, 2008). Since managers and other project stakeholders have a vested interest in evaluating the value that projects deliver to organizations, there is a requirement for proper evaluation and support tools to effectively monitor and control projects (Brynjolfsson, 1993; Melville et al., 2004).

According to Barclay and Osei-Bryson (2010), one of the major problems in evaluating project performance is the traditional system of measures that adheres to cost, time, and specifications. These measures have been considered incomplete by many researchers (e.g., Wateridge (1998); Atkinson (1999); Cohen and Graham (2001); Atkinson et al. (2006)). This incompleteness has been identified as a basis for a high incidence of project failure (Glass, 2005). Despite these claims, these measures still dominate practice (Agarwal and Rathod, 2006). Thus, there is a need to provide more robust measures of project performance.

Researchers have offered various solutions to performance evaluation, especially in developing critical success factors and measures to consider in assessing project performance (Barclay and Osei-Bryson, 2010). One of the many factors that contribute to project success is the entrepreneurial orientation of individuals. Because of the importance of entrepreneurship in economic growth, productivity, and job creation, many practitioners and researchers have been interested in this concept (Ge and Peng, 2012; Stamboulis and Barlas, 2014; Ambad and Damit, 2016; Barba-Sánchez and Atienza-Sahuquillo, 2018). Some theoretical views have been introduced to analyze the relationship between entrepreneurship skillsets and performance outcomes (Souitaris et al., 2007). For example, in the context of project management, Martes et al. (2015) have shown there is a positive relationship between project management and entrepreneurship orientation (EO) through integration, scope, time, cost, quality, human resources, communications, risk,

* Corresponding author.
E-mail addresses: Ssabahi@aggies.ncat.edu (S. Sabahi), mahour.parast@asu.edu (M.M. Parast).

https://doi.org/10.1016/j.ijpe.2020.107621
Received 16 April 2019; Received in revised form 16 December 2019; Accepted 13 January 2020
Available online 15 January 2020
0925-5273/

and procurement management. Also, Martens et al. (2018) developed a model showing the positive impact of EO on project success in the Brazilian context.

Despite the importance of EO on project performance, most of the studies have focused on the relationship between EO and firm performance in the entrepreneurship literature (Rauch et al., 2009; Filser and Eggers, 2014), with performance and entrepreneurial attitude being two of the subjects most investigated in EO-related studies for about 30 years (Martens et al., 2016). A central contention in this relationship is that EO can be seen as an entrepreneurial strategy, making processes that key leaders use to establish their organizational goals, support their vision, and create competitive advantage (Rauch et al., 2009). Although the relationship between EO and firm performance has been well studied in the literature (Moreno and Casillas, 2008; Rauch et al., 2009), the relationship between project performance and EO is not clear.

Apart from understanding the success factors contributing to project performance, it is also important to be able to predict the outcome of a project from different perspectives. Although the project management literature has introduced different methods for project performance evaluation, shortcomings associated with these traditional methods suggest a need for improvement in project performance prediction in a project management context (Cheng et al., 2010). Over the past few decades, there have been notable advances in predictive modeling techniques and concepts from machine learning, statistics, and computer science that have value for organizational researchers and practitioners (Putka et al., 2018). Although there are several studies that have used modern predictive techniques in the context of project management (e.g., Chulani et al., 1999; Dvir et al., 2006; Paliwal and Kumar, 2009; Wang and Gibson, 2010), there is still a need to develop more understanding regarding the potential use of these methods in project management. A new approach uses a perspective that is primarily concerned with the prediction of project performance using the methods of predictive analytics. (Shmueli and Koppius, 2011; Kleinberg et al., 2015).

The contribution of this study is threefold: First, this paper evaluates the impact of several aspects of entrepreneurship orientation (EO) on project performance (PP) using predictive analytics grounded in a machine learning approach. In contrast to traditional hypothesis testing and theory building research that aim to examine the *causality* between the input (EO) and the output (PP), our goal is to develop a predictive analytics model that can predict the effect of EO on PP and identify aspects of EO that are important factors in the prediction (Shmueli and Koppius, 2011; Kleinberg et al., 2015). We predict project performance based on the degree of EO, using four sets of machine learning algorithms: multiple linear regression, support vector machines, neural networks, and random forests. We chose these four sets of algorithms due to their popularity in the machine learning literature (Oztekin et al., 2013) as well as the evidence we found on their applicability in the entrepreneurship literature (Fellnhofer et al., 2016; Iskender and Batı, 2015; Li, 2018; Xu et al., 2018). We then select the machine learning algorithm that is best at predicting project performance, and use the results to identify the aspects of entrepreneurial orientation and attitude that are the most important factors in the prediction.

Second, in a similar vein to Gemünden et al. (2018) and Martens et al. (2018), we argue that EO's impact on project performance is not limited to entrepreneurship projects; the impact of EO can be examined for any type of project. Third, the technique variation and the applicability of our approach makes our models easily replicated by scholars and practitioners. Thus, our models can be implemented in real-world situations, such as the case of a project manager who can better understand and analyze project performance, and the case of entrepreneurship institutes that can improve their functionality based on the results of machine learning models.

This paper is organized as follows. In the next section, we first provide the literature on determinants of project performance. After identifying the importance of behavioral and entrepreneurial aspects of project management, we will discuss EO and its different dimensions. Then we provide the literature on project performance evaluation methods. After identifying the benefits of machine learning in the field, we provide a brief review of machine learning and the algorithms that we will use in this study. In the third section, we outline our methodology, our evaluation method, and our approach for comparing the algorithms and identifying the aspects of EO and attitude that are important factors in the prediction. In the fourth section, we explain our data collection method, present our empirical experimentation results, analyze the performance of each of the predictive models, and identify the most important predictor variables. In the fifth section, we discuss our findings, the implications for research and practice, and future directions.

2. Literature review

2.1. Determinants of project performance

Since the 1960s, project management researchers have been attempting to find the factors that determine project success (Cooke-Davis, 2002).[1] Baker et al. (1974) added the issue of client satisfaction to the traditional triangle of project success criteria: cost, time, and quality. Project success criteria, therefore, became a square of criteria (Ika, 2009): cost, time, quality, and client satisfaction. Then, other authors such as Shenhar et al. (1997), Baccarini (1999), and Lim and Mohamed (1999) added more criteria to the traditional success criteria dimensions, including acknowledgment of the strategic objectives of the client organization that started the project, end-user satisfaction, and stakeholder satisfaction.

Although the literature generally agrees on success criteria measures, these criteria have been largely criticized particularly in the context of complex projects (Rezvani et al., 2016). This is because such criteria tend to draw on shortsighted constructs that do not reflect the experience in large, complex projects (Toor and Ogunlana, 2010). Also, these criteria fail to address more extensive elements such as behavioral skills or strategic management objective criteria (Jugdev and Müller, 2005). In the 1990s, authors began to conduct studies demonstrating that project success is a multi-dimensional context and consequently new models for project performance management should mirror the multidimensionality of a project (Todorovic et al., 2015). In this view, which is more people-focused, success is estimated by the behavioral and interpersonal skills of project teams as well as stakeholder and client satisfaction (Pinto, 1990; Jugdev and Müller, 2005; Mazur et al., 2014). Bryde (2005) demonstrates that the determination of key performance indicators with no respect for the project team and the organizational environment in which the project is implemented might lead to serious impediments to enhancing the performance of a project. These are the reasons that necessitate considering success factors.

Success factors are measures related to a management system that directly or indirectly impact the success of a project (Wit, 1988). Success factors center around "soft" issues (Rezvani et al., 2016), such as the behavioral skills of a project team and the satisfaction of stakeholders and clients; hence, success factors demonstrate a more realistic, dynamic way to deal with project success (Pinto, 1990; Jugdev and Müller, 2005). Research on critical success factors started by concentrating on various characteristics of project control. Baker et al. (1974) recommended replacing the triangle of cost, time, and quality by a measure called "perceived success". Later, Slevin and Pinto (1986) proposed a scientific

[1] It is important to distinguish between success criteria and critical success factors. According to Wit (1988), success criteria are related to measures by which success or failure of a project will be decided. Success criteria center around measures such as cost, time, and quality (Pinto and Slevin, 1987). Success factors, on the other hand, are measures related to a management system that directly or indirectly impact the success of a project (Wit, 1988).

basis for project success that is manageable by the project team (Ika, 2009). This scientific basis for project success includes ten critical success factors: project mission, top management support, project schedules/plan, client consultation, personnel, technical tasks, client acceptance, monitoring and feedback, troubleshooting, and communication. This list was subsequently extended by Pinto and Slevin (1988) with four additional factors: characteristics of the project team leader, power and politics, environmental events, and urgency. The lists of most-repeated critical success factors were introduced by Cooke-Davis (2002), Jugdev and Müller (2005), Fortune and White (2006), and Ika et al. (2012). However, the general conclusion is that there is no list of critical success factors that is valid for all projects (Todorovic et al., 2015).

Many other success factors and frameworks have been proposed by different authors; the behavioral skills and personality traits of the individuals who perform the project is one of these success factors. It is fast becoming accepted philosophy that it is individuals who deliver projects, not systems or procedures (Cooke-Davis, 2002). As the title of the paper by Lechler (1998) says, "When it comes to project management, it's the people that matter."

One of the many success factors that contributes to project success is the entrepreneurial orientation of individuals. A recent study by Martens et al. (2018), using a survey conducted among project managers in Brazilian organizations, showed that EO explains 20% of the effects on project success. Ahmed et al. (2014) used data collected from IT firms to demonstrate that having entrepreneurial individuals among team members enhances project performance. In addition, by collecting data from project managers in the Brazilian software market, Martes et al. (2015) showed that there is a positive relationship between project management and entrepreneurial orientation through integration, scope, time, cost, quality, human resources, communications, risk, and procurement management. As shown in all these examples, the EO impact on project performance is not limited to entrepreneurship projects; any project may benefit from the entrepreneurship skillset.

Apart from the direct impact of the entrepreneurial orientation of individuals on project success, there is evidence in the literature showing the positive impact of different dimensions of entrepreneurial orientation on project success. For example, in a study by Vezzoni et al. (2013) that aimed to identify project success factors, one of the most important factors enhancing project success was identified as preparation to confront risks, which might be related to the risk-taking dimensions of entrepreneurial orientation. Innovation capability, which is another dimension of entrepreneurial orientation, also has been shown to contribute to long-term project success (Biedenbach and Müller, 2012). Teams frequently need innovation to succeed in a project, and a positive team environment for innovation results in better project performance: teams that have a positive orientation for innovativeness are able to complete projects faster than teams that do not have it (Pirola-Merlo, 2010). Another dimension of entrepreneurial orientation is proactiveness. Kerzner (2004) points out proactivity as something expected of project managers that can add to the achievements of the project. Also, self-efficacy, which is related to entrepreneurship attitude orientation, has been identified in the project management literature as having a potential influence on several aspects of project performance (Dainty et al., 2003), commitment to the project (Jani, 2011), and knowledge sharing (Lin and Huang, 2010). As a result, there is enough evidence in the project management literature to show that there is a positive relationship between entrepreneurial orientation and project success. However, there is a lack of understanding of this relationship (Kuura et al., 2014; Lundin et al., 2015); this study addresses this gap. In the next section, we will elaborate on entrepreneurship orientation and its dimensions.

2.2. Entrepreneurship orientation

Entrepreneurship orientation is defined as "a firm's strategic posture towards entrepreneurship" (Anderson et al., 2015, p. 1579). EO is identified with the essential arrangements and practices for the improvement of entrepreneurial activities, along with the choices and procedures that decision-makers use to upgrade their organization's goals, bolster their vision, and create competitive advantage (Rauch et al., 2009; Freitas et al., 2012). The behaviors that characterize EO include innovativeness, risk-taking, proactiveness, competitive aggressiveness, and autonomy (Lumpkin and Dess, 1996). Most of the research in the context of EO has paid attention to innovativeness, risk-taking, and proactiveness; competitive aggressiveness and autonomy have been examined less regularly (Lyon et al., 2000; Rauch et al., 2009). Depending on the area of research, it has been found that these five dimensions of EO can be considered either together (Runyan et al., 2008; Lumpkin et al., 2009) or separately (Lumpkin and Dess, 1996, 2001; Wang, 2008).

In this study, we will consider the three dimensions of EO that have been examined the most: innovativeness, risk-taking, and proactiveness. Innovativeness is an inclination to take part in experimentation through research and development (Rauch et al., 2009) and to help activities that can lead to new processes, services, or products (Lumpkin and Dess, 1996). Risk-taking is associated with the desire to acquire significant outcomes (Lumpkin and Dess, 1996) and is near to that of innovativeness, including bold activities that venture into the obscure or dedicate noteworthy assets to questionable endeavors (Rauch et al., 2009). Risk-taking has been noticeably studied in the literature, and its positive impacts on the characteristics of entrepreneurs have been investigated (e.g., Sexton and Bowman, 1983; Schwer and Yucelt, 1984; Wong et al., 2005; Gürol and Atsan, 2006).

Proactiveness is described as an organization's propensity to be in front of the challenge when it comes to launching new technologies, services, or products, as opposed to simply pursuing market activities (Miller, 1983). It is associated with the ability to envision and look for new opportunities (Lumpkin and Dess, 1996; Setiawan et al., 2015) and with the desire for a share in developing markets (Martens et al., 2018). It is defined as "an opportunity-seeking, forward-looking perspective characterized by the introduction of new products and services ahead of the competition and acting in anticipation of future demand" (Rauch et al., 2009, p. 763).

Additionally, entrepreneurship attitude orientation (EAO) is regarded as an important concept in entrepreneurship literature. Gasse (1985) and Douglas and Shepherd (2002) found that a progressive attitude toward risk and independence prompts more grounded entrepreneurial intentions. According to Gasse (1985), entrepreneurs frequently have a more noteworthy internal locus of control than the general population. According to trait theory, entrepreneurs have unique personalities and characteristics, so researchers can distinguish entrepreneurs from non-entrepreneurs by developing methods to locate these characteristics (Low and MacMillan, 1988; Scherer et al., 1989).

Research on individual entrepreneurial orientation has shown that self-esteem is one of the basic personality traits that is correlated with the intention to become an entrepreneur (Harris and Gibson, 2008). It has been shown that factors related to entrepreneurship attitude and self-esteem play a significant role in achieving higher individual entrepreneurship orientation (Vogelsang, 2015). In the realm of entrepreneurship, self-efficacy is an important trait of an entrepreneur because it addresses an individual's capacity to actualize or execute explicit plans and activities to accomplish explicit results in an explicit setting (Erikson, 2002; Goddard et al., 2004; Henry et al., 2005). Robinson et al. (1991) consider self-esteem as one of the items measuring entrepreneurship attitude. In our study, we will examine self-esteem in social efficacy, appearance, and comparativeness as the dimensions of entrepreneurship attitude orientation.

2.3. Project performance evaluation methods

Existing studies on project performance evaluation have used various methods that depend on various sets of evaluation measures and factors. For instance, earned value management (EVM) is a traditional method in the context of project management that enables program managers, project managers, and other top-level stakeholders to evaluate and visualize the status of a project during the project life cycle (Salari and Khamooshi, 2016). Multicriteria decision making is another technique to evaluate the overall performance of projects; this technique has been used to aggregate multiple performance criteria under different application settings (Pillai et al. (2002); Marques et al. (2011); Barfod (2012)). In assessing the relative performance efficiency of the finished projects, data envelopment analysis (DEA) has been broadly used as a performance measurement technique that incorporates input and output impactful variables (Linton and Cook, 1998; Busby and Williamson, 2000; Verma and Sinha, 2002; Revilla et al., 2003; Stensrud and Myrtveit, 2003; Eilat et al., 2006, 2008; Farris et al., 2006; Vitner et al., 2006; Cao and Hoffman, 2011).

These traditional approaches may not be sufficiently effective to reflect all the dimensions of projects. For example, the EVM technique is based only on the cost deviation and schedule deviation of projects; there are many other dimensions that may directly or indirectly affect project performance. According to Marques et al. (2011), in a project management context, new methods are needed to predict project performance based on different multiple criteria. Although multi-criteria decision-making methods consider multiple criteria to assess project performance, these criteria are based on decision-makers' preferences that might be subjective. Furthermore, these methods are based on providing some potential decisions and selecting the best decision provided. However, since the potential decisions are created by decision-makers, they may not encompass all possible options and criteria. There are shortcomings associated with Data Envelopment Analysis (DEA) methods as well, since DEA techniques are based on a benchmarking approach. One possible decision or criteria that works well for one project may not be compatible with another project because each project has its unique characteristics.

Thus, shortcomings in currently used methods suggest a need and opportunity for improvement in a project management context for project performance prediction (Cheng et al., 2010). As Putka et al. (2018) suggested in their recent study, there have been significant advances in predictive modeling from computer science and machine learning that can benefit management and organizational studies.

2.3.1. The advantages of machine learning and predictive analytics

The advantages of machine learning techniques over traditional methods in predicting project performance have been addressed in project management literature. For example, in predicting software development project performance, Chulani et al. (1999) have shown Bayesian analysis has a significantly better performance than traditional multiple regression models, as Bayesian networks take advantage of a formal learning mechanism for prediction. In another study by Dvir et al. (2006) that aims to predict defense project success, the authors compared the prediction performance of neural networks with a traditional linear regression model. Their results showed that neural networks predict project success more accurately than the traditional regression model, as a neural networks model can better fit the data. In the same vein, Wang and Gibson Jr (2010) have compared the performance of neural networks with simple linear regression in predicting construction project success; their results showed that neural networks produce better prediction results. Furthermore, in a literature review by Paliwal and Kumar (2009) that is based on the results of 73 papers, they found that in 56% of cases, a neural network as a machine learning technique outperformed the traditional methods.

Using a machine learning approach based on survival analysis, Li et al. (2016) have identified the weakness of the traditional regression model in predicting the performance of crowdfunding projects: traditional regression models ignore the valuable information embedded in failed projects and are not capable of learning from failures. Finally, in their recent literature review, Martínez and Fernández-Rodríguez (2015) showed that artificial intelligence and machine learning methods work better than traditional techniques in estimating project performance, as they are able to more effectively deal with project uncertainty and today's complex environment.

In conclusion, based on the evidence in the project management literature, machine learning techniques can potentially work better than traditional methods (including regression models) for several reasons: 1) they have a learning mechanism to learn and communicate past knowledge (Chulani et al., 1999), 2) they may work well in situations where the data is scarce and incomplete (Chulani et al., 1999), and 3) unlike traditional models, they have the capabilities of self-learning and self-updating, which make them much simpler and more effective to establish the relationship between the input and the output (Dvir et al., 2006; Wang and Gibson Jr, 2010). In this study, we aim to address the shortcomings in the methods used for predicting project performance by applying modern methods in a management and organizational context using machine learning.

2.4. Machine learning

In machine learning approaches, knowledge is extracted by training the model (Wang et al., 2011); there is no need for benchmarking or assumptions regarding the importance of each variable. When using traditional prediction methods, ambiguity regarding the relationship between influential factors and project performance complicates the problem of predicting project performance. Is the relationship linear? If nonlinearity is needed, which functional form should be assumed? Must interaction between the features be considered? In contrast to traditional prediction methods, machine learning is well suited for problems with ambiguous functional forms (Gu et al., 2018).

Machine learning enables computer programs to recognize and acquire information from the real world and improve the execution of some assignments based on the new information (Portugal et al., 2018). A broadly used definition of machine learning is provided by Mitchell (1997, p. 2): "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E". Although computer programs simulate human learning, the process of learning is not by reasoning but by algorithms. These algorithms can be categorized based on the approach they use for learning: supervised, unsupervised, semi-supervised, and reinforcement learning (Portugal et al., 2018). The category of supervised learning (Kotsiantis, 2007; Zhang and Tsai, 2007) is used when training data and correct responses are given to algorithms. The output of the algorithms would be learning the training data and predicting the responses of unseen data. The category of unsupervised learning is used when there is no training data and subsequently, there are no correct and seen responses. In this case, algorithms need to learn the data and try to find hidden patterns (Celebi and Aydin, 2016). The third category is semi-supervised learning: algorithms are provided with training data, but the data is missing some information. The machine learning algorithms in this category need to learn and predict responses, even though the training data is incomplete (Chapelle et al., 2006). Finally, the category of reinforcement learning refers to cases when algorithms learn using the external response provided either by the environment or by a thinking entity (Sutton and Barto, 2018). An example is teaching dogs when they receive a treat for correct action and no treat for a wrong action (Portugal et al., 2018).

In this study, the response (project performance) is known, so our problem falls into the category of supervised learning (James et al., 2013), and we need to select machine learning algorithms suitable for this category. We will use four predictive modeling techniques to predict the performance of projects based on the degree of entrepreneurial orientation. The four techniques were selected based on their popularity in the context of machine learning (Oztekin et al., 2013) as well as evidence of their applicability in the context of entrepreneurship. The four techniques are multiple linear regression (lasso and ridge), support vector machines (support vector regressor), neural networks, and random forests (Fellnhofer et al. (2016); Iskender and Batı, 2015; Li (2018); Xu et al. (2018)). Below we provide an overview of these techniques in management research.

Multiple linear regression. In social and behavioral science research, multiple linear regression has been considered as a powerful technique for identifying hidden complex correlations among data (Huberty, 1989; Fox, 1991). Multiple linear regression has also been a popular technique in the context of entrepreneurial studies. For example, to measure the factors contributing to corporate entrepreneurship cultivation and examine the relationship between corporate entrepreneurship and innovation performance, Chen et al. (2005) used a series of multiple linear regression analyses and identified the factors contributing to corporate entrepreneurship. Fellnhofer et al. (2016) conducted a study using multiple linear regression to examine the different perceptions of the entrepreneurial orientation of females compared to those of their male counterparts. Their multiple linear regression analysis revealed that in different industries, the entrepreneurial orientation differs between genders. In a study conducted by Thapa (2015) to identify microenterprise performance determinants, the authors used multiple linear regression and found entrepreneur-related factors were the key factors determining the performance of microenterprises in Nepal. The last example is a study by Yao et al. (2016), who examined the impact of Chinese university students' perceived entrepreneurial environment on their entrepreneurial tendency in the context of Chinese economic transition; the researchers found multiple linear regression a proper method for their analysis.

Multiple linear regression analyzes the correlation between a set of predictor/independent variables and a single response/dependent variable that are supporting research questions that deal with prediction, theory testing, or explanation (Cohen et al., 2014). The traditional linear regression uses ordinary least square (OLS) estimates that are obtained by minimizing the residual squared error; however, the limitations regarding the prediction accuracy and interpretation of OLS estimates have often caused data analysts not to be satisfied with OLS estimates (Tibshirani, 1996). Thus, lasso and ridge, which are extensions of simple linear regression (James et al., 2013), have been introduced to improve the OLS estimates by shrinking the coefficients and retaining the more important features. For more details about lasso and ridge, we refer the reader to Tibshirani (1996) and Kim et al. (2007).

Artificial neural network. In recent years, the technique of artificial neural networks has become a popular method for analyzing data related to organizational contexts (Minbashian et al., 2010). This technique has been extensively applied to behavioral research in organizations (Scarborough and Somers, 2006). For example, Somers (2001) and Somers and Casal (2009) applied neural networks to investigate the relationship between work attitudes and job performance. Similarly, Palocsay and White (2004) used a neural network to study the relationships between dimensions of culture and perceptions of organizational justice. As another example, Minbashian et al. (2010) applied a neural network in the context of research on personality and work performance; this study showed some of the benefits of this method compared to multiple regression for conducting exploratory research.

The findings from all these studies suggest that neural networks perform at least as well if not better than traditional techniques and, in some cases, lead to an intuitive understanding that would not be detected otherwise. Furthermore, throughout many entrepreneurship studies, neural networks have been used in a variety of ways, including innovation and entrepreneurship education (Li, 2018) and business failure (Williams, 2016). Neural networks enable researchers to capture nonlinear and configural relationships between sets of predictors and response variables (Bishop, 1995). Neural networks are made from layers of units that are interconnected by the means of weights; an input layer represents the predictor variable, and an output layer represents the response to be predicted (Minbashian et al., 2010). For more detail about neural networks, we refer the reader to Haykin (2009).

Support **vector machines**. Support vector machine is a machine learning technique that is very effective in pattern recognition and has been used by scholars in the context of entrepreneurial studies. For example, Nasution et al. (2018) used support vector machines to predict the entrepreneurial intention of graduates and alumni. Marijana et al. (2010) developed a model using support vector machines to classify first-year students and predict their entrepreneurial intention based on a survey collected from a Croatian university. Iskender and Batı, 2015 used support vector machines to classify universities based on the degree to which they are oriented toward entrepreneurship and innovation. The support vector machine technique was first developed by Vapnik (1998) and used to create predictor-output mapping functions based on a training set (Oztekin et al., 2013). It is a nonlinear generalization of the Generalized Portrait algorithm developed in the 1960s (Vapnik and Lerner, 1963). It is well-grounded in statistical learning theory, which has been developed by Vapnik (1982, 1995). For more detail about support vector machines, we refer the reader to Smola and Schölkopf (2004).

Random forests. The random forest technique has been used in a variety of studies in the context of entrepreneurship. For example, Xu et al. (2018) studied the relationship between enterprise credit and entrepreneur credit; their results showed the random forest technique has good performance in developing model robustness. Carter et al. (2019) used random forests to search for the source of heterogeneity in entrepreneurial programs in Nicaragua and identified the benefits of these programs to households. In a study by Tu et al. (2019), the application of random forests combined with support vector machines and a sine cosine algorithm was shown to predict college students' entrepreneurial intention in advance. Random forests are a set of tree predictors; each tree depends on the values of random vectors sampled independently and with the same distribution for all trees (Breiman, 2001). Based on a preset criterion, random forests iteratively split the data into branches to increase the prediction accuracy, which consequently results in a tree-shaped structure (Quinlan, 1986). For more detail about this algorithm, we refer the reader to Breiman (2001).

3. Methodology

In this study, we aim to predict project performance based on the degree of entrepreneurial orientation of the individuals in the project. In order to address this problem, we use four sets of machine learning algorithms: multiple linear regression (lasso and ridge), support vector machines, neural networks, and random forests. For support vector machines, the use of different kernel functions generates various algorithms; in this study, we will use four different kernels: linear, rbf, sigmoid, and polynomial. For neural networks, we will use multi-layer perceptron (MLP) neural networks with different layers, particularly, 1, 2, 3, 4, and 5 layers.

One of the important preliminary steps of machine learning is to understand the splitting configuration for estimation and hyperparameter tuning (Gu et al., 2018). Hyperparameter tuning includes, for example, deciding the number of trees in the random forests method, the alpha in lasso and ridge, and the minimum sample leaf in neural networks. One of the common approaches in splitting data is to select tuning parameters from the validation set (Gu et al., 2018). In prediction models, there are no globally accepted best parameters that work for every any kind of problem. Thus, the best model and the best parameters for each model depends on the scenarios and the data being used (Oztekin et al., 2013). However, we can obtain the best parameters through trial and error experimentation (Ruiz and Nieto, 2000).

More specifically, we examined different values of alpha for the two methods of lasso and ridge, different values of C and gamma for the support vector machines method, different values of alpha with different learning rates, solver, and activation function for the neural networks method, and different values for minimum sample leaf and number of estimators for the random forests method. The best parameters in terms of minimizing the mean squared error were selected. We split our sample into three sets: training, validating, and testing. The validation set is used for parameter tuning, the training set is used for fitting the model, and the testing set is used to evaluate the performance of each of the models in predicting the response.

Since the best configuration of splitting differs for each problem, in the case study that we will present in the following section, we have considered different configurations and evaluated each one in order to get the best splitting scheme. Fig. 1 shows an overview of our methodology.

3.1. Model evaluation

To compare the performance of the aforementioned predictive models, we used the mean squared error (MSE) given by Equation (1); this is a popular method for analyzing the predictive performance of machine learning algorithms (Botchkarev, 2018). In Equation (1), y_i is the actual output variable, $\hat{y}_i$ represents the predicted output variable, and n is the total number of samples. MSE is used as a metric to select the best model; a smaller MSE indicates a better prediction by the model (Makridakis et al., 2008).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{1}$$

Since this research is a human-related study, we used the correlation between the predicted performance and the actual performance as another metric for evaluating the performance of each of the predictive models. For human-related studies, it is recommended that the correlation should be at least 0.3 (Cohen et al., 2014). We expect that better predictive models will have higher correlation values. The correlation formulation is given by Equation (2). We can use the correlation between the estimated output variable and the actual output variable in order to determine the satisfactory models. In Equation (2), $\overline{y_i}$ is the mean of the actual output variable, $\overline{\hat{y}_i}$ is the mean of the estimated output variable, s_y is the standard variation of the actual output variable, and $s_{\hat{y}}$ is the standard deviation of the estimated output variable.

$$r_{\hat{y},y} = \sum_{i=1}^{n}\frac{(y_i - \overline{y_i})(\hat{y}_i - \overline{\hat{y}}_i)}{(n-1)\left(s_y * s_{\hat{y}}\right)} \tag{2}$$

3.2. Statistical comparison of machine learning models

After obtaining the performance results of each machine learning algorithm, we take another step to decide whether there is a significant difference between the algorithm performances. In order to address this issue, we use statistical testing. According to Demšar (2006), when we do not have information about the distribution of samples and the sphericity of samples, the best approach for comparison of multiple machine learning algorithms is a non-parametric statistical test. The Friedman test was introduced by Friedman (1937, 1940), and many machine learning studies have used it as a method for comparing the performance of different algorithms (e.g. Vink and Haan, 2015; Malhotra, 2016; Douzas and Bacao, 2018). The Friedman test is a non-parametric method for comparison of multiple sample means that is capable of ranking machine learning algorithms for each data set separately (Demšar, 2006).

Equation (3) shows the Friedman statistic, which is distributed according to χ_F^2 with k-1 degrees of freedom. N is the number of datasets and R_j is the rank allocated to the jth algorithm and calculated using Equation (4). In Equation (4), r_i^j is the rank of the jth machine learning algorithm on the ith dataset. The best performing algorithm is assigned the rank of 1; in case of equality in performance, the average rank is assigned (Demšar, 2006; Trawiński et al., 2012; Brown and Mues, 2012; Zhang and Suganthan, 2016).

$$\chi_F^2 = \frac{12N}{k(k+1)}\left[\sum_j R_j^2 - \frac{k(k+1)^2}{4}\right] \tag{3}$$

$$R_j = \frac{1}{N}\sum_j r_i^j \tag{4}$$

In the Friedman test, the null hypothesis states that all the R_j are equal, i.e., all the algorithms are equivalent in terms of performance (Demšar, 2006). According to Demšar (2006), if the null hypothesis is rejected, we need to do a post-hoc test that will enable us to identify the algorithms with different performances. This post-hoc test is called the Nemenyi test (Nemenyi, 1963); it is similar to the Tukey test for ANOVA. Using this test, we would be able to understand which pairs of algorithms have different performance. The Nemenyi test includes the calculation of a critical distance, which is then compared to the differences between the average ranks of the two algorithms assigned by the Friedman test (Malhotra, 2016). Equation (5) shows the critical distance for the Nemenyi test.

$$CD = q_\alpha\sqrt{\frac{n(n+1)}{6k}} \tag{5}$$

In Equation (5), n corresponds to the total of the number of compared algorithms, k is the number of datasets on which the comparison is performed, and q_α is the studentized range statistic divided by $\sqrt{2}$.

3.3. Feature importance

After identifying the best predictive model, we analyze the importance of features on project performance using two criteria: the Bayesian Information Criterion (BIC) (Schwarz, 1978) and the Akaike Information Criterion (AIC) (Akaike, 1974). These are two of the most popular methods for identifying the importance of predictors that are believed to be related to the response or output (James et al., 2013). BIC and AIC both provide information regarding the combination of features that best explain a model with the least prediction error. The most significant subset of features is associated with the lowest BIC or AIC values. The

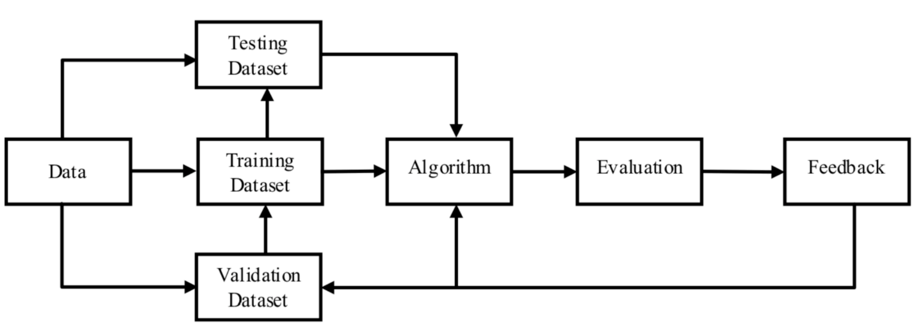

Fig. 1. An overview of the machine learning approach.

procedure for selecting the most significant features is shown in Table 1. For more information regarding these criteria, we refer the reader to James et al. (2013).

4. Case study

4.1. Data collection

Our case study is based on data collected from university students who were working on projects related to their coursework. The reason we selected university students for our case study is threefold: First, it is generally acknowledged that institutions are the rules of the game in a society; in other words, institutions are the humanly devised constraints that form human interaction (North, 1990). The institutional context influences the performance of economies, especially through the impact on the entrepreneur's behavior (Veciana et al., 2005). According to Institutional Economic Theory, "institutions include any form of constraint that human beings devise to shape their interaction" (Veciana et al., 2005, p. 165), particularly in terms of attitude and behavior. Since universities are one of the most important institutions in society, they can be a good source of understanding entrepreneurship behavior. Second, according to Kuratko (2005), the most entrepreneurial generation of the 21st century is the younger generation, with age between 18 and 34. As stated by Thomas and Mueller (1998) and Gürol and Atsan (2006), a significant portion of the pool of potential entrepreneurs is related to university students. Therefore, they can be considered as a reliable source for studies in the context of entrepreneurial behavior. Third, higher education students have shown an ongoing interest in entrepreneurship education in order to prepare themselves for today's competitive environment (Küttim et al., 2014). In fact, university students can be future entrepreneurs if they are not already (Yurtkoru et al., 2014).

Hence, a survey instrument was developed to collect data on student entrepreneurship skillsets. The survey captures measures developed based on trait theory, entrepreneurship orientation, and individual entrepreneurship orientation that are suitable to assess a student's entrepreneurship skillsets. We used a measurement scale that was created by Bolton and Lane (2012) and modified by Vogelsang (2015) to test the validity of the scales for entrepreneurial orientation. Items were measured using a seven-point Likert scale (1 = strongly disagree, to 7 = strongly agree). Apart from features related to EO and entrepreneurship attitude, we considered the demographic/individual characteristics of the students, including race, GPA, gender, and contextual/project characteristics that include the context of the course project. The list of features used in the prediction of student performance is presented in Table 2.

The number of samples we obtained is 185; each sample contains 10 features representing the predictors, and the student performance in the course project represents the response variable. In this study, student performance is defined as an individual's overall subjective evaluation of his/her performance in the course project. Table 3 presents a summary of the descriptive statistics of the sample.

4.2. Survey administration

The sample for this study contains undergraduates at a university in the southeastern United States during the 2017–2018 academic year from the following departments: BIOL (biology), CHEM (chemistry), CM (construction management), CST (computer systems technology), GCS (graphic design technology), and AET (applied engineering technology). Table 4 provides some characteristics of the participants in this study.

Prior to administering the survey, the research team identified courses that have the project as an integral part of the course. This information was obtained by contacting the Dean of the school and through communicating with the instructors of the courses. The instructor of the courses was also informed about the nature of the project and was asked whether the project is an integral part of the courses. After further screening, several courses were identified for survey administration from the College of Science and Technology and the College of Engineering.

Data collection was conducted through administering a paper-based survey during class time. Prior to administering the survey, one of the research team members explained the significance of the project to the class, in order to motivate students to complete the survey.

We received 243 completed surveys from the students. We eliminated 58 responses due to missing values. Thus, our total sample size for this study is 185. This provides a response rate of 76.1%, which is significantly greater than the average response rate of 52.7% for empirical research from individuals in management research (Baruch and Holtom, 2008).

Since the complex relationship between the response variable and the features that are listed in Table 2 is not known a priori, the next step is to find a predictive model that best describes this relationship.

Table 1
Algorithm for identifying the most significant features using BIC (AIC).

Selecting the most significant features
1 Assume M_0 is the null model containing no features.
2 Fit all the models containing exactly k features for k = 1, 2, …, p.
3 Calculate the BIC (AIC) for all the models and select a set of features with the smallest BIC(AIC).

Table 2
Survey constructs.

Feature	Description
Entrepreneurship features	
Risk-taking (RT)	Individual's willingness to take a personal risk and make commitments
Innovativeness (Inn)	Individual's creativity and the ability to pursue new opportunities
Proactiveness (Pro)	Individual's ability to anticipate future problems or demands
Appearance self-efficacy (App)	Individual's overall subjective evaluation of his/her appearance
Social self-efficacy (Soc)	Individual's overall subjective evaluation of his/her social value
Comparativeness (Comp)	Individual's subjective evaluation of him/herself compared to others
Demographic/Individual features	
Race (R)	Individual's race
Gender (G)	Individual's gender
GPA	Individual's GPA
Contextual feature	
Project context (P)	The context of the project that the student is working on

4.3. Comparison of machine learning performance

As the nature of machine learning algorithms is to learn from data, often no specific assumptions regarding the characteristics of data are made. Thus, in order to choose a model that best fits the nature of the data and more accurately predicts project performance, we need to compare different algorithms so that we would understand which algorithm and which model is able to predict the unseen data with higher accuracy. Table 5 and Table 6 present the comparison of machine learning techniques we used with respect to their MSE and correlation, respectively. All the algorithms are coded in Python. As explained in the methodology section, we split our data set into three sets: validating, training, and testing. The validating set is used for tuning the model parameters, the training set is used for fitting the model, and the testing set is used for testing the performance of the models. Since the best splitting configuration differs for each problem, we examined 20 splitting schemes. The first column of Table 5 shows the splitting schemes we

Table 3
Descriptive statistics.

	count	mean	Std.	min	25%	50%	75%	max	Kurt	skew
RT	185	5.24	1.11	1.33	4.67	5.33	6.00	7.00	0.65	−0.68
Inn	185	5.06	0.95	1.75	4.50	5.00	5.75	7.00	0.71	−0.48
Pro	185	5.72	1.03	2.67	5.33	6.00	6.33	7.00	0.35	−0.92
App	185	5.24	1.29	1.33	4.33	5.33	6.33	7.00	−0.40	−0.48
Soc	185	3.67	1.12	1.43	2.86	3.86	4.43	6.14	−0.70	−0.09
Comp	185	5.49	0.96	3.00	4.83	5.50	6.17	7.00	−0.40	−0.39
GPA	185	3.14	0.47	1.80	2.90	3.20	3.50	4.00	0.02	−0.39
G	185	0.44	0.50	0.00	0.00	0.00	1.00	1.00	−1.97	0.23
P	185	2.42	1.63	1.00	1.00	2.00	4.00	6.00	−0.36	0.93
Performance	185	4.97	1.07	2.29	4.14	5.00	5.71	7.00	−0.49	−0.13

Table 4
Sample distribution.

Department	Course Title	Course Code	Number of Responses
Biology	General Microbiology	BIOL 221	77
Chemistry	General Chemistry	CHEM 106 CHEM 117	41
Construction Management	Introduction to Construction Management Senior Seminar	CM 100 CM 400	16
Computer Systems Technology	Senior Project I: A Capstone Experiment	CST 498	26
Graphics Design Technology	Senior Capstone for Graphics	GCS461	9
Applied Engineering Technology	Manufacturing Planning and Management	AET 232	16

used for this study. For example, 80:10:10 indicates that we used 80% of the data for training, 10% for validating, and 10% for testing. In total, we compared 12 models: lasso; ridge; support vector regressor (SVR)-linear (SVR-l); SVR-rbf (SVR-r); SVR-poly (SVR-p); SVR-sigmoid (SVR-s); neural networks with one to five layers (NN1, NN2, NN3, NN4, and NN5); and random forests (RF).

In Table 5, the numbers in parentheses show the ranking of the algorithm in each of the splitting configurations, with 1 as the highest rank and 12 as the lowest. The last row of the table shows the overall ranking of each algorithm using Equation (4). The results show that the three best methods in terms of minimizing MSE are lasso, ridge, and NN1, with an MSE value of 0.524 with a 58:21:21 split for lasso and ridge, and a 66:17:17 split for NN1. Also, the results show that as we decrease the percentage of training and increase the percentage of validating and testing samples, approximately we get better MSE and correlation values for lasso, ridge, NN2, and the random forests methods. However, this trend stops when the decrease gets to 56% for training samples. In contrast, other methods display a chaotic response to the percentage of training, validating, and testing. As shown in Table 5, based on average ranking (R_j), lasso has the best performance in terms of MSE.

As shown in Table 6, the correlation values for most of the prediction models and most of the splitting schemes are well above the recommended value of 0.3 for human-related studies (Cohen et al., 2014). However, NN1 results in negative values for all the splitting schemes, which means it is not a promising method for explaining the relationship between the testing samples and the predicted response. The best result in terms of correlation was obtained from the SVR-r method with a value of 0.695.

4.4. Statistical test on the performance of machine learning algorithms

In order to understand whether the differences between the performance of machine learning algorithms are random or there are some algorithms that perform better compared to the others, we compared the difference in performance among the machine learning algorithms using the Friedman test. We assumed the significance level (α) for our statistical analysis is 0.05 and conducted the Friedman test for the MSE results of the algorithms. As shown in Table 5, each of the twelve algorithms is ranked based on their MSE performance (rank number in parentheses) and the average ranking for each of the algorithms is calculated using Equation (4) and presented in the last row of Table 5. Since twelve algorithms are being compared, the degrees of freedom is 11. As shown in Table 7, the computed Friedman statistic is 83.654.

The null hypothesis of the Friedman test states that there is no significant difference between the MSE of the machine learning algorithms. However, as shown in Table 7, the p-value is less than 0.05, which means we reject the null hypothesis and accept the alternative hypothesis, which states there is a significant difference between the performance of some of the algorithms with respect to MSE. The results from the average rank of each of the twelve algorithms show that lasso is the best algorithm followed by ridge and SVR-s. NN1, NN3, and NN5 received the lowest ranks among the machine learning algorithms. Thus, the results confirm the possibility of differences between some of the algorithms.

After obtaining the Friedman test results, which show a significant difference in the performance of the algorithms, we proceed in our analysis by doing a Nemenyi post-hoc test (Demšar, 2006) in order to understand which pairs of algorithms have significantly different performance.

We assumed the significance level for the Nemenyi test is 0.05. The critical distance computed for the Nemenyi test is 3.726 at $\alpha = 0.05$. The distance between the average rank of each of the algorithms is calculated for each pair of the algorithms and then compared to the Nemenyi critical distance. If the difference is greater than or equal to the Nemenyi critical distance, we conclude that there is a significant difference between their performances at $\alpha = 0.05$.

The results of the pair-wise comparison of the machine learning algorithms are presented in Table 8. Out of 66 pairs of algorithms, 14 were found to have significantly different performance. As shown in Table 8, lasso is significantly superior to all the neural network algorithms (NN1, NN2, NN3, NN4, and NN5); ridge is significantly superior to NN3, NN4, and NN5; SVR-l, SVR-r, SVR-s, and RF are significantly superior to NN1; and SVR-s is significantly superior to NN3 and NN5 (and NN1, already mentioned in this sentence). Table 8 also confirms that lasso is the top machine learning algorithm, with the highest number of significant pairs.

Based on the results of Table 8, Fig. 2 shows for each algorithm the number of pairs in which that algorithm outperforms. As shown in Fig. 2, lasso is significantly better than five other methods: NN1, NN2, NN3, NN4, and NN5. Ridge and SVR-s are each significantly better than three other methods. SVR-l, SVR-r, and RF are each significantly better than one other method. SVR-p, NN1, NN2, NN3, NN4, and NN5 do not significantly outperform any other algorithms.

Contrary to our initial expectation, we see multiple linear regression models, particularly lasso and ridge, have better performance in learning and capturing the training data. As support vector machines,

Table 5
MSE prediction performance with an average ranking.

Splitting	lasso	ridge	SVR-l	SVR-r	SVR-s	SVR-p	NN1	NN2	NN3	NN4	NN5	RF
80:10:10	0.691 (1.5)	0.691 (1.5)	0.749 (5)	0.832 (8)	0.744 (4)	0.898 (9)	1.188 (11)	1.188 (11)	0.788 (6)	1.188 (11)	0.830 (7)	0.727 (3)
78:11:11	0.767 (2.5)	0.767 (2.5)	0.776 (6)	0.773 (5)	0.772 (4)	0.765 (1)	0.972 (10)	1.202 (11)	1.324 (12)	0.947 (9)	0.812 (7)	0.926 (8)
76:12:12	0.690 (2)	0.691 (3)	0.897 (8)	0.696 (5)	0.695 (4)	0.643 (1)	1.129 (10.5)	1.129 (10.5)	0.948 (9)	0.719 (6)	1.144 (12)	0.873 (7)
74:13:13	0.753 (2)	0.752 (1)	0.783 (3)	1.093 (9.5)	1.093 (9.5)	2.653 (12)	1.108 (11)	0.895 (8)	0.843 (7)	0.837 (5)	0.842 (6)	0.795 (4)
72:14:14	0.702 (4)	0.701 (3)	0.906 (7)	1.062 (10.5)	0.750 (5)	1.062 (10.5)	1.089 (12)	0.663 (1)	1.001 (9)	0.694 (2)	0.934 (8)	0.805 (6)
70:15:15	0.605 (2.5)	0.605 (2.5)	0.601 (1)	0.739 (9)	0.617 (4)	0.720 (8)	0.621 (5)	0.645 (7)	0.958 (11)	0.643 (6)	1.039 (12)	0.786 (10)
68:16:16	0.567 (1)	0.568 (2)	0.729 (4)	0.911 (8)	0.911 (8)	0.911 (8)	0.910 (6)	0.739 (5)	1.187 (12)	0.975 (10)	1.158 (11)	0.712 (3)
66:17:17	0.549 (2.5)	0.549 (2.5)	0.796 (10)	0.580 (5.5)	0.580 (5.5)	0.557 (4)	0.524 (1)	0.581 (7)	0.740 (9)	0.946 (12)	0.887 (11)	0.639 (8)
64:18:18	0.590 (2)	0.591 (4)	0.588 (1)	0.591 (4)	0.591 (4)	0.645 (9)	0.944 (12)	0.608 (5)	0.697 (11)	0.619 (7)	0.642 (8)	0.674 (10)
62:19:19	0.659 (4)	0.660 (5)	0.789 (8)	0.609 (2.5)	0.609 (2.5)	0.772 (7)	0.961 (12)	0.764 (6)	0.811 (9)	0.904 (10)	0.954 (11)	0.607 (1)
60:20:20	0.559 (3.5)	0.559 (3.5)	0.756 (10)	0.823 (11)	0.549 (1)	0.696 (9)	0.893 (12)	0.596 (7)	0.550 (2)	0.635 (8)	0.572 (5)	0.588 (6)
58:21:21	0.524 (1.5)	0.524 (1.5)	0.793 (8)	0.586 (5.5)	0.586 (5.5)	0.581 (4)	0.897 (11)	0.629 (7)	0.861 (9)	1.011 (12)	0.876 (10)	0.560 (3)
56:22:22	0.528 (1)	0.529 (2)	0.855 (10)	0.780 (8)	0.710 (5.5)	0.828 (9)	0.981 (11)	0.573 (4)	0.710 (5.5)	0.736 (7)	1.085 (12)	0.530 (3)
54:23:23	0.701 (8)	0.703 (9)	0.696 (6)	0.598 (2)	0.699 (7)	0.735 (10)	1.000 (12)	0.660 (3)	0.851 (11)	0.664 (4)	0.685 (5)	0.542 (1)
52:24:24	0.571 (5.5)	0.571 (5.5)	0.563 (2)	0.564 (3)	0.570 (4)	0.548 (1)	1.067 (10)	0.742 (7)	1.085 (11)	0.946 (9)	1.161 (12)	0.761 (8)
50:25:25	0.677 (5)	0.678 (6)	0.582 (4)	0.554 (1)	0.566 (2)	0.577 (3)	0.974 (11)	0.690 (8)	0.692 (9)	0.713 (10)	2.145 (12)	0.687 (7)
48:26:26	0.600 (3)	0.601 (4)	0.702 (7)	0.594 (2)	0.585 (1)	0.618 (5)	1.053 (10)	1.043 (9)	1.114 (11)	0.935 (8)	2.962 (12)	0.656 (6)
46:27:27	0.895 (7)	0.899 (8)	0.836 (5)	0.970 (11)	0.837 (6)	0.823 (4)	0.945 (10)	0.911 (9)	1.319 (12)	0.622 (2)	0.682 (3)	0.606 (1)
44:28:28	0.759 (1.5)	0.761 (3)	0.999 (8)	0.825 (4.5)	0.825 (4.5)	0.846 (6)	1.100 (11)	1.055 (10)	0.928 (7)	1.030 (9)	1.297 (12)	0.759 (1.5)
42:29:29	0.835 (6)	0.838 (7)	0.705 (2)	0.770 (4)	0.734 (3)	0.830 (5)	1.112 (11)	0.906 (8)	1.102 (10)	1.143 (12)	0.998 (9)	0.670 (1)
Average rank	3.3	3.825	5.75	5.95	4.5	6.275	9.975	7.175	9.125	7.95	9.25	4.875

Table 6
Correlation performance.

Splitting	lasso	ridge	SVR-l	SVR-r	SVR-s	SVR-p	NN1	NN2	NN3	NN4	NN5	RF
80,10,10	0.585	0.584	0.563	0.497	0.569	0.561	−0.402	0.395	0.528	−0.128	0.479	0.495
78,11,11	0.528	0.528	0.527	0.534	0.535	0.539	−0.135	0.340	0.362	0.350	0.470	0.298
76,12,12	0.642	0.642	0.658	0.695	0.659	0.665	−0.150	0.389	0.507	0.611	0.386	0.491
74,13,13	0.566	0.566	0.561	0.597	−0.424	0.155	−0.199	0.505	0.531	0.535	0.526	0.494
72,14,14	0.575	0.576	0.589	0.588	0.561	0.475	−0.202	0.589	0.431	0.586	0.383	0.478
70,15,15	0.597	0.598	0.588	0.548	0.597	0.570	−0.231	0.559	0.363	0.561	0.336	0.426
68,16,16	0.611	0.611	0.564	0.533	−0.397	0.380	−0.259	0.555	0.384	0.469	0.356	0.467
66,17,17	0.633	0.633	0.557	0.597	0.597	0.601	−0.224	0.614	0.545	0.466	0.499	0.526
64,18,18	0.583	0.583	0.607	0.605	0.605	0.591	−0.251	0.574	0.534	0.568	0.577	0.505
62,19,19	0.554	0.554	0.621	0.617	0.617	0.529	−0.268	0.497	0.483	0.375	0.359	0.592
60,20,20	0.634	0.634	0.584	0.490	0.634	0.575	−0.282	0.602	0.652	0.565	0.618	0.623
58,21,21	0.669	0.668	0.592	0.630	0.630	0.631	−0.278	0.597	0.428	0.394	0.454	0.637
56,22,22	0.661	0.661	0.635	0.634	0.631	0.494	−0.253	0.646	0.629	0.557	0.467	0.648
54,23,23	0.542	0.541	0.559	0.618	0.560	0.568	−0.265	0.578	0.510	0.575	0.560	0.641
52,24,24	0.622	0.622	0.644	0.656	0.668	0.653	−0.276	0.475	0.267	0.347	0.067	0.437
50,25,25	0.520	0.519	0.614	0.641	0.649	0.615	−0.215	0.508	0.505	0.528	0.128	0.501
48,26,26	0.618	0.617	0.569	0.654	0.666	0.655	−0.196	0.338	0.311	0.399	−0.110	0.540
46,27,27	0.413	0.412	0.558	0.563	0.558	0.433	−0.212	0.479	0.257	0.583	0.541	0.560
44,28,28	0.506	0.505	0.544	0.558	0.559	0.460	−0.261	0.415	0.392	0.351	0.228	0.471
42,29,29	0.499	0.498	0.590	0.597	0.603	0.498	−0.267	0.466	0.387	0.370	0.469	0.570

neural networks, and random forests are all more advanced and flexible compared to multiple linear regression methods, we had expected one or more of these algorithms would have the best prediction performance. However, we see lasso and ridge outperform other algorithms; this may be because of the possibility of a linear relationship between the features and the response. In addition, another reason could be the relatively small number of observations, as lasso works well when the number of observations is small compared to the number of features (Fonti and

Table 7
Result of Friedman test at α = 0.05

Friedman chi-squared	Degree of freedom	P-value
83.654	11	2.888e-13

Belitser, 2017).

4.5. Identifying the most important features

After identifying lasso as the best predictive model, we analyzed the importance of features on project performance using the BIC and AIC values. Fig. 3 shows the minimum BIC and minimum AIC values for each number of features. The BIC and AIC values are calculated for each possible model that contains a subset of the ten features in our dataset. In this figure, the red line tracks the best model for a given number of features according to the BIC, and the blue line tracks the best model for a given number of features according to the AIC. As shown in the figure, the minimum values are 1.018 (five features) for AIC and 1.091 (three features) for BIC. We see that the BIC value shows an increase when more than three features are selected, and the AIC value shows an increase when more than five features are selected. In both cases, however, the most important features are related to entrepreneurship orientation and attitude, as shown in Table 9. Table 9 presents the most significant set of features identified by the BIC and the AIC.

As shown in Table 9, both the BIC and the AIC identify entrepreneurship features as the most significant ones. More specifically, all the features related to entrepreneurship attitude (social self-efficacy, appearance self-efficacy, and comparativeness) have been identified as the most significant features. Besides the entrepreneurship attitude features, proactiveness (one of the dimensions of entrepreneurship orientation) has been identified as one of the significant features when predicting project performance. Thus, our results support the findings of previous studies on the significant role of entrepreneurship attitude and orientation on project performance, though with a new approach.

5. Discussion

In this study, we examined the relationship between EO and project performance by using the data collected from students about their performance in their course projects. Our study makes several contributions to the theory and practice in project/operations management, which we discuss below.

5.1. Theoretical contributions

Our first finding is related to the positive impacts of EO and entrepreneurship attitude on project performance. Based on the results of our case study, we showed that EO is important in any type of project, not

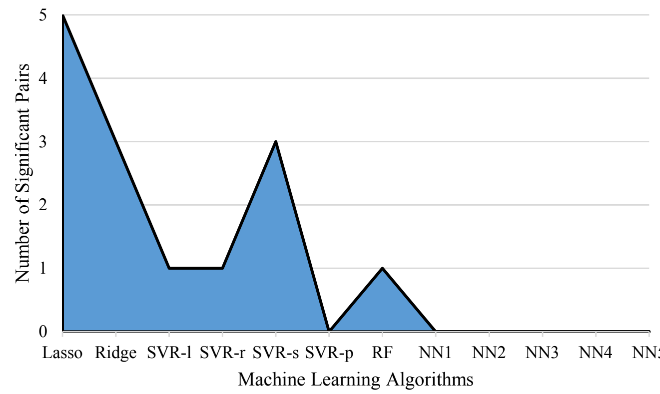

Fig. 2. Number of significant pairs for each machine learning algorithm.

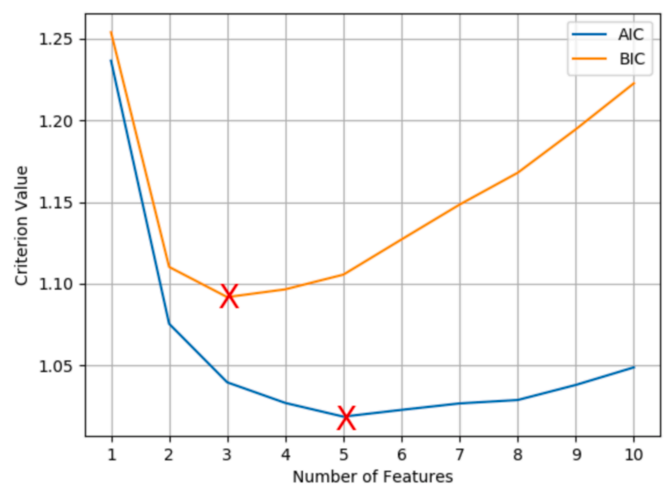

Fig. 3. Minimum values of AIC and BIC associated with different numbers of features.

Table 9
Most important features.

Criterion	Most important features
BIC	Soc, App, Comp
AIC	Pro, Soc, App, Comp, GPA

Table 8
Results of Nemenyi test with α = 0.05

	lasso	ridge	SVR-l	SVR-r	SVR-s	SVR-p	RF	NN1	NN2	NN3	NN4	NN5
lasso		←	←	←	←	←	←	←●	←●	←●	←●	←●
ridge			←	←	←	←	←	←	←	←●	←●	←●
SVR-l				←	↑	←	↑	←●	←	←	←	←
SVR-r					↑	←	↑	←●	←	←	←	←
SVR-s						←	←	←●	←	←●	←	←●
SVR-p							↑	←	←	←	←	←
RF								←●	←	←	←	←
NN1									↑	↑	↑	↑
NN2										←	←	←
NN3											↑	←
NN4												←

Indicates the algorithm that the arrow is pointing to is better, but the difference in their performance is not significant. ←● indicates the algorithm that the arrow is pointing to is significantly better than the other. ↑ indicates the algorithm that the arrow is pointing to is better, but the difference in their performance is not significant.

just in entrepreneurship projects. In our case study, we considered university students who were working on different types of projects; our results showed that the entrepreneurship dimensions we considered have a positive impact on how students perform in their projects.

With respect to the effect of entrepreneurship orientation on project performance, our results show that among the different features we considered (including entrepreneurship, demographic/individual, and contextual factors), the most important set of features when predicting project performance belongs to the entrepreneurship features, according to the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). Our results support previous findings in the literature regarding the important role of entrepreneurship orientation of individuals in project performance; our approach used machine learning, which is a novel approach for studying this relationship.

Our second finding is related to using machine learning principles and the methodological approach to address our research problem. With the use of machine learning, the problem of ambiguity regarding the relationship between influential factors and project performance can be resolved, which makes the use of machine learning less challenging compared to traditional prediction models (Gu et al., 2018). Since the response in the problem we are dealing with is known (project performance), we used some of the popular supervised learning methods (Oztekin et al., 2013) including linear regression models (lasso and ridge), support vector regressor (with different kernels), random forests, and neural networks with different numbers of hidden layers. A summary of the comparison of the algorithms is provided in Table 10.

The results of our analysis showed the best algorithm with respect to MSE is lasso, followed by ridge, SVR with sigmoid kernel, and random forests. One of the reasons that lasso worked better compared to the other algorithms could be due to the number of observations. Lasso is useful when the number of observations is small compared to the number of features (Fonti and Belitser, 2017). Also, since lasso and ridge have better performance, there is a possibility that the relationship between the features and the response is linear. The data shows that neural networks are not an appropriate method for this problem, as the worst average rankings belong to these methods. This may be because of the possibility of a linear relationship between the variables. With respect to correlation results, most of the prediction models (except neural network with one hidden layer) and splitting schemes resulted in a correlation well above the recommended minimum value of 0.3 for human-related studies (Cohen et al., 2014).

Our third finding is regarding the appropriate percentage of training versus validation and testing in conducting machine learning algorithms. As in prediction models, there is not a globally accepted best training size that works for any kind of problem. Cramer et al. (2017) emphasized the need to test the models on variable lengths of training and testing. In order to address this issue, we tested the machine learning methods in 20 different splitting configurations of the sample.

We split our sample into three sets of training, validating, and testing. The results showed that as we decrease the percentage of training and increase the percentage of validating and testing samples, we often get better MSE and correlation values for lasso, ridge, NN2 (neural networks with two hidden layers), and random forests methods. This might be because of the fact that a larger percentage of the training set cause overfitting. As we decrease the length of the training set, the models become fitter.

However, this trend stops when it gets to 56% of the samples for training; this may be because of underfitting created by overly decreasing the size of the training set. In contrast, the other methods display a chaotic response to changes in the percentage of training, validating, and testing. A summary of our findings in this study is presented in Table 11.

Table 10
Summary of the machine learning algorithm comparison.

Algorithm	Comparing factors		
	Prediction accuracy	Sensitivity to training length	Correlation[a]
lasso	Good	Sensitive	above 0.3
ridge	Good	Sensitive	above 0.3
SVR-l	Average	Sensitive	above 0.3
SVR-r	Average	Sensitive	above 0.3
SVR-s	Between Average and Good	Sensitive	Mostly above 0.3
SVR-p	Average	Relatively sensitive	Mostly above 0.3
RF	Between Average and Good	Sensitive	Mostly above 0.3
NN1	Weak	Relatively sensitive	Below 0.3
NN2	Limited	Sensitive	above 0.3
NN3	Weak	Sensitive	Mostly above 0.3
NN4	Limited	Sensitive	Mostly above 0.3
NN5	Weak	Relatively sensitive	Close to 0.3

[a] 0.3 is the recommended minimum value of correlation for human-related studies (Cohen et al., 2014).

5.2. Research implications

Results obtained in this study provide new insight for scholars in the area of project management in understanding the impact of factors related to EO that can also directly influence project performance. This study validates the theoretical models regarding the alignment of the practices of project management to the entrepreneurial orientation that a firm or an organization may pursue for competitive advantage. Moreover, developing ways to measure project performance provides valuable information that covers short-, medium-, and long-term perspectives of enterprises. Project performance information is used not only for control and monitoring but also to analyze operations planning and policies. Information obtained from the performance of projects can be used in different decision areas of operations management such as operations planning and control and human resource management. Thus, our findings in this study can provide new insight to researchers in the field of operations management and supply chain management, in order to have a new perspective on individuals who are involved in activities within the enterprise.

5.3. Practical implications

Our findings provide new insights into the relationship between entrepreneurship skills and project performance. An immediate

Table 11
Summary of the findings.

Subject	Description
The positive impact of entrepreneurship orientation and attitude on project performance	Using the data collected from our case study, we showed that entrepreneurship orientation is important in any type of project, not just in entrepreneurship projects.
Method of study	We used machine learning to address the problem of understanding the entrepreneurial factors in project performance; this is a new methodology in the existing body of knowledge.
length of the training set versus testing and validating sets	We tested the machine learning methods in 20 different training, testing, and validating lengths and showed that using 56% data for training and 22% for testing and validating will give us better prediction performance for the lasso, ridge, NN2, and RF methods. Since other methods show a chaotic response, this conclusion is not valid for the other methods.

outcome of this study is to show that entrepreneurship skills are not limited to the success of entrepreneurship projects; these skills are critical in the success of any project. Given the results obtained in this study, it is critical that project managers approach the problem of project performance or project success with new methods, considering the limitations associated with traditional approaches. Project managers should realize the importance of soft skills such as those examined in this study on project performance. While project managers still need to assess the success of a project with respect to time, cost, and quality, it is important to identify the determinants of project performance from an individual's perspective. In that regard, our study identifies new metrics that impact project performance. Second, turning to supply chain managers and those who are involved in new product development projects with their suppliers or have developed collaborative projects (Petersen et al., 2005; Cheng and Carrillo, 2012), it is also critical to ensure that project members across the supply chain exhibit the entrepreneurial skills that were discussed in this paper. This is one of the ways that operations and supply chain managers can ensure that projects can achieve their desired outcomes.

5.4. Limitations and future directions

As with any study, our results are limited by project context, which in this study were course projects carried out by university students. Future studies using different project contexts are needed to verify and extend our results. In the same vein, the results are limited to a particular population, which in our study was university students doing six different projects. A deeper understanding may be provided by a more diverse population. As the sample size increases, we may get new insights regarding the validity of the results and the performance of machine learning algorithms.

Another limitation is the metric we applied for measuring project performance. In our study, we used a self-assessment of students regarding their performance in course projects, which may be subjective in most cases. Thus, future studies are needed to use other metrics for measuring project performance. For example, one metric could be the performance of students in the course project from the perspective of their instructor.

We used 10 features to develop our predictive models. Additional features related to EO, added singly or in combination, can enhance our understanding of the effect of EO on project performance. Similarly, including other individual and contextual features, such as an individual's age, project time, or project resources may improve the accuracy of the results regarding project performance. Moreover, in this study, we have used only four sets of machine learning algorithms. Future studies are needed to test other algorithms. We also believe that future studies should examine the relationship between entrepreneurship orientation and project performance in organizations. A major challenge in this regard is to identify organizations with multiple types of projects (e.g., product development, process improvement). Finally, future studies can examine other methods to determine feature importance, in order to validate and improve our understanding regarding the importance of the features considered in this study.

Author contribution section

Sima Sabahi: Methodology, Formal Analysis; Data curation; Investigation; Software; Writing- Original draft preparation; Visualization, Mahour Parast: Conceptualization; Writing- Original draft preparation; Supervision; Writing - Review & Editing; Funding Acquisition; Project Administration.

Acknowledgments

This research is based on work supported by the National Science Foundation (NSF) under Grant No. 1533681. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The authors are thankful to Ali Jamali for assisting with the conduct of the project.

References

Agarwal, N., Rathod, U., 2006. Defining 'success' for software projects: an exploratory revelation. Int. J. Proj. Manag. 24 (4), 358–370.
Ahmed, I., Ali, G., Ramzan, M., 2014. Leader and organization: the impetus for individuals' entrepreneurial orientation and project success. J. Global Entrepren. Res. 4 (1), 1–11.
Akaike, H., 1974. A new look at the statistical model identification. In: Selected Papers of Hirotugu Akaike. Springer, New York, pp. 215–222.
Ambad, S.N.A., Damit, D.H.D.A., 2016. Determinants of entrepreneurial intention among undergraduate students in Malaysia. Procedia Econ. Finan. 37, 108–114.
Anderson, B.S., Kreiser, P.M., Kuratko, D.F., Hornsby, J.S., Eshima, Y., 2015. Reconceptualizing entrepreneurial orientation. Strateg. Manag. J. 36 (10), 1579–1596.
APM, 2012. APM Body of Knowledge, sixth ed. Association for Project Management, Princes Risborough.
Atkinson, R., 1999. Project management: cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria. Int. J. Proj. Manag. 17 (6), 337–342.
Atkinson, R., Crawford, L., Ward, S., 2006. Fundamental uncertainties in projects and the scope of project management. Int. J. Proj. Manag. 24 (8), 687–698.
Baccarini, D., 1999. The logical framework method for defining project success. Proj. Manag. J. 30 (4), 25–32.
Baker, B.N., Murphy, D.C., Fisher, D., 1974. Factors affecting project success. In: Project Management Handbook. Van Nostrand Reinhold, New York, pp. 902–919.
Barba-Sánchez, V., Atienza-Sahuquillo, C., 2018. Entrepreneurial intention among engineering students: the role of entrepreneurship education. European Res. Manag. Bus. Econ. 24 (1), 53–61.
Barclay, C., 2008. Towards an integrated measurement of is project performance: the project performance scorecard. Inf. Syst. Front. 10 (3), 331–345.
Barclay, C., Osei-Bryson, K.M., 2010. Project performance development framework: an approach for developing performance criteria & measures for information systems (IS) projects. Int. J. Prod. Econ. 124 (1), 272–292.
Barfod, M.B., 2012. An MCDA approach for the selection of bike projects based on structuring and appraising activities. Eur. J. Oper. Res. 218 (3), 810–818.
Baruch, Y., Holtom, B.C., 2008. Survey response rate levels and trends in organizational research. Hum. Relat. 61 (8), 1139.
Biedenbach, T., Müller, R., 2012. Absorptive, innovative and adaptive capabilities and their impact on project and project portfolio performance. Int. J. Proj. Manag. 30 (5), 621–635.
Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Clarendon Press; Oxford University Press, Oxford.
Bolton, D.L., Lane, M.D., 2012. Individual entrepreneurial orientation: development of a measurement instrument. Educ + Train 54 (2/3), 219–233.
Botchkarev, A., 2018. Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology. Cornell University Library, Ithaca (arXiv.org).
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Brown, I., Mues, C., 2012. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Syst. Appl. 39 (3), 3446–3453.
Bryde, D.J., 2005. Methods for managing different perspectives of project success. Br. J. Manag. 16 (2), 119–131.
Brynjolfsson, E., 1993. The productivity paradox of information technology. Commun. ACM 36 (12), 66–77.
Busby, J.S., Williamson, A., 2000. The appropriate use of performance measurement in non-production activity the case of engineering design. Int. J. Oper. Prod. Manag. 20 (3), 336–358.
Cao, Q., Hoffman, J.J., 2011. A case study approach for developing a project performance evaluation system. Int. J. Proj. Manag. 29 (2), 155–164.
Carter, M.R., Tjernström, E., Toledo, P., 2019. Heterogeneous impact dynamics of a rural business development program in Nicaragua. J. Dev. Econ. 138, 77–98.
Celebi, M.E., Aydin, K., 2016. Unsupervised Learning Algorithms. Springer, Berlin.
Chapelle, O., Schölkopf, B., Zien, A., 2006. Semi-supervised Learning. MIT Press, Cambridge, Mass.
Chen, J., Zhu, Z., Anquan, W., 2005. A system model for corporate entrepreneurship. Int. J. Manpow. 26 (6), 529–543.
Cheng, L.C., Carrillo, E.E., 2012. Assessing supplier performances under partnership in project-type procurement. Ind. Manag. Data Syst. 112 (2), 290–312.
Cheng, M.Y., Wu, Y.W., Wu, C.F., 2010. Project success prediction using an evolutionary support vector machine inference model. Autom. ConStruct. 19 (3), 302–307.
Chulani, S., Boehm, B., Steece, B., 1999. Bayesian analysis of empirical software engineering cost models. IEEE Trans. Softw. Eng. 25 (4), 573–583.
Cohen, D.J., Graham, R.J., 2001. The Project Manager's MBA: How to Translate Project Decisions into Business Success. John Wiley & Sons, New York.
Cohen, P., West, S.G., Aiken, L.S., 2014. Applied Multiple Regression/correlation Analysis for the Behavioral Sciences, second ed. Psychology Press, New York.
Collins, A., Baccarini, D., 2004. Project success—a survey. J. Constr. Res. 5 (02), 211–231.

Cooke-Davis, T., 2002. The real success factors on projects. Int. J. Proj. Manag. 20 (3), 185–190.
Cramer, S., Kampouridis, M., Freitas, A.A., Alexandridis, A.K., 2017. An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. Expert Syst. Appl. 85, 169–181.
Dainty, A., Cheng, M.I., Moore, D., 2003. Redefining performance measures for construction project managers: an empirical evaluation. Constr. Manag. Econ. 21 (2), 209–218.
Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. 7 (Jan), 1–30.
Douglas, E.J., Shepherd, D.A., 2002. Self-employment as a career choice: attitudes, entrepreneurial intentions, and utility maximization. Entrep. Theory Pract. 26 (3), 81–90.
Douzas, G., Bacao, F., 2018. Effective data generation for imbalanced learning using conditional generative adversarial networks. Expert Syst. Appl. 91, 464–471.
Dvir, D., Ben-David, A., Sadeh, A., Shenhar, A.J., 2006. Critical managerial factors affecting defense projects success: a comparison between neural network and regression analysis. Eng. Appl. Artif. Intell. 19 (5), 535–543.
Eilat, H., Golany, B., Shtub, A., 2006. Constructing and evaluating balanced portfolios of R&D projects with interactions: a dea based methodology. Eur. J. Oper. Res. 172 (3), 1018–1039.
Eilat, H., Golany, B., Shtub, A., 2008. R&D project evaluation: an integrated dea and balanced scorecard approach. Omega 36 (5), 895–912.
Erikson, T., 2002. Goal-setting and entrepreneurial self-efficacy. Int. J. Entrep. Innov. 3 (3), 183–189.
Farris, J.A., Groesbeck, R.L., Aken, E.M.V., Letens, G., 2006. Evaluating the relative performance of engineering design projects: a case study using data envelopment analysis. IEEE Trans. Eng. Manag. 53 (3), 471–482.
Fellnhofer, K., Puumalainen, K., Sjögrén, H., 2016. Entrepreneurial orientation and performance–are sexes equal? Int. J. Entrepr. Behav. Res. 22 (3), 346–374.
Filser, M., Eggers, F., 2014. Entrepreneurial orientation and firm performance: a comparative study of Austria, Liechtenstein and Switzerland. S. Afr. J. Bus. Manag. 45 (1), 55–65.
Fonti, V., Belitser, E., 2017. Feature selection using lasso. VU Amst. Res. Paper Bus. Anal. 1–25.
Fortune, J., White, D., 2006. Framing of project critical success factors by a systems model. Int. J. Proj. Manag. 24 (1), 53–65.
Fox, J., 1991. Regression Diagnostics: an Introduction. Sage Publications, Newbury Park, Calif.
Freitas, H., Martens, C.D.P., Boissin, J.P., Behr, A., 2012. Guidance elements for the entrepreneurial orientation of software firms. Rev. Adm. 47 (2), 163–179.
Friedman, M., 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Am. Stat. Assoc. 32 (200), 675–701.
Friedman, M., 1940. A comparison of alternative tests of significance for the problem of m rankings. Ann. Math. Stat. 11 (1), 86–92.
Gasse, Y., 1985. A strategy for the promotion and identification of potential entrepreneurs at the secondary school level. Front. Entrepren. 538–559.
Ge, L., Peng, X.M., 2012. Research fronts of international entrepreneurship education in the visual threshold of knowmetrics. J. Knowl. Bas. Innov. China 4 (1), 55–65.
Gemünden, H.G., Lehner, P., Kock, A., 2018. The project-oriented organization and its contribution to innovation. Int. J. Proj. Manag. 36 (1), 147–160.
Glass, R.L., 2005. IT failure rates - 70% or 10-15%? IEEE Software 22 (3), 110–112.
Goddard, R.D., Hoy, W.K., Hoy, A.W., 2004. Collective efficacy beliefs: theoretical developments, empirical evidence, and future directions. Educ. Res. 33 (3), 3–13.
Gu, S., Kelly, B., Xiu, D., 2018. Empirical Asset Pricing via Machine Learning. National Bureau *of Economic Research*, pp. 1–67.
Gürol, Y., Atsan, N., 2006. Entrepreneurial characteristics amongst university students: some insights for entrepreneurship education and training in Turkey. Educ + Train 48 (1), 25–38.
Harris, M.L., Gibson, S.G., 2008. Examining the entrepreneurial attitudes of US business students. Educ + Train 50 (7), 568–581.
Haykin, S.S., 2009. Neural Networks and Learning Machines, third ed. Prentice Hall, New York.
Henry, C., Hill, F., Leitch, C., 2005. Entrepreneurship education and training: can entrepreneurship be taught? Part I. Educ + Train 47 (2), 98–111.
Huberty, C.J., 1989. Problems with stepwise methods—better alternatives. In: Advances in Social Science Methodology, vol. 1. JAI, Greenwich, CT, pp. 43–70.
Ika, L.A., 2009. Project success as a topic in project management journals. Proj. Manag. J. 40 (4), 6–19.
Ika, L.A., Diallo, A., Thuillier, D., 2012. Critical success factors for world bank projects: an empirical investigation. Int. J. Proj. Manag. 30 (1), 105–116.
Iskender, E., Batı, G.B., 2015. Comparing Turkish universities entrepreneurship and innovativeness index's rankings with sentiment analysis results on social media. Procedia - Soc. Behav. Sci. 195, 1543–1552.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning. Springer, New York.
Jani, A., 2011. Escalation of commitment in troubled it projects: influence of project risk factors and self-efficacy on the perception of risk and the commitment to a failing project. Int. J. Proj. Manag. 29 (7), 934–945.
Jugdev, K., Müller, R., 2005. A retrospective look at our evolving understanding of project success. Proj. Manag. J. 36 (4), 19–31.
Kerzner, H., 2004. Advanced Project Management: Best Practices on Implementation. John Wiley & Sons, Hoboken, N.j.
Kim, S.J., Koh, K., Lustig, M., Boyd, S., Gorinevsky, D., 2007. An interior-point method for large-scale l1 regularized least squares. IEEE J. Select. Top. Sign. Process. 1 (4), 606–617.
Kleinberg, J., Ludwig, J., Mullainathan, S., Obermeyer, Z., 2015. Predicting policy problems. Am. Econ. Rev. 105 (5), 491–495.
Kotsiantis, S.B., 2007. Supervised machine learning: a review of classification techniques. Informatica-Ljubljana 31 (Part 3), 249–268.
Kuratko, D.F., 2005. The emergence of entrepreneurship education: development, trends, and challenges. Entrep. Theory Pract. 29 (5), 577–597.
Küttim, M., Kallaste, M., Venesaar, U., Kiis, A., 2014. Entrepreneurship education at university level and students' entrepreneurial intentions. Procedia-Soc. Behav. Sci. 110, 658–668.
Kuura, A., Blackburn, R.A., Lundin, R.A., 2014. Entrepreneurship and projects—linking segregated communities. Scand. J. Manag. 30 (2), 214–230.
Lechler, T., 1998. When it comes to project management, it's the people that matter: an empirical analysis of project management in Germany. In: The Nature and Role of Projects in the Next 20 Years: Research Issues and Problems. University of Calgary, pp. 205–215.
Li, F.J., 2018. Quality evaluation method of college graduates' innovation and entrepreneurship education based on the principle of brain neurology. Kuram ve Uygulamada Egitim Bilimleri 18 (6), 3114–3124.
Li, Y., Rakesh, V., Reddy, C.K., 2016. Project success prediction in crowdfunding environments. In: Ninth ACM International Conference on Web Search and Data Mining, San Francisco, CA. February 22–25.
Lim, C., Mohamed, M.Z., 1999. Criteria of project success: an exploratory re-examination. Int. J. Proj. Manag. 17 (4), 243–248.
Lin, T.C., Huang, C.C., 2010. Withholding effort in knowledge contribution: the role of social exchange and social cognitive on project teams. Inf. Manag. 47 (3), 188–196.
Linton, J.D., Cook, W.D., 1998. Technology implementation: a comparative study of canadian and U.S. Factories. Infor 36 (3), 142–150.
Low, M.B., MacMillan, I.C., 1988. Entrepreneurship: past research and future challenges. J. Manag. 14 (2), 139–161.
Lumpkin, G.T., Cogliser, C.C., Schneider, D.R., 2009. Understanding and measuring autonomy: an entrepreneurial orientation perspective. Entrep. Theory Pract. 33 (1), 47–69.
Lumpkin, G.T., Dess, G.G., 1996. Clarifying the entrepreneurial orientation construct and linking it to performance. Acad. Manag. Rev. 21 (1), 135–172.
Lumpkin, G.T., Dess, G.G., 2001. Linking two dimensions of entrepreneurial orientation to firm performance: the moderating role of environment and industry life cycle. J. Bus. Ventur. 16 (5), 429–451.
Lundin, R.A., Arvidsson, N., Brady, T., Ekstedt, E., Midler, C., 2015. Managing and Working in Project Society. Cambridge University Press, Cambridge.
Lyon, D.W., Lumpkin, G.T., Dess, G.G., 2000. Enhancing entrepreneurial orientation research: operationalizing and measuring a key strategic decision making process. J. Manag. 26 (5), 1055–1085.
Makridakis, S., Wheelwright, S.C., Hyndman, R.J., 2008. Forecasting Methods and Applications. John wiley & sons, New York.
Malhotra, R., 2016. An empirical framework for defect prediction using machine learning techniques with android software. Appl. Soft Comput. 49, 1034–1050.
Marijana, Z.S.a., Sanja, P., Ivana, D., 2010. Classification of entrepreneurial intentions by neural networks, decision trees and support vector machines. Croat. Oper. Res. Rev. 1 (1), 62–71.
Marques, G., Gourc, D., Lauras, M., 2011. Multi-criteria performance analysis for decision making in project management. Int. J. Proj. Manag. 29 (8), 1057–1069.
Martens, C.D.P., Lacerda, F.M., Belfort, A.C., Freitas, H.M. R.d., 2016. Research on entrepreneurial orientation: current status and future agenda. Int. J. Entrepr. Behav. Res. 22 (4), 556–583.
Martens, C.D.P., Machado, F.J., Martens, M.L., Freitas, H.M.d.R., 2018. Linking entrepreneurial orientation to project success. Int. J. Proj. Manag. 36 (2), 255–266.
Martes, C.D., Carneiro, K., Martens, M., Silva, D., 2015. Relationship between entrepreneurial orientation and project management maturity in brazilian software firms. Rev. Ibero-Am. Estratég. 14 (2), 72–91.
Martínez, D.M., Fernández-Rodríguez, J.C., 2015. Artificial intelligence applied to project success: a literature review. Int. J. Artif. Intell. Interact. Multimedia 3 (5), 77–84.
Mazur, A., Pisarski, A., Chang, A., Ashkanasy, N.M., 2014. Rating defence major project success: the role of personal attributes and stakeholder relationships. Int. J. Proj. Manag. 32 (6), 944–957.
Melville, N., Kraemer, K., Gurbaxani, V., 2004. Information technology and organizational performance: an integrative model of IT business value. MIS Q. 28 (2), 283–322.
Miller, D., 1983. The correlates of entrepreneurship in three types of firms. Manag. Sci. 29 (7), 770–791.
Minbashian, A., Bright, J.E.H., Bird, K.D., 2010. A comparison of artificial neural networks and multiple regression in the context of research on personality and work performance. Organ. Res. Methods 13 (3), 540–561.
Mitchell, T.M., 1997. Machine Learning. McGraw-Hill, New York.
Moreno, A.M., Casillas, J.C., 2008. Entrepreneurial orientation and growth of smes: a causal model. Entrep. Theory Pract. 32 (3), 507–528.
Nasution, M.D.T.P., Siahaan, A.P.U., Rossanty, Y., Aryza, S., 2018. Entrepreneurship intention prediction using decision tree and support vector machine. In: Paper Presented at the International Conference on Advance & Scientific Innovation, Medan, Indonesia.
Nemenyi, P., 1963. Distribution-free Multiple Comparisons. Doctoral Dissertation. Princeton University.
North, D.C., 1990. Institutions, Institutional Change and Economic Performance. Cambridge University Press, Cambridge.
Oztekin, A., Delen, D., Turkyilmaz, A., Zaim, S., 2013. A machine learning-based usability evaluation method for elearning systems. Decis. Support Syst. 56, 63–73.

Paliwal, M., Kumar, U.A., 2009. Neural networks and statistical techniques: a review of applications. Expert Syst. Appl. 36 (1), 2–17.
Palocsay, S.W., White, M.M., 2004. Neural network modeling in cross-cultural research: a comparison with multiple regression. Organ. Res. Methods 7 (4), 389–399.
Petersen, K.J., Handfield, R.B., Ragatz, G.L., 2005. Supplier integration into new product development: coordinating product, process and supply chain design. J. Oper. Manag. 23 (3,4), 371–388.
Pillai, A.S., Joshi, A., Rao, K.S., 2002. Performance measurement of R&D projects in a multi-project, concurrent engineering environment. Int. J. Proj. Manag. 20 (2), 165–177.
Pinto, J.K., 1990. Project implementation profile: a tool to aid project tracking and control. Int. J. Proj. Manag. 8 (3), 173–182.
Pinto, J.K., Slevin, D.P., 1987. Critical factors in successful project implementation. IEEE Trans. Eng. Manag. 34 (1), 22–27.
Pinto, J.K., Slevin, D.P., 1988. Critical success factors across the project life cycle. Proj. Manag. J. 19 (3), 67–74.
Pirola-Merlo, A., 2010. Agile innovation: the role of team climate in rapid research and development. J. Occup. Organ. Psychol. 83 (4), 1075–1084.
PMI, 2017. A Guide to the Project Management Body of Knowledge (PMBOK Guide), sixth ed. Project Management Institute, Newtown Square, PA.
Portugal, I., Alencar, P., Cowan, D., 2018. The use of machine learning algorithms in recommender systems: a systematic review. Expert Syst. Appl. 97, 205–227.
Putka, D.J., Beatty, A.S., Reeder, M.C., 2018. Modern prediction methods: new perspectives on a common problem. Organ. Res. Methods 21 (3), 689–732.
Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1 (1), 81–106.
Rauch, A., Wiklund, J., Lumpkin, G.T., Frese, M., 2009. Entrepreneurial orientation and business performance: an assessment of past research and suggestions for the future. Entrep. Theory Pract. 33 (3), 761–787.
Revilla, E., Modrego, A., Sarkis, J., 2003. Evaluating performance of public-private research collaborations: a DEA analysis. J. Oper. Res. Soc. 54 (2), 165–174.
Rezvani, A., Chang, A., Wiewiora, A., Ashkanasy, N.M., Jordan, P.J., Zolin, R., 2016. Manager emotional intelligence and project success: the mediating role of job satisfaction and trust. Int. J. Proj. Manag. 34 (7), 1112.
Robinson, P.B., Huefner, J.C., Hunt, H.K., 1991. Entrepreneurial research on student subjects does not generalize to real world entrepreneurs. J. Small Bus. Manag. 29 (2), 42.
Ruiz, E., Nieto, F.H., 2000. A note on linear combination of predictors. Stat. Probab. Lett. 47 (4), 351–356.
Runyan, R., Droge, C., Swinney, J., 2008. Entrepreneurial orientation versus small business orientation: what are their relationships to firm performance? J. Small Bus. Manag. 46 (4), 567–588.
Salari, M., Khamooshi, H., 2016. A better project performance prediction model using fuzzy time series and data envelopment analysis. J. Oper. Res. Soc. 67 (10), 1274–1287.
Scarborough, D., Somers, M.J., 2006. Neural Networks in Organizational Research: Applying Pattern Recognition to the Analysis of Organizational Behavior, first ed. American Psychological Association, Washington, DC.
Scherer, R.F., Adams, J.S., Carley, S.S., Wiebe, F.A., 1989. Role model performance effects on development of entrepreneurial career preference. Entrep. Theory Pract. 13 (3), 53–72.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6 (2), 461–464.
Schwer, R.K., Yucelt, U., 1984. A study of risk-taking propensities among small business entrepreneurs and managers: an empirical evaluation. Am. J. Small Bus. 8 (3), 31–40.
Setiawan, H., Erdogan, B., Ogunlana, S.O., 2015. Proactiveness of contractors: a study of Indonesia. Procedia Eng. 125, 60–67.
Sexton, D.L., Bowman, N., 1983. Determining entrepreneurial potential of students. Acad. Manag. Proc. 1983 (1), 408–412.
Shenhar, A.J., Levy, O., Dvir, D., 1997. Mapping the dimensions of project success. Proj. Manag. J. 28 (2), 5–13.
Shmueli, G., Koppius, O.R., 2011. Predictive analytics in information systems research. MIS Q. 35 (3), 553–572.
Slevin, D.P., Pinto, J.K., 1986. The project implementation profile: new tool for project managers. Proj. Manag. J. 17 (4), 57–70.
Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14 (3), 199–222.
Somers, M.J., 2001. Thinking differently: assessing nonlinearities in the relationship between work attitudes and job performance using a bayesian neural network. J. Occup. Organ. Psychol. 74 (1), 47–61.
Somers, M.J., Casal, J.C., 2009. Using artificial neural networks to model nonlinearity: the case of the job satisfaction - job performance relationship. Organ. Res. Methods 12 (3), 403.
Souitaris, V., Zerbinati, S., Al-Laham, A., 2007. Do entrepreneurship programmes raise entrepreneurial intention of science and engineering students? The effect of learning, inspiration and resources. J. Bus. Ventur. 22 (4), 566–591.
Stamboulis, Y., Barlas, A., 2014. Entrepreneurship education impact on student attitudes. Int. J. Manag. Educ. 12 (3), 365–373.
Stensrud, E., Myrtveit, I., 2003. Identifying high performance ERP projects. IEEE Trans. Softw. Eng. 29 (5), 398–416.
Sutton, R.S., Barto, A.G., 2018. Reinforcement Learning: an Introduction. MIT press, Cambridge.
Thapa, A., 2015. Determinants of microenterprise performance in Nepal. Small Bus. Econ. 45 (3), 581–594.
Thomas, A.S., Mueller, S., 1998. Are entrepreneurs the same across cultures. In: USASBE Annual National Conference. January 5-18.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B 58 (1), 267–288.
Todorovic, M.L., Petrovic, D.C., Mihic, M.M., Obradovic, V.L., Bushuyev, S.D., 2015. Project success analysis framework: a knowledge-based approach in project management. Int. J. Proj. Manag. 33 (4), 772–783.
Toor, S.U.R., Ogunlana, S.O., 2010. Beyond the 'iron triangle': stakeholder perception of key performance indicators (KPIs) for large-scale public sector development projects. Int. J. Proj. Manag. 28 (3), 228–236.
Trawiński, B., Smętek, M., Telec, Z., Lasota, T., 2012. Nonparametric statistical analysis for multiple comparison of machine learning regression algorithms. Int. J. Appl. Math. Comput. Sci. 22 (4), 867–881.
Tu, J., Lin, A., Chen, H., Li, Y., Li, C., Cuevas, E., 2019. Predict the entrepreneurial intention of fresh graduate students based on an adaptive support vector machine framework. Math. Probl. Eng. 2019, 1–16.
Vapnik, V., 1982. Estimation of Dependences Based on Empirical Data Berlin. Springer, Berlin.
Vapnik, V., 1995. The Nature of Statistical Learning Theory, second ed. Springer, New York.
Vapnik, V., Lerner, A., 1963. Pattern recognition using generalized portrait method. Autom. Rem. Contr. 24, 774–780.
Vapnik, V., 1998. Statistical Learning Theory. Wiley, New York.
Veciana, J.M., Aponte, M., Urbano, D., 2005. University students' attitudes towards entrepreneurship: a two countries comparison. Int. Entrep. Manag. J. 1 (2), 165–182.
Verma, D., Sinha, K.K., 2002. Toward a theory of project interdependencies in high tech R&D environments. J. Oper. Manag. 20 (5), 451–468.
Vezzoni, G., Pacagnella, A.C., Banzi, A.L., Silva, S.L.D., 2013. Identification and analysis of critical success factors in projects. Revista de Gestao e Projectos 4 (1), 116–137.
Vink, J.P., Haan, G.D., 2015. Comparison of machine learning techniques for target detection. Artif. Intell. Rev.: Int. Sci. Eng. J. 43 (1), 125–139.
Vitner, G., Rozenes, S., Spraggett, S., 2006. Using data envelope analysis to compare project efficiency in a multi-project environment. Int. J. Proj. Manag. 24 (4), 323–329.
Vogelsang, L., 2015. Individual Entrepreneurial Orientation: an Assessment of Students. Doctoral dissertation. Humboldt State University.
Wang, C.L., 2008. Entrepreneurial orientation, learning orientation, and firm performance. Entrep. Theory Pract. 32 (4), 635–657.
Wang, G., Hao, J., Ma, J., Jiang, H., 2011. A comparative assessment of ensemble learning for credit scoring. Expert Syst. Appl. 38 (1), 223–230.
Wang, Y.R., Gibson Jr., G.E., 2010. A study of preproject planning and project success using anns and regression models. Autom. ConStruct. 19 (3), 341–346.
Wateridge, J., 1998. How can IS/IT projects be measured for success? Int. J. Proj. Manag. 16 (1), 59–63.
Williams, D.A., 2016. Can neural networks predict business failure? Evidence from small high tech firms in the UK. J. Dev. Enterpren. 21 (01), 1–17.
Wit, A.D., 1988. Measurement of project success. Int. J. Proj. Manag. 6 (3), 164–170.
Wong, W.K., Cheung, H.M., Venuvinod, P.K., 2005. Individual entrepreneurial characteristics and entrepreneurial success potential. Int. J. Innov. Technol. Manag. 2 (03), 277–292.
Xu, B., Yang, J., Sun, B., 2018. A nonparametric decision approach for entrepreneurship. Int. Entrep. Manag. J. 14 (1), 5–14.
Yao, X., Wu, X., Long, D., 2016. University students' entrepreneurial tendency in China effect of students' perceived entrepreneurial environment. J. Entrepren. Emerg. Econ. 8 (1), 60–81.
Yurtkoru, E.S., Kuşcu, Z.K., Doğanay, A., 2014. Exploring the antecedents of entrepreneurial intention on Turkish university students. Procedia-Soc. Behav. Sci. 150, 841–850.
Zhang, D., Tsai, J.J.P., 2007. Advances in Machine Learning Applications in Software Engineering. Idea Group Pub, Hershey, PA.
Zhang, L., Suganthan, P.N., 2016. A comprehensive evaluation of random vector functional link networks. Inf. Sci. 367–368, 1094–1105.