**The Determinants of Entrepreneurial Intentions and Activity: Opportunities and Challenges from the Application of Machine Learning**

***Byron Graham* a[[1]](#footnote-1)\* *and Karen Bonner* b**

**a *Queens Management School, Queens University Belfast, BT9 5EE***

**b *Economic Policy Centre, Ulster University, Jordanstown, BT37 0QB***

Keywords: Entrepreneurship; Machine Learning ; Dominance Analysis

***Abstract***

Advances in analytics have created opportunities for entrepreneurship researchers to apply machine learning techniques to address entrepreneurship questions, contributing to entrepreneurship theory and practice. Although these opportunities have been recognised in the business and entrepreneurship literature, challenges remain which limit their adoption in entrepreneurship research. This study aims to elaborate on these challenges, and to illustrate some of the opportunities available from the application of machine learning techniques to entrepreneurship research questions. Drawing on data from the Global Entrepreneurship Monitor (GEM), this study adopts a machine learning methodology to examine the relative importance of the determinants of entrepreneurial intentions (EI) and total early stage entrepreneurship (TEA), and the ability to predict these outcomes. The machine learning approaches are compared with a more traditional regression model. The results show that the more advanced algorithm has higher predictive accuracy, and also provides variable importance measures, which enable us to examine the dominance of determinants. Across all models, self-perceptions, experience and age are found to be relatively more important determinants, with cultural factors and other demographics less important. Overall, TEA can be modelled more accurately than EI, but it remains challenging to accurately predict both EI and TEA. The results contribute to our understanding of the determinants of EI and TEA, as well as highlighting the application of the machine learning methodology.

***1.0 Introduction***

Entrepreneurship has been found to play an important role in economic growth (Stel, Carree, & Thurik, 2005), innovation, and social development. A substantial body of literature has focused on the determinants of entrepreneurial activity (see Audretsch and Erdem, 2005; Walter and Heinrichs, 2015 for reviews) although there is still debate over which set of factors has greater explanatory power (Amini Sedeh, Beck & Bajestani, 2020). Models focusing on the determinants of entrepreneurial activity have also had limited predictive accuracy, leading Kruger to propose focusing on entrepreneurial intentions (EI) (Krueger, Reilly, & Carsrud, 2000). Since the work of Shapero and Sokol (1982), EI have become an important framework to understand entrepreneurship, with entrepreneurial intentions preceding activity. A large body of research has contributed to the understanding of EI, and the role of individual predictors such as demographics and psychological factors (Arenius & Minniti, 2005; Beynon, Jones, & Pickernell, 2018; Koellinger, Minniti, & Schade, 2013) as well as contextual and institutional predictors (Engle, Schlaegel, & Dimitriadi, 2011; Fayolle & Liñán, 2014; Liñán, Urbano, & Guerrero, 2011; Moriano, Gorgievski, Laguna, Stephan, & Zarafshani, 2012).

A large portion of the understanding of EI has been underpinned by core theoretical frameworks, including Bird's (1988) model of entrepreneurial ideas, Shapero and Sokol’s (1982) model of the entrepreneurial event and Ajzen's (1991) theory of planned behaviour (Fayolle & Liñán, 2014; Kautonen, van Gelderen, & Fink, 2015; Krueger & Carsrud, 1993; Krueger et al., 2000; Liñán & Fayolle, 2015). Although this work provides a theoretical and empirical foundation for the study of EI, the literature on EI still lacks systemization (Liñán & Fayolle, 2015), and debate remains about the most important determinants. Competing theoretical perspectives leave questions as to which factors are most important in predicting EI.

Previous studies have often found conflicting results, particularly around the socio-cultural impacts (Ahmad, Xavier & Bakar, 2013; Liñán & Fayolle, 2015; Liñán, Urbano & Guerrero, 2011) and demographic variables such as age (Tsai et al, 2016) and education levels (van der Sluis, van Praag & Vijverberg, 2005.). These studies also often have low predictive ability. In discussing the determinants of entrepreneurship, Arin, Huang, Minniti, Nandialath, & Reich, (2015) attribute conflicting findings to model uncertainty, resulting from the choices made by researchers around model specifications. Model uncertainty is likely to also be a particular issue in the intentions literature due to the aforementioned competing frameworks, resulting in a wide range of model specifications. We aim to contribute to these debates by applying a machine learning approach to understand the determinants of entrepreneurial intentions and activity. The machine learning algorithms adopted in this study allow for the inclusion and evaluation of a large number of variables in the model building process, through the use of an automatic feature selection process. This reduces the model uncertainty introduced by the researchers’ specification of the model. This approach also generates variable importance measures, which allows us to identify the most dominant variables in predicting EI and TEA.

The machine learning approach has been facilitated by advances in computer science, which have created opportunities for entrepreneurship researchers to draw on new data and techniques to answer new and existing research questions. Our study is set within the emerging ‘analytics paradigm’ (Delen & Zolbanin, 2018). In this paradigm, new types of data and methods, such as machine learning approaches, can be used to answer interesting research questions and provide new insights. In the entrepreneurship literature, Obschonka & Audretsch, (2020) refer to this as a ‘new era’ of entrepreneurship research, suggesting that AI and big data provide opportunities for the potential to progress the entrepreneurship field, albeit with some challenges.

Although these opportunities have been well discussed in recent literature (Obschonka & Audretsch, 2020; Schwab & Zhang, 2019), they have only been implemented to a limited extent in entrepreneurship research (Lévesque, Obschonka, & Nambisan, 2020; Obschonka & Audretsch, 2020), and have tended to focus on a narrow set of problems. One example where substantial recent progress has been made is in the emerging body of entrepreneurship literature which draws on textual data and text analytics approaches (Singh, Tanev & Bailetti, 2020; Schuelke-Leech and Barry, 2016). In the wider business literature, machine learning approaches have been applied in some areas, for example to problems in marketing (Coussement & De Bock, 2013; Duchessi & Lauría, 2013), accounting (Amani & Fadlalla, 2017), banking (Dincer, Hacioglu, Tatoglu, & Delen, 2019), and human resource management (Huselid, 2018). Obschonka & Audretsch, (2020) suggest that one reason for the limited adoption of analytics in entrepreneurship research is due to entrepreneurship researchers’ lack of familiarity with techniques in the analytics paradigm. There are also cultural differences between the more traditional statistical approach and the more algorithmic approach illustrated in this study (Breiman, 2001). In this paper, we therefore aim to demonstrate some of the common types of machine learning techniques available to entrepreneurship researchers, the questions that these techniques can be used to address, as well as elaborating on the challenges that entrepreneurship researchers might face in implementing these techniques.

To highlight the potential of machine learning techniques in entrepreneurship research we apply common machine learning approaches to identify the most important factors in predicting entrepreneurial intentions and early-stage entrepreneurship. Our central question is thus: Which factors are most important in predicting entrepreneurial intentions and early-stage entrepreneurship? The machine learning approach is implemented using three common algorithms: recursive partitioning, logistic regression, and gradient boosting. These approaches are compared with a traditional logistic regression model, which is often used to study the drivers of entrepreneurship (Coduras, Clemente, & Ruiz, 2016). This allows us to highlight differences between these approaches and algorithms, the challenges faced, and their potential utility in addressing entrepreneurship research questions. Although we apply different algorithms, as with the approach taken by Arin et al. (2015)**,** our study also provides measures of variable importance. This allows us to contribute to the literature on dominance analysis (Budescu, 1993), by identifying the relative importance of the determinants. Our empirical study, in which we implement these algorithms, draws on data from the 2017 GEM. Although the GEM framework has been studied extensively, few studies have applied machine learning techniques.

The results of our empirical study show that entrepreneurship is difficult to predict, in terms of both entrepreneurial intention and activity. Relationships are non-linear, as identified in the wider literature (Beynon, Jones, & Pickernell, 2016, 2018, 2019; Coduras, Clemente, & Ruiz, 2016). The most important factors determining entrepreneurial activity are perceptual, whereby self-efficacy and knowing an entrepreneur are consistently the most important predictors. Similarly, we find that self-efficacy, measured as the individuals’ perceptions of having the skills to start a business, is an important predictor of entrepreneurial intentions, alongside age, and knowing an entrepreneur.

Our study makes several theoretical, methodological, and practical contributions to the entrepreneurship literature. We highlight an opportunity to apply machine-learning approaches to gain insight about entrepreneurship questions, as well as some of the challenges with this approach. In particular, our approach allows us to build more accurate models than the traditional approach, as well as identifying the relative importance of predictors. The visualisation of the decision tree models also allows us to identify more complex relationships in the data.

In addition to highlighting the predictive accuracy and non-linear relationships in the data, the machine learning techniques produce measures of variable importance. This allows us to carry out a dominance analysis, by identifying the most important determinants of entrepreneurial intentions. Although dominance analysis has been carried out in the entrepreneurship research (Arin et al., 2015)**,** to our knowledge this is the first study that has focused on a dominance analysis of the determinants of EI. By carrying out this analysis we also contribute to the wider methodological literature on dominance analysis (Budescu, 1993)**.**

However, challenges remain from both a theoretical and a practical perspective. Most machine learning algorithms have the aim of maximising the predictive accuracy of the model, rather than generating insight in social science research. Moreover, some machine learning models are less interpretable than the traditional approach. Recent advances in explainable AI have helped to open the black box, creating more potential to use these methods in entrepreneurship research. There are also practical challenges, particularly around the skills needed to apply machine learning techniques. We discuss these challenges in this paper, with the aim of demonstrating some of the key considerations when addressing questions within the analytics paradigm.

Our study also has practical and policy implications. We find that it is challenging to accurately predict entrepreneurship, even when using advanced machine learning techniques. However, the GBM is more accurate than the traditional statistical approach. This has practical implications for policy makers, as more accurate models could help to make better decisions. From a policy perspective the models developed in this paper could help policy makers to better understand the people most likely to start a business, as well as the most important predictors. This information could enable more effective targeting of resources and improved focus of policy interventions to increase entrepreneurship.

The paper proceeds as follows. Section two discusses the analytics paradigm in entrepreneurship research. We then discuss the setting for the empirical research, briefly summarising past research using the GEM data. This is followed by the methodology, results, and conclusions.

***2.0 Adopting the Analytics Paradigm in Entrepreneurship Research***

Before proceeding to the empirical study, we discuss the methodological underpinnings of the approach used to address our research question. The development of a wide variety of machine learning algorithms, largely in the computer science field, have provided new tools on for addressing important business and entrepreneurship research questions. In the wider business literature, Delen & Zolbanin, (2018), term this ‘the analytics paradigm’, which differs from the traditional scientific method in important ways. In the traditional scientific method, hypotheses are specified in advance before being formally tested using data, whereas in the analytics paradigm, learning algorithms are used to identify relationships in the data. The metrics used to evaluate models are also different, with analytics methods focusing on measures of predictive accuracy on unseen test data, and traditional approaches focusing on p-values and R squared (Delen & Zolbanin, 2018). Although there are substantial differences between the traditional scientific method and the analytics paradigm, Delen & Zolbanin, (2018) view the analytics paradigm in business research as complementary to the traditional scientific method.

The methods that are used in the analytics paradigm draw on algorithms and approaches which are more commonly developed and applied in computer science. These methods are often unfamiliar to business researchers, which could be one reason for the low levels of usage in both the business and entrepreneurship literature (Delen & Zolbanin, 2018; Obschonka & Audretsch, 2020). Machine learning algorithms are also typically developed to maximise predictive accuracy rather than the testing of causal relationships, which is the usual focus of quantitative studies. Obschonka & Audretsch, (2020) posit that there are opportunities to apply analytics to entrepreneurship problems to gain insights through the use of ‘new research methods, datasets, and study designs’ (pg 531). These techniques can also be used to build and test theory, as well as evaluating predictability (Lévesque et al., 2020; Shmueli & Koppius, 2011).

In this paper, we draw on one type of analytics technique: machine learning. Machine learning involves the application of an algorithm to learn relationships between variables in a dataset. There are two broad categories of machine learning: supervised learning, and unsupervised learning. Supervised learning involves using an algorithm to learn the relationships between input features (variables) and a target. There are many learning algorithms that can be used, such as regression, support vector machines, decision trees, random forests, gradient boosting, and artificial neural networks. For example, previous studies have applied machine learning techniques to predict employee churn, by learning the relationships between employee characteristics and churn (Saradhi & Palshikar, 2011). In contrast to supervised learning, unsupervised learning involves the application of an algorithm to learn the patterns in the data. The key difference between supervised and unsupervised learning is that there is no target variable, rather the aim is to identify patterns of relationships in the data. Common algorithms include k-means and hierarchical clustering. This study specifically draws on supervised learning techniques to predict EI and TEA.

In addition to the different methods used in the analytics paradigm, there is also the opportunity to drawn on ‘big data’; that is larger datasets, and datasets with more heterogeneous types of data. As discussed in the following sections, studies in the entrepreneurship field have begun to draw on more heterogeneous data types, such as textual data from social media (Singh, Tanev & Bailetti, 2020; Schuelke-Leech and Barry, 2016). This big data can require different storage, processing, and analysis techniques. (Braun, Kuljanin, & DeShon, 2018). However, the supervised and unsupervised learning algorithms discussed previously can also be applied to gain insights from small datasets. For example, in the marketing field, Duchessi & Lauría, (2013) apply decision trees to a sample of 66 observations to profile marketing activities.

*2.1 Challenges in Carrying out Entrepreneurship Research within the Analytics Paradigm*

Machine learning approaches can be used to describe the characteristics and relationships in a dataset; to make predictions about future events; and to make statistical inferences. Addressing entrepreneurship research questions within each of these areas comes with its own set of challenges. The different methods also suit different types of research questions. Many of the challenges with the application of machine learning approaches arise in part because the key algorithms were developed within computer science with the main aim of prediction, rather than interpretability or inference (Efron, 2020). The difficulty in making causal inferences from machine learning models has been recognised in the wider literature (Hünermund, Kaminski, & Schmitt, 2021), and in the entrepreneurship field, with Obschonka & Audretsch (2020) identifying ‘issues of prediction vs. explanation; inductive, data driven approaches vs. deduction, theory-driven approaches’ as one of the conceptual research priorities around the use of analytics in entrepreneurship (pg 3).

This highlights a key underpinning difference between the machine learning approach and the traditional statistical approach, where machine learning approaches tend to focus on maximising predictive accuracy rather than on causal inference (Athey & Imbens, 2019; Efron, 2020). Machine learning approaches usually therefore do not adopt some of the same formal approaches that would be found in traditional econometrics techniques, such as the use of confidence intervals (Athey & Imbens, 2019). Factors such as selection bias and confounding effects of unobserved variables can make it difficult to establish causality in a data driven approach (Bareinboim & Tian, 2015). However, in some domains, such as healthcare, machine learning approaches have been found to have more accurate predictive ability than traditional statistical techniques (Graham, Bond, Quinn, & Mulvenna, 2018) and can help to identify non-linear and other complex relationships in the data (Athey & Imbens, 2019). These benefits could outweigh the disadvantages, depending on the goals of the analysis (Athey & Imbens, 2019).

One of the criticisms of machine learning is that it has operated using so called ‘black box’ algorithms, which make it difficult to explain how the algorithm arrived at a particular prediction or decision (Hair & Sarstedt, 2021). In response to these challenges, a growing body of literature has focused on the use of machine learning techniques for causality and inference, and in ‘Explainable AI’ (Adadi & Berrada, 2018; Doran, Schulz, & Besold, 2018; Guo, Cheng, Li, Hahn, & Liu, 2020). From a scientific perspective, the aim of these approaches is to understand what the model has learned, which could lead to new scientific discoveries (Dybowski, 2020). The key question is to interpret what the model has learnt about the relationships between the input variables and the target; and if possible, to understand what causes the target. However, even though it is often possible to explain how a model arrived at a particular prediction, this does not necessarily go as far as identifying the cause of an outcome. Recent research in machine learning algorithms have started to develop techniques for inference based on machine learning algorithms. This opens up the possibility of using machine learning models for causation as well as prediction. For example, the random forest algorithm has been extended to allow for the testing of treatment effects in the generalized random forest algorithm (Athey, Tibshirani, & Wager, 2019).

Another advantage of the machine learning approach is that some machine learning techniques, such as tree-based methods, ignore irrelevant variables when building the model, which acts as a method of feature selection (Kuhn & Johnson, 2013). This could help to avoid model uncertainty. Model uncertainty can result from the specification of the model, for example, through omitting a variable (Chatfield, 1995). Use of inbuilt feature selection reduces the opportunity for the researcher to omit a variable that could have predictive value. An example, of a study seeking to reduce model uncertainty in the entrepreneurship field is carried out by Arin et al. (2015), who apply Bayesian model averaging to study the determinants of entrepreneurship.

The differences and tensions between the machine learning approach and the traditional approach have been recognised in both the business (Delen & Zolbanin, 2018) and entrepreneurship domains (Lévesque et al., 2020). There are conflicts between traditional theory driven approach and the analytics paradigm in entrepreneurship research that need to be addressed (Lévesque et al., 2020). Addressing these factors has been termed a ‘grand challenge’ in entrepreneurship research (Lévesque et al., 2020 pg. 2). Lévesque et al. (2020) also point to the concerns around using AI approaches in analytics, which need to be carefully considered when planning and carrying out research in the analytics paradigm. The first is ensuring the research makes a contribution to theory. The availability of existing data and the potential to easily implement advanced analytics techniques, could result in descriptive research without a clear contribution to theory. It is therefore crucial that research within the analytics paradigm adopts a high level of rigour (Lévesque et al., 2020).

Entrepreneurship researchers may also face practical challenges in the implementation of machine learning techniques. Although some of the machine learning approaches can be implemented using graphical user interface tools, many algorithms require the use of coding languages, the two most popular of which are currently R and Python. Although becoming more popular in undergraduate and postgraduate programmes in the social sciences, these are areas more common to computer science. This means that many entrepreneurship researchers will be unfamiliar with the coding skills and algorithms needed to implement many of the techniques utilised in the analytics paradigm. Thanks to platforms offering Massive Online Open Courses, there are many excellent free courses that provide enough information to get started with the techniques applied in this paper.

The second practical consideration centres on the storage and processing of ‘big data’ (Braun et al., 2018). The data used in the present study is a large dataset, but it is well structured, and is not so large that it cannot be stored and processed on a single machine. In contrast, ‘big data’ can be unstructured (such as textual data) or can be too large to store and process on a single machine. Overcoming these challenges often requires specialised data engineering skills, and potentially collaborations with computer scientists. However, the coding languages previously mentioned can also be used to carry out analytics on ‘big data’.

*2.2 Studies Applying Analytics Approaches in Entrepreneurship Research*

Despite the challenges in applying analytics in entrepreneurship research, a growing body of literature has emerged. Previous studies adopting analytics approaches in entrepreneurship have focused on two broad areas. The first are studies that have drawn on novel sources of data, or ‘big data’. These studies have utilised data from sources such as social media and crowdfunding platforms to address entrepreneurship questions. The second group of studies have drawn on more traditional data sources, such as surveys, but have applied novel techniques such as machine learning, to analyse the data.

Much of the recent applications of analytics in entrepreneurship research have drawn on textual data to gain new insights.Several recent studies have drawn on social media data from sources such as twitter to address entrepreneurship research questions (Olanrewaju, Hossain, Whiteside, & Mercieca, 2020). These studies have focused on diverse research questions in entrepreneurship, such as examining personality characteristics (Obschonka, Fisch, & Boyd, 2017), predicting new venture failure (Antretter, Blohm, Grichnik, & Wincent, 2019), analysing the effect of failure on twitter identity (Tata, Martinez, Garcia, Oesch, & Brusoni, 2017), and analysing entrepreneurs’ social networks (F. Wang, Mack, & Maciewjewski, 2017).

Other studies have drawn on textual data from crowdfunding platforms (Courtney, Dutta, & Li, 2017). For example, drawing on textual data from the crowdfunding platform Kickstarter, W. Wang, Chen, Zhu, & Wang, (2020) use text analytics to examine the relationship between text content and crowdfunding success. Kim, Kim, & Sohn, (2020) use text analytics and deep learning to identify the technological propositions of start-ups on Crunchbase, which would allow companies to screen for potential collaborators. Drawing on textual data from online posts, Williamson, Drencheva, & Battisti, (2020) use machine learning techniques to study the causes of entrepreneurial disappointment and its relationship to depression. Prüfer & Prüfer, (2020) analyse job descriptions to identify the most important entrepreneurship skills.

Other studies have focused on applying novel techniques to analyse more traditional numerical data sources. These studies have focused on a wide variety of entrepreneurship research questions, and have drawn on techniques such as LASSO, artificial neural networks, Bayesian networks and fuzzy set qualitative comparative analysis (fsQCA). Drawing on a large number of variables, Coad & Srhoj, (2019) use LASSO to predict high growth firms. Despite the use of the big data technique, the models had limited accuracy. Studies drawing on machine learning approaches have focused on areas such as the success of entrepreneurs in rural areas (Celbiş, 2021), predicting entrepreneurial success (McKenzie & Sansone, 2019), classifying individuals into entrepreneurs and non-entrepreneurs (Montebruno, Bennett, Smith, & Lieshout, 2020), predicting entrepreneurial intentions of students (Wei et al., 2020) and predicting project performance based on entrepreneurial orientation (Sabahi & Parast, 2020). Although most of these studies have been carried out recently, earlier work, (Haworth, Brearley, & Chell, 1991) used neural networks to classify entrepreneurs into different categories.

A more recent study drawing on neural networks was carried out by Santos et al. (2020) who use artificial neural networks to explore non-linear relationships between affect and business start-up. This study highlights the benefits of using non-linear methods to study entrepreneurship. In one of the few studies applying analytics techniques to the GEM data, Sohn & Lee, (2013) use Bayesian networks to predict Total Early-stage Entrepreneurship (TEA) at the national level.

A recent body of literature has focused on applying fsQCA to study entrepreneurship at the country level identifying more nuanced and non-linear relationships (Beynon, Jones, & Pickernell, 2016, 2018a, 2019; Coduras et al., 2016). These studies draw on the GEM conceptual framework to identify the combinations of variables that result in a binary entrepreneurial outcome; TEA, and separately, the relationship between entrepreneurial climate and self-perceptions. Although these studies do not draw on machine learning techniques, they illustrate the potential to gain new insights through the application of novel techniques.

For example, focusing on the attitudinal and social variables form the 2014 GEM survey, Coduras et al. (2016) compare stepwise regression with fsQCA, arguing that fsQCA provides more information than regression. Comparing the traditional regression approach with the more novel fsQCA allows the researchers to demonstrate the advantages of the novel approach. In their regression model, perceptions of skills, good career choice, and media coverage were statistically significant predictors of TEA. The fsQCA highlights the importance of the individual perceptions and social variables in TEA, whilst drawing out the more nuanced relationships in the data. They apply a new method to study a well-researched area in the entrepreneurship literature– the relationship between early-stage entrepreneurship and attitudes. Applying this novel methodology results in new insights about this problem, enabling more complex causal relationships to be identified.

***3.0 Examining the Determinants of Entrepreneurial Intentions and Activity***

Our empirical study focuses on the determinants of entrepreneurial intentions and early-stage entrepreneurial activity. We focus initially on entrepreneurial intentions as intention is regarded as an antecedent of actual behaviour. Entrepreneurial intention therefore provides an important framework within which to understand and predict entrepreneurial activity (Krueger et al., 2000).

Entrepreneurial intentions have received substantial attention in the entrepreneurship literature, one strand theoretically drawing on the social psychology literature whereby intention has been shown to best predict action. The second strand specifically focuses on entrepreneurial intention (Liñán and Fayolle, 2015). Ajzen’s (1991) theory of planned behaviour (TPB) is an example of the former and has been used extensively in research on entrepreneurial intention and entrepreneurial activity (Karimi et al., 2016). The argument is that attitudes towards behaviour, social norms, and behavioural control influence intentions, and intentions subsequently lead to behaviour. These three antecedents are formed by an individual’s set of beliefs and also influenced by an individual’s background, culture, demographics and experiences. In the entrepreneurial sphere, the model is applicable as the decision to become an entrepreneur and create a new business is said to require time, planning and a degree of cognitive processing, it is therefore a deliberate and conscious decision that can be considered as a planned behaviour (Ozaralli and Rivenburgh, 2016). Papers using this methodology to predict entrepreneurial intention, particularly among university students, have shown consistent findings across countries with the three antecedents found to be significant predictors of intention along with factors such as age, work experience, entrepreneurial education and entrepreneurial role models (Autio et al., 2001; Varamaki et al, 2013; Karimi et al., 2016;, Ozaralli and Rivenburgh, 2016).

Alternatively, the entrepreneurial event model (EEM) (Shapero and Sokol, 1982; Shapero, 1984) specifically focuses on entrepreneurial intention. The argument is that perceived desirability, propensity and feasibility are the underlying factors that influence entrepreneurial intention which in turn influences the behaviour of an individual to start a business. Perceived feasibility and perceived desirability are further guided by situational variables, social factors, individual characteristics and the cultural environment. As with TPB there is empirical support for the EEM with the three antecedents found to be significant predictors of entrepreneurial intention (Kreuger, 1993; Kreuger et al., 2000).

Previous research has further confirmed that entrepreneurial intentions lead to entrepreneurial action (Kautonen et al., 2015; Kolvereid & Isaksen, 2006). Indeed, Kreuger ([1993](https://link.springer.com/article/10.1007/s11365-006-0006-z#ref-CR33)) identifies that an individual’s intent to create a venture precedes the search for and discovery of new venture opportunities. Yet, despite a substantial body of research on the drivers of entrepreneurial intentions, the literature is conflicted around the importance of some variables and whether they moderate or mediate the relationship. Previous studies have argued that personal and situational variables alone are poor predictors of entrepreneurship (Krueger et al., 2000). Indeed, it is suggested that to properly understand entrepreneurial intentions both individual-level and contextual antecedents must be considered, and indeed the multiple interactions between these factors (Schmutzler et al., 2019). In the following section we discuss the predictors of EI and activity identified from the literature.

*3.1 Determinants of Entrepreneurial intentions and activity*

Following Bird (1988), a range of individual and contextual influences have been identified as the key antecedents of entrepreneurial intention. Recent empirical research has further reinforced the importance of these as determinants of entrepreneurship, to include demographic factors, past experience, self-perceptions, and perceptions of the entrepreneurial climate (Beynon, Jones, & Pickernell, 2018) (Arenius & Minniti, 2005) (Koellinger, Minniti, & Schade, 2013). Despite the commonality, however, what is less clear is which are the most important predictors, and whether the relative importance differs between predicting entrepreneurial intention and predicting entrepreneurial activity.

We group these key predictors into three broad categories to identify the individual-level and contextual-level influences, namely self-perceptions; cultural perceptions, and demographics and past experience. We focus on those factors that utilise constructs included in the GEM survey framework to provide comparability for our results.

*3.1.1 Self-perceptions*

Bandura’s (1977) work highlighted the relevance of two important perceptions in social learning: role model perception and self-efficacy. Conceptually, the latter is linked to Ajzen’s (1991) perceived behavioural control element of the TPB in which it is perceived that behaviour is within an individual’s control. Both self-efficacy and role model perception have been recognised as important drivers of entrepreneurial intention and activity (Arenius and Minniti, 2005; Chen et al., 1998; Kreuger, 2003; Kreuger et al., 2000; Liñán, Urbano & Guerrero, 2011; Scherer et al., 1989). Cognitive factors including opportunity and threat perception have also been identified as part of an individual’s entrepreneurship decision process and are driven by perceived competence (Chell, 2013; Krueger and Dickson, 1994). They have been found to be particularly important predictors of entrepreneurial intention (Krueger et al., 2000). Other studies have found similar relationships between perceptual variables and entrepreneurial activity (Koellinger et al., 2013).

Opportunity perception – Entrepreneurship can be viewed as a process of exploiting opportunities (Shapero, 1984), making opportunity perception and identification central in the decision to start a business. A substantial body of research has focused specifically on opportunity identification (see George et al., 2016 for a review), and empirical work has linked opportunity perception with entrepreneurial activity (Arenius & Minniti, 2005).

Self-efficacy – This is based on individuals’ self-perceptions of their skills and abilities. It reflects their thoughts on whether they have the abilities perceived as important to undertake a task, as well as the belief that they will be able to effectively convert those skills into a chosen outcome (Bandura, 1989, 1997). Entrepreneurs need to have the necessary skills to start a business, as well as belief in their skills (Boudreaux, Nikolaev, & Klein, 2019). Empirical research has found self-efficacy to be positively related to opportunity driven entrepreneurship (Boudreaux et al., 2019), and entrepreneurial intentions (SEQUEIRA, MUELLER, & MCGEE, 2007; Wilson, Kickul, & Marlino, 2007; Zhao, Hills, & Seibert, 2005). Indeed, the impact of self-efficacy on entrepreneurial intention is regarded as axiomatic (Schmutzler et al., 2019). Tsai et al (2016) further refine the linkage between self-efficacy and entrepreneurial intention by identifying the mediating role of perceived opportunity.

Networks – Entrepreneurship is regarded as a social phenomenon (Aldrich & Zimmer, 1986) with networks and social capital gained through ties identified as important facilitators of entrepreneurship, assisting the exchange of resources and information. Following Bandura (1977) entrepreneurial role models, mentors, and entrepreneurs within an individual’s social network may enhance an individual’s desire to become an entrepreneur and the entrepreneurial self-efficacy of individuals (Van Auken et al., 2006) as individuals are assumed to learn in a social context through the observation of others with whom they can identify and who perform well. They also help shape entrepreneurial attitudes and behaviour (Hoang & Yi, 2015) and transmit legitimacy to the perceived social value of entrepreneurship (Kacperczyk, 2013). Past research has shown networks to be important in determining entrepreneurial intentions (SEQUEIRA et al., 2007), and activity (Casson and Giusta, 2007; Thornton, Ribeiro-Soriano & Urbano, 2011). Role models have been found to significantly impact entrepreneurial intention (Ahmad, Xavier and Bakar, 2013).

Fear – Risk perceptions, as introduced in entrepreneurial cognition research, have been identified as an important factor influencing the start-up process (Simon et al., 2000). Fear of failure is considered a type of risk perception and typically viewed as a barrier to entrepreneurship (Cacciotti & Hayton, 2015). Risk has been found to play an important role in entrepreneurial intentions (Shane et al. 2003) as potential entrepreneurs are expected to perceive lower risks and show lower fear of failure and, therefore, their intentions of becoming entrepreneurs are higher. Empirical research has also found this negative relationship between fear of failure and different aspects of entrepreneurship, such as opportunity driven entrepreneurship (Boudreaux et al., 2019).

*3.1.2 Cultural Perceptions*

Contextual or situational antecedents of entrepreneurial intention and activity have largely been measured via perceptions of the cultural context whereby culture represents the underlying value systems and norms in a society shaping individual behaviour. Entrepreneurship may be positively influenced by culture particularly where it promotes positive attitudes to business creation and offers social legitimisation to the behaviour (Liñán, Urbano & Guerrero, 2011).

Theoretically, cultural perceptions reflect the subjective norm and behavioural control aspects of Ajzen’s (1991) TPB. In a cross-country comparison, cultural influences can be measured empirically via Hofstede’s (1980) cultural dimensions aspect. In the GEM framework, they are captured via socio-cultural variables relating to perceptions about the social legitimisation of entrepreneurship and include questions on perceptions of whether entrepreneurship is a desirable career choice, whether entrepreneurs have a high status in society and media coverage of successful new businesses. Empirically, the impact of these variables on entrepreneurial intention has been mixed and, where they have been found to be significant, odds ratios are low (Liñán, Urbano & Guerrero., 2011).

Perceptions of entrepreneurship as a desirable career choice has been found to be an important determinant of entrepreneurial intention (Ahmad, Xavier & Bakar, 2013; Beynon et al., 2016, 2018a, 2019; Coduras et al., 2016), although Liñán, Urbano & Guerrero. (2011) find the impact to be weak. Perceptions of the high status and respect of entrepreneurs has been found to have a significant impact on entrepreneurial intention (Liñán, Urbano & Guerrero, 2011) but the results are not consistent (Ahmad, Xavier & Bakar, 2013). Tsai et al. (2016) suggest that there is an indirect relationship, with status and respect for people who have successfully started a new business indirectly influencing entrepreneurial intention through perceived opportunity.

Media coverage of successful entrepreneurs was also found to exert the highest impact of the socio-cultural factors (Beynon et al., 2016, 2018a, 2019; Coduras et al., 2016); Liñán, Urbano & Guerrero. 2011) but other papers found no significant impact (Ahmad, Xavier & Bakar, 2013).

*3.1.3 Demographics and past experience*

Demographic variables have been identified as important determinants of intentions and activity as they reflect the human and resource capital and necessary experience associated with successful business creation. These factors include income, education, employment status, age, and gender (Arenius & Minniti, 2005; Boudreaux et al., 2019; Koellinger et al., 2013; Pathak & Muralidharan, 2021).

Age has been found to have conflicting impacts on entrepreneurial intention (Tsai et al., 2016). Positive effects of age are related to an individual’s quantity of financial and human capital which increases with age (Arenius and Minniti 2005). However entrepreneurial activity may also decrease with age because entrepreneurship is regarded as a more risky employment option (Boden 1999; Parker 2009) which often requires longer working hours (Blanchflower 2004). Levesque and Minniti (2006) identified an inverted-U relationship between age and starting a business, whereas Reynolds et al. (2003) showed that the period of greatest entrepreneurial activity is for those aged 25-34 years old, with activity subsequently declining with age. Evidence for younger people being more likely to become entrepreneurs has also been observed (Arenius and Minniti 2005; Carter et al. 2001, Tsai et al., 2016).

There has also been no clear relationship determined between an individual’s level of education or training and entrepreneurial initiative (Carter et al., 2001). It is suggested that it is not always necessary to have a specific education for possessing entrepreneurial abilities (Leazar, 2005; Murphy et al., 1991). Indeed, Khan et al. (2019) find that education is negatively related to entrepreneurial intention, suggesting that the higher the education level the lower the level of entrepreneurial start-up. Conversely, Levie and Autio (2008) and Davidsson (2015) suggest that education plays a very important role, with entrepreneurship-related human capital allowing individuals to successfully discover, identify, exploit and manage entrepreneurial opportunities.

Previous studies have found that females are less likely to be engaged in entrepreneurial activity than males (Koellinger et al., 2013; Roper & Scott, 2009). It is further suggested that intention to start a business also differs according to gender (Shinnar, Giacomin and Janssen, 2012) with males having stronger intentions than females (see Haus et al., 2013 for a review). In fact, men are associated with higher self-efficacy, risk tolerance, and willingness to start up a business (Díaz-García and Jiménez-Moreno, 2010; Fernández-Serrano et al., 2009; Verheul et al., 2012; Wilson et al., 2007; Zellweger et al. 2011); while women have both lower entrepreneurial self-efficacy and lower entrepreneurial intentions (Chowdhury & Endres, 2005; Gatewood, Shaver, Powers, & Gartner, 2002; Kourilsky & Walstad, 1998).

It is acknowledged that it is difficult for start-up businesses to access external financing. Thus, there is an expectation that business start-up is associated with individuals with higher household incomes. It has been shown empirically that in developed countries entrepreneurial activity increases among women with high incomes (Allen et al., 2008; Vossenberg, 2013).

It is theorised that the likelihood of an individual becoming an entrepreneur is higher if they are in paid work compared to being unemployed, retired or still in education (Sanchez-Escobedo et al., 2014). This is due to the fact that being in work, along with associated previous experience, enables access to the resources, ideas and social capital necessary to set up a business (Allen et al., 2008).

Past failure has been found to be a significant determinant of entrepreneurial intentions (He, Bai, & Xiao, 2020) as it is suggested that both past failure and success reflect useful learning effects and enhance entrepreneurs’ experience and skills, so they are more likely to start again (Stam et al. 2008). Furthermore, entrepreneurs are thought to have resilient qualities and abilities to recover from failure, indicating a likelihood of serial entrepreneurship (Tugade and Fredrikson, 2004). Indeed, having shut a business in the past twelve months is found to be positively related to entrepreneurial activity (Koellinger et al., 2013).

***4.0 Methodology***

*4.1 Data*

The study draws on data from the 2017 GEM survey. The GEM survey is based on an underpinning conceptual framework, which focuses on the factors that influence entrepreneurial behaviour, and in particular the influence of individual characteristics, social values, and national framework conditions (GEM, 2018). National framework conditions focus on the country level factors that might influence entrepreneurship, in particular social, cultural, political and economic factors (GEM, 2018). Individual level factors include demographics, as well as perceptual and motivating factors such as whether the individual knows an entrepreneur, whether they have the skills to start a business, opportunities to start a business, and fear of failure (GEM, 2018). Social values focus on whether entrepreneurship is a good career choice, media portrayal, status of entrepreneurs in society, and the ease of start-up (GEM, 2018). The GEM framework considers entrepreneurship from multiple perspectives, including nascent entrepreneurship, new and established business, and discontinuation (GEM, 2018). In addition, the framework also considers the impact and type of entrepreneurship (GEM, 2018). We draw on this framework, alongside the literature reviewed previously, as a guide to select variables for inclusion in the models.

This study draws on the full annual population dataset from the 2017 GEM. The full list of variables included in the study are shown in Table 1. Two binary dependent variables are considered: entrepreneurial intentions and TEA. Entrepreneurial intentions are measured using the survey question: ‘Are you, alone or with others, expecting to start a new business, including any type of self-employment, within the next three years?’. TEA is a widely used measure encompassing nascent and new business ownership. The nascent stage reflects businesses within the first three months of start-up, and new businesses as those between 3 and 42 months old. Both dependent variables are measured on binary scales where 0 = no and 1 = yes.

For the independent variables, we focus on individual level determinants from the GEM framework. This includes variables relating to self-perceptions, cultural factors, and individual demographics and experience. The self-perceptions include self-efficacy, networks, opportunities, and fear of failure. As shown in table 2, these variables are measured on binary scales, where 0=no and 1=yes. The cultural factors include perception of good opportunities for start-up; perception of entrepreneurship as a good career choice; perception of status of entrepreneurs; status of entrepreneurs in the media; ease of starting a business and social entrepreneurship. These variables are also measured on binary scales, where 0=no and 1=yes.

The individual demographics include age, gender, ethnicity, occupation, household size, whether the person has acted as a business angel, whether they have closed a business in the past two years. Occupation and business ownership is excluded from the models focusing on TEA as it would reveal information about the outcome. The descriptive statistics for the variables are shown in table 2.

Table 1: List of Variables

|  |  |  |
| --- | --- | --- |
| **Gem Variable** | **Name** | **Description** |
| **Dependent Variables** | | |
| TEAyy | Total Early-stage Entrepreneurial Activity | Binary variable where 1= engaged in Total Early-stage Entrepreneurial Activity (nascent or new business owner), and 0=not |
| futsup | Intention to start a business | Binary variable where 1=intending to start a business within 3 years, and 0=not |
| **Self-Perceptions** | | |
| suskill | Skills perception | Binary variable where 1=has the skills, knowledge and experience to start a business, and 0=not |
| knowent | Networks/role model | Binary variable where 1=know someone personally who has started a business in the last 2 years, and 0=not |
| opport | Opportunity perception | Binary variable where 1=believes there will be good opportunities for starting a business in local area in next 6 months, and 0=not |
| fearfail | Fear of failure | Binary variable where 1=believes that fear of failure would prevent them from starting a business and 0=not |
| **Cultural Perceptions** | | |
| nbgoodcq | Good career choice | Binary variable where 1=believes most people would consider starting a new business a desirable career choice, and 0=not |
| easystart |  | Binary variable where 1= believes it is easy to start a business in my country, and 0=not |
| nbstatus | High status | Binary variable where 1=believes that those successful at starting a new business have a high level of status and respect and 0=not |
| nbsocent | Social enterprise | Binary variable where 1=in my country you often see businesses that primarily aim to solve social problems, and 0=not |
| nbmedia | Media coverage | Binary variable where 1=believes that you often see stories in the media about successful new businesses and 0=not |
| **Experience Variables** | | |
| discent | Business exit | Binary variable where 1=in last 12 months has shut down, discontinued or quit a business they owned AND the business did not continue its activities, and 0=not |
| busang | Business angel | Binary variable where 1=business angel in last 3 years (personally provided funds for a business started by someone else), and 0= not |
| ownmge | Business owner | Binary variable where 1=current owner of a business, and 0=not |
| **Demographic Variables** | | |
| hhsize | Household size | Continuous variable indicating number of members of permanent household |
| gender | Gender | Binary variable where 1=male, 0=female |
| GEMHHINC | Household income | Household income recoded into thirds where:  1=lowest 33% tile  2=middle 33% tile  3=upper 33%tile |
| age | Age of respondent | Continuous variable for age in years |
| GEMOCCU | Work status | Categorical variable for work status where:  1 =full: full or part time  2=part time work only  3=retired / disabled  4=homemaker  5=student  6=not working  7=self-employed  8=other |
| UNEDUC | Educational attainment | Highest educational attainment where  1=pre-primary education  2=primary education or first stage of basic education  3=lower secondary or second stage of basic education  4=(upper) secondary education  5=post-secondary non-tertiary education  6=first stage of tertiary education  7=second stage of tertiary education |

Table 2: Descriptive Statistics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grouping** | **Variable** | **Categories** | **Frequency (n)** | **Percentage (/Mean)** |
| Dependent | TEA | Yes | 19012 | 10.92 |
|  |  | No | 155116 | 89.08 |
|  |  | Total | 174128 | 100.00 |
|  | EI | Yes | 35289 | 21.44 |
|  |  | No | 129281 | 78.56 |
|  |  | Total | 164570 | 100.00 |
| Self Perceptions | Opportunity Perception | Yes | 62263 | 43.00 |
|  |  | No | 82549 | 57.00 |
|  |  | Total | 144812 | 100.00 |
|  | Fear of Failure | Yes | 69950 | 54.11 |
|  |  | No | 59320 | 45.89 |
|  |  | Total | 129270 | 100.00 |
|  | Networks | Yes | 65184 | 37.51 |
|  |  | No | 108613 | 62.49 |
|  |  | Total | 173797 | 100.00 |
|  | Skills | Yes | 80758 | 48.45 |
|  |  | No | 85918 | 51.55 |
|  |  | Total | 166676 |  |
| Cultural Perceptions | Good Career Choice | Yes | 97097 | 62.08 |
|  |  | No | 59320 | 37.92 |
|  |  | Total | 156417 | 100.00 |
|  | High Status | Yes | 104933 | 66.55 |
|  |  | No | 52739 | 33.45 |
|  |  | Total | 157672 | 100.00 |
|  | Social Enterprise | Yes | 47777 | 36.30 |
|  |  | No | 83834 | 63.70 |
|  |  | Total | 131611 | 100.00 |
|  | Media Coverage | Yes | 93199 | 12.97 |
|  |  | No | 625559 | 87.03 |
|  |  | Total | 718758 | 100.00 |
| Experience and Demographics | Business Angel | Yes | 9290 | 5.38 |
|  |  | No | 163516 | 94.62 |
|  |  | Total | 172806 | 100.00 |
|  | Business Owner | Yes | 29773 | 17.19 |
|  |  | No | 143408 | 82.81 |
|  |  | Total | 173181 | 100.00 |
|  | Business Exit | Yes | 7143 | 4.13 |
|  |  | No | 165720 | 95.87 |
|  | Total |  | 172863 | 100.00 |
|  | Work Status | 1 =full: full or part time | 74340 | 43.53 |
|  |  | 2=part time work only | 15688 | 9.19 |
|  |  | 3=retired / disabled | 16866 | 9.88 |
|  |  | 4=homemaker | 12686 | 7.43 |
|  |  | 5=student | 8140 | 4.77 |
|  |  | 6=not working | 14716 | 8.62 |
|  |  | 7=self-employed | 28356 | 16.60 |
|  |  | Total | 170792 | 100.00 |
|  | Educational Attainment | 1=pre-primary education | 4033 | 2.36 |
|  |  | 2=primary education or first stage of basic education | 15562 | 9.11 |
|  |  | 3=lower secondary or second stage of basic education | 26863 | 15.73 |
|  |  | 4=(upper) secondary education | 54869 | 32.13 |
|  |  | 5=post-secondary non-tertiary education | 23568 | 13.80 |
|  |  | 6=first stage of tertiary education | 41812 | 24.48 |
|  |  | 7=second stage of tertiary education | 4059 | 2.38 |
|  |  | Total | 170766 | 100.00 |
|  | Household Income | 1=lowest 33% tile | 48739 | 34.91 |
|  |  | 2=middle 33% tile | 43462 | 31.13 |
|  |  | 3=upper 33%tile | 47408 | 33.96 |
|  |  | Total | 139609 | 100.00 |
|  | Gender | 1 | 87769 | 50.40 |
|  |  | 2 | 86359 | 49.60 |
|  |  | Total | 174128 | 100.00 |
|  | Age |  | 171878 | 41.51 |
|  | Household Size |  | 171192 | 3.56 |

*4.2 Machine Learning Method*

This study draws on a machine learning approach to predict and profile entrepreneurial intentions and TEA. Machine learning involves using algorithms to learn patterns in data. There are two broad categories of machine learning problems: supervised and unsupervised. Supervised learning involves applying an algorithm to learn the relationships between input data and a target outcome. In contrast, unsupervised learning involves grouping observations into categories that emerge from patterns in the data. In the present study, we adopt a supervised learning approach to learn the relationships between the input data, and EI and, separately, TEA. The machine learning methodology for this study involves a series of steps, which include data and feature selection; data splitting; training and testing the model; evaluating accuracy; interpretation of output. These stages are discussed in more detail below.

*4.3 Machine Learning Algorithms and Model Training*

The data for training and testing the models is split into two parts using random stratified sampling. 70% of the data is used to train the machine learning models, and 30% of the data is used to test the performance of the model. This split allows for a sufficient quantity of data to train the model, as well as providing an objective assessment of the models’ predictive performance. Evaluating the accuracy of the model using a holdout test set allows us to obtain a more objective assessment of the models’ performance.

Three machine learning algorithms are used to build the models: logistic regression; a single decision tree (recursive partitioning); and gradient boosted machines (GBM). This allows us to answer different research questions using the same data. These techniques were chosen as they highlight some of the key differences between the machine learning approaches, and allow us to draw on both inferential and purely predictive approaches, contrasting the different insights that can be gained from each approach.

Logistic regression is widely used to model data with a categorical dependent variable. Logistic regression has the advantages of being relatively simple, widely available in software packages, and often performs well (Coussement & De Bock, 2013). However, logistic regression assumes a linear relationship between the logit of the dependent variable and the independent variables (Coussement & De Bock, 2013).

The decision tree algorithm builds a model by recursively partitioning the data into groups which are increasingly homogenous across the target variable. The tree starts from a root node, which contains all of the observations. The algorithm then searches across all variables to find the best variable to split on, which is found by measuring the purity of the leaf node that results from splitting on each variable. The same process is applied at each node until a stopping criteria is reached or splitting the data further does not increase the node purity. This results in a terminal node, which is used to predict the outcome.

This results in a model consisting of a series of “if, then, else” rules which can be used to make predictions about the target outcome in cases where this is unknown. The model that is built by the decision tree algorithm has the advantage of being readily interpretable both graphically and using the variable importance measures. Decision trees can be particularly useful in modelling non-linear relationships in the data. In contrast to logistic regression, no assumptions are made about the linearity of the relationship between the dependent and independent variables. Decision tree algorithms have been applied in other areas of the business literature, such as predicting business failure (Gepp, Kumar, & Bhattacharya, 2010), and profiling marketing activities (Duchessi & Lauría, 2013).

GBM’s are an extension of the single decision tree. GBM is an ensemble technique, which fits multiple decision trees to the data. The algorithm works by fitting a series of decision trees to the data, with each subsequent tree aiming to correct the error from the previous trees.

*4.3.1 Parameter Tuning*

Recursive partitioning and gradient boosting both have parameters that should be tuned to maximise the accuracy of the model. For recursive partitioning we tune the complexity parameter. For gradient boosting we tune the tree depth, minimum number of observations in the terminal node and the number of trees. As the parameters must be learned from the data we adopt a 10-fold cross validation approach, repeated five times (Kuhn & Johnson, 2013). A grid of plausible tuning parameters are tested on the training data to identify the parameters that result in the most accurate model. To tune the parameters, the training data is split into 10 parts, with each model built using 9 parts of the data, and tested using the one part that was left out. Each of the candidate parameters is evaluated for accuracy. The process is repeated five times for robustness. The overall process identifies the model parameters that maximise the predictive accuracy of the model.

Both dependent variables suffer from class imbalance, which biases the predictions towards the majority class. To reduce the impact of this problem, up sampling of the minority class is used during the cross-validation process. Each cross-validation fold was up sampled separately, rather than up sampling the entire training dataset prior to the cross validation. This ensures that the individual folds are balanced, and allows performance to be assessed on the unbalanced data. The training dataset was not up sampled, allowing us to evaluate the model performance based on the original distribution of the data. Other common resampling techniques were also evaluated during the analysis process, including SMOTE, ROSE and down sampling. SMOTE and ROSE resulted in similar performance to up sampling, and down sampling was slightly worse. We therefore elected to present results from models built using the up sampled data as it is both intuitive and of comparable accuracy.

*4.3.2 Evaluating and Interpreting the Models*

The predictive accuracy of the models are evaluated using the hold out test set. Due to the class imbalance of the dependent variables, the main metric used to evaluate the model accuracy is the Area under the curve of the receiver operating characteristic (AUC-ROC). We also present kappa and the overall accuracy, although would caution against using accuracy by itself, due to class imbalance.

Models are interpreted using the variable importance scores.The variable importance measures are calculated differently for each of the three machine learning algorithms. In all cases they are standardised to have a maximum value of 100. For the logistic regression model, the variable importance scores are calculated based on the model coefficients. For RPART, the variable importance scores are calculated based on the reduction in the loss function from the splits and surrogate splits on the variable (Thereneau, T. M. and Atkinson, 2015). The variable importance measures for the GBM are calculated using the same principle, and summed across the boosting iterations (Kuhn, 2017).

It is also worth noting that the GEM dataset has a considerable amount of missing data, as shown later in table 2 of the results section. This has implications for the interpretation of the models, due to differences in how missing data is handled. For the logistic regression models, observations with any missing data are excluded listwise which reduces the number of observations included in the model building. However, for RPART and GBM, observations with missing data are still included in the modelling process. This is achieved by using surrogate splits. If an observation has missing data on a splitting variable, the next best variable will be used instead. This means more data is used in the model building process. It is important to note that observations with missing data on a particular variable are not used to calculate node impurity when splitting on that variable (Thereneau, T. M. and Atkinson, 2015).

*4.4 Implementing a Traditional Approach for Comparison*

In addition to the machine learning approach we also implement a traditional logistic regression approach to identify the variables that are statistically significant determinants of EI and TEA. It should also be noted that whilst we use the same logistic regression algorithm in the machine learning approach, the approach taken in this stage of the analysis is quite different. In our traditional approach we use the full dataset to build the models. Three models are built for each dependent variable. The variables are entered into models in blocks, with the demographic and experiential variables added in the first model, the individual perceptual variables in the second, and the cultural variables in the third. The change in the log likelihood was monitored at each stage of the model building process.

***5.0 Results***

Although we do not discuss the results of the traditional logistic regression models in detail, we present these in Table 3 to enable comparison between the coefficients in the logistic regression models and the variable importance scores from the machine learning approach. The results of model 3 show that all except two variables are significantly related to EI. The two variables that do not show a statistically significant association are having an education level of first stage of tertiary education and being in the upper third of household income. Model 6 shows that all except one of the variables included are statistically significant determinants of TEA. Having a household income in the top third is the only variable found not to be statistically significant in the model. The results of the logistic regression models are largely in line with the expectations based on the literature review, with the majority of demographic and attitudinal variables having a statistically significant effect on entrepreneurial intentions and TEA in the expected direction. We therefore turn our focus to the machine learning approach and the relative importance of the variables.

Table 3: Traditional Logistic Regression Models for Comparison

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Dependent variable:* | | | | | |
|  |  | | | | | |
|  | EI | | | TEA |  | |
|  | 1 | 2 | 3 | 4 | 5 | 6 |
|  | | | | | | |
| **Self-Perceptions** | |  |  |  |  |  |
| Networks |  | 0.318\*\*\* | 0.229\*\*\* |  | 0.349\*\*\* | 0.262\*\*\* |
|  |  | -0.017 | -0.019 |  | -0.016 | -0.019 |
|  |  |  |  |  |  |  |
| Opportunities |  | 0.462\*\*\* | 0.397\*\*\* |  | 0.462\*\*\* | 0.399\*\*\* |
|  |  | -0.016 | -0.019 |  | -0.016 | -0.019 |
|  |  |  |  |  |  |  |
| Self-efficacy |  | 0.874\*\*\* | 0.828\*\*\* |  | 0.936\*\*\* | 0.885\*\*\* |
|  |  | -0.018 | -0.02 |  | -0.017 | -0.02 |
|  |  |  |  |  |  |  |
| Fear of failure |  | -0.138\*\*\* | -0.156\*\*\* |  | -0.142\*\*\* | -0.160\*\*\* |
|  |  | -0.016 | -0.019 |  | -0.016 | -0.019 |
|  |  |  |  |  |  |  |
| **Cultural Perceptions** | |  |  |  |  |  |
| Good career choice |  |  | 0.266\*\*\* |  |  | 0.264\*\*\* |
|  |  |  | -0.02 |  |  | -0.02 |
|  |  |  |  |  |  |  |
| High status |  |  | 0.166\*\*\* |  |  | 0.167\*\*\* |
|  |  |  | -0.02 |  |  | -0.02 |
|  |  |  |  |  |  |  |
| Media coverage |  |  | 0.160\*\*\* |  |  | 0.163\*\*\* |
|  |  |  | -0.02 |  |  | -0.02 |
|  |  |  |  |  |  |  |
| Easy to start |  |  | 0.084\*\*\* |  |  | 0.084\*\*\* |
|  |  |  | -0.019 |  |  | -0.019 |
|  |  |  |  |  |  |  |
| Social Enterprise |  |  | 0.051\*\*\* |  |  | 0.046\*\* |
|  |  |  | -0.02 |  |  | -0.019 |
| **Experience and Demographics** | |  |  |  |  |  |
| Own manager |  | 0.079\*\*\* | 0.081\*\*\* |  |  |  |
|  |  | -0.024 | -0.027 |  |  |  |
|  |  |  |  |  |  |  |
| Business Exit |  | 0.831\*\*\* | 0.831\*\*\* |  | 0.850\*\*\* | 0.847\*\*\* |
|  |  | -0.033 | -0.037 |  | -0.033 | -0.037 |
|  |  |  |  |  |  |  |
| Business Angel |  | 0.720\*\*\* | 0.692\*\*\* |  | 0.724\*\*\* | 0.704\*\*\* |
|  |  | -0.03 | -0.034 |  | -0.03 | -0.033 |
|  |  |  |  |  |  |  |
| Age | -0.029\*\*\* | -0.028\*\*\* | -0.025\*\*\* | -0.017\*\*\* | -0.030\*\*\* | -0.028\*\*\* |
|  | -0.001 | -0.001 | -0.001 | -0.001 | -0.001 | -0.001 |
|  |  |  |  |  |  |  |
| Gender | -0.216\*\*\* | -0.117\*\*\* | -0.099\*\*\* | -0.298\*\*\* | -0.117\*\*\* | -0.113\*\*\* |
|  | -0.015 | -0.017 | -0.019 | -0.017 | -0.016 | -0.018 |
|  |  |  |  |  |  |  |
| Household size | 0.126\*\*\* | 0.107\*\*\* | 0.090\*\*\* | 0.070\*\*\* | 0.113\*\*\* | 0.095\*\*\* |
|  | -0.004 | -0.004 | -0.005 | -0.005 | -0.004 | -0.005 |
|  |  |  |  |  |  |  |
| Education: Pre primary | -0.335\*\*\* | -0.415\*\*\* | -0.304\*\*\* | -0.031 | -0.426\*\*\* | -0.314\*\*\* |
|  | -0.051 | -0.057 | -0.063 | -0.066 | -0.056 | -0.062 |
|  |  |  |  |  |  |  |
| Education: primary | -0.373\*\*\* | -0.501\*\*\* | -0.385\*\*\* | -0.139\*\* | -0.527\*\*\* | -0.412\*\*\* |
|  | -0.049 | -0.054 | -0.06 | -0.063 | -0.053 | -0.059 |
|  |  |  |  |  |  |  |
| Education: lower secondary | -0.294\*\*\* | -0.433\*\*\* | -0.341\*\*\* | -0.111\* | -0.460\*\*\* | -0.364\*\*\* |
|  | -0.047 | -0.053 | -0.058 | -0.061 | -0.052 | -0.057 |
|  |  |  |  |  |  |  |
| Education: secondary | -0.151\*\*\* | -0.296\*\*\* | -0.185\*\*\* | -0.053 | -0.322\*\*\* | -0.197\*\*\* |
|  | -0.049 | -0.055 | -0.061 | -0.063 | -0.054 | -0.06 |
|  |  |  |  |  |  |  |
| Education: post secondary | -0.019 | -0.225\*\*\* | -0.083 | 0.034 | -0.254\*\*\* | -0.104\* |
|  | -0.048 | -0.054 | -0.059 | -0.062 | -0.052 | -0.058 |
|  |  |  |  |  |  |  |
| Education: 1st stage tertiary | 0.167\*\*\* | -0.031 | 0.214\*\*\* | 0.074 | -0.031 | 0.237\*\*\* |
|  | -0.062 | -0.07 | -0.08 | -0.078 | -0.068 | -0.078 |
|  |  |  |  |  |  |  |
| Household income: middle 1/3 | -0.053\*\*\* | -0.087\*\*\* | -0.071\*\*\* | 0.087\*\*\* | -0.087\*\*\* | -0.070\*\*\* |
|  | -0.018 | -0.02 | -0.023 | -0.023 | -0.02 | -0.023 |
|  |  |  |  |  |  |  |
| Household income: upper 1/3 | 0.096\*\*\* | -0.041\*\* | -0.01 | 0.410\*\*\* | -0.033 | -0.0002 |
|  | -0.018 | -0.02 | -0.023 | -0.022 | -0.02 | -0.023 |
|  |  |  |  |  |  |  |
| Occupation: part time | 0.055\*\* | 0.120\*\*\* | 0.137\*\*\* |  |  |  |
|  | -0.025 | -0.029 | -0.033 |  |  |  |
|  |  |  |  |  |  |  |
| Occupation: retired/disabled | -1.009\*\*\* | -0.876\*\*\* | -0.917\*\*\* |  |  |  |
|  | -0.046 | -0.051 | -0.06 |  |  |  |
|  |  |  |  |  |  |  |
| Occupation: homemaker | -0.151\*\*\* | -0.006 | -0.089\*\* |  |  |  |
|  | -0.031 | -0.035 | -0.039 |  |  |  |
|  |  |  |  |  |  |  |
| Occupation: student | -0.494\*\*\* | -0.249\*\*\* | -0.276\*\*\* |  |  |  |
|  | -0.037 | -0.042 | -0.048 |  |  |  |
|  |  |  |  |  |  |  |
| Occupation: not working | 0.133\*\*\* | 0.230\*\*\* | 0.163\*\*\* |  |  |  |
|  | -0.026 | -0.029 | -0.034 |  |  |  |
|  |  |  |  |  |  |  |
| Occupation: self-employed | 0.610\*\*\* | 0.184\*\*\* | 0.149\*\*\* |  |  |  |
|  | -0.018 | -0.025 | -0.028 |  |  |  |
|  |  |  |  |  |  |  |
| Constant | -0.294\*\*\* | -0.934\*\*\* | -1.308\*\*\* | -1.585\*\*\* | -0.843\*\*\* | -1.236\*\*\* |
|  | -0.057 | -0.066 | -0.076 | -0.071 | -0.063 | -0.072 |
|  |  |  |  |  |  |  |
|  | | | | | | |
| Observations | 126,463 | 101,550 | 71,799 | 135,201 | 103,414 | 73,216 |
| Log Likelihood | -63,438.05 | -49,472.25 | -37,223.97 | -47,550.83 | -50,651.43 | -38,108.73 |
| Akaike Inf. Crit. | 126,912.10 | 98,994.49 | 74,507.94 | 95,125.66 | 101,338.90 | 76,263.46 |
|  | | | | | | |
| *Note:* | \*p\*\*p\*\*\*p<0.01 | | | | | |

One of the aims of our machine learning approach is to identify the most important variables in predicting EI and TEA. This is similar to the dominance analysis approach adopted in the wider management literature (Arin et al., 2015; Hakanen, Bakker, & Turunen, 2021; Kumar, Kee, & Charles, 2010). We draw on the variable importance measures produced through the machine learning approach to identify the relative importance of the independent variables. Table 4 presents the results of the three models that were built using a machine learning approach to predict entrepreneurial intentions. The variable importance scores and the model performance metrics are shown. Self-efficacy is the most important predictor in the logistic regression model, followed by age, having closed a business, and perceiving good opportunities for entrepreneurship. In the recursive partitioning model, age and self-efficacy are also the two most important predictors of EI, followed by a work status of retired, disabled, and perceiving good opportunities for entrepreneurship.

In terms of the perceptual variables, the variable importance scores for the logistic regression model show that having the skills to start a business is the most important predictor of intentions. Perceiving good opportunities to start a business and knowing an entrepreneur are also important attitudinal variables. Relevant experience are also important predictors, in terms of being a business angel, having discontinued a business, and having an occupation of retired / disabled. Demographics are also important, particularly age and household size. In the GBM, the most important predictors are an occupation of retired, disabled, having closed a business, experience as a business angel, and self-efficacy.

The structure of the final decision tree that was built using recursive partitioning is show in Figure 1. This tree structure can be interpreted alongside the variable importance scores to add additional information about the structure of the final model. The tree shows that the variable that best separates the data by EI is self-efficacy, with individuals less likely to intend to start a business when they feel they do not have the skills to do so (i.e. suskillyes = 0). This is followed by age on both the left-hand side and right-hand side of the tree. On the left-hand side of the tree, people who do not have the skills to start a business and who are aged 41 or over are predicted not to have entrepreneurial intentions. Traversing the right-hand side of the tree shows that people with the skills to start a business, who are aged under 50 and who perceive good opportunities are predicted to have entrepreneurial intentions. The decision rules for lower branches of the tree involve a more complex set of rules, and we therefore focus on the higher levels of the tree, which show the most important splits. Overall, the most accurate model at predicting EI was the GBM, with an AUC-ROC of 0.759.

Table 4: Entrepreneurial Intentions variable importance and model performance

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dependent Variable: Entrepreneurial Intentions | | | | | |
| Logistic Regression | | Recursive Partitioning | | Gradient Boosted Machine | |
| Variable | Importance | Variable | Importance | Variable | Importance |
| Self-efficacy | 100 | Age | 100 | Occ: retired/ disabled | 100 |
| Age | 79.135 | Self-efficacy | 88.4263 | Business exit | 46.5364 |
| Business exit | 55.498 | Network | 85.974 | Business Angel | 33.4158 |
| Opportunity | 49.942 | Occ: retired/ disabled | 81.3262 | Self-efficacy | 21.6196 |
| Business angel | 49.434 | Opportunity | 79.4921 | Educ: tertiary | 8.0279 |
| Household size | 43.414 | Household size | 45.8678 | Age | 6.5218 |
| Occ: retired/ disabled | 37.056 | Business Exit | 23.8965 | Opportunity | 5.7117 |
| Good career | 29.27 | Business Angel | 21.47 | Owner manager | 4.7028 |
| Network | 28.442 | Owner manager | 11.7751 | Occ: self-employed | 4.3237 |
| Good status | 21.292 | Good career | 7.8389 | Occ: homemaker | 3.2493 |
| Media coverage | 19.384 | Occ: self-employed | 3.9187 | Household size | 2.4087 |
| Fear of Failure | 18.214 | Easy to start | 1.9164 | Network | 1.9431 |
| Educ: primary | 16.616 | High status | 1.5125 | Good career | 1.7942 |
| Educ: lower secondary | 16.044 | Social enterprise | 0.6488 | Educ: post secondary | 1.7198 |
| Educ: pre primary | 13.021 | Media coverage | 0.5763 | Occ: not working | 1.0868 |
| Occ: student | 12.422 | Occ: Student | 0.4727 | Media coverage | 1.0025 |
| Gender | 11.633 | gender2 | 0.3476 | High status | 0.9288 |
| Occ: not working | 11.437 | Educ: secondary | 0.3053 | Easy to start | 0.8949 |
| Occ: self-employed | 11.432 | Educ: post-secondary | 0.2451 | Edu: upper secondary | 0.7937 |
| Easy to start | 9.848 | Educ: lower-secondary | 0.1878 | Educ: pre primary | 0.7438 |
|  |  |  |  |  |  |
| ROC | 0.7378 |  | 0.7248 |  | 0.759 |
| Kappa | 0.2092 |  | 0.2707 |  | 0.3014 |
| Accuracy | 0.7343 |  | 0.6627 |  | 0.7706 |

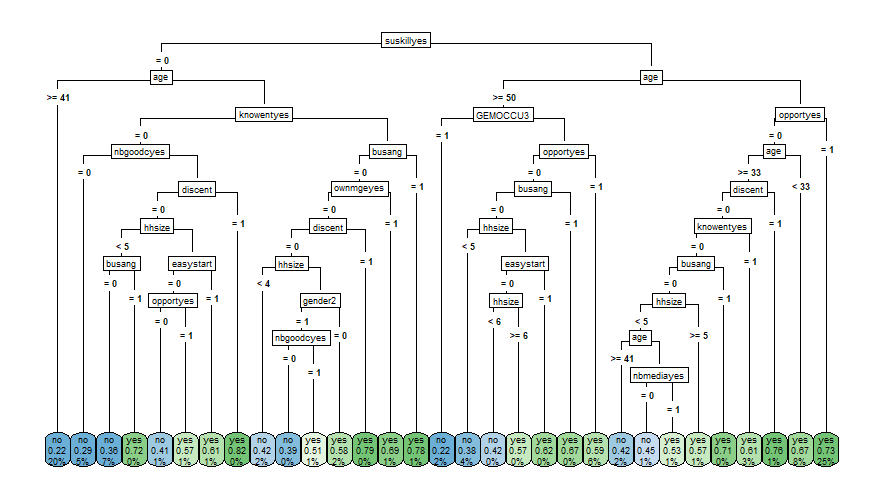


Figure 1: Decision tree model predicting EI.

Table 5 shows the variable importance measures for the models aimed at predicting TEA. The three perceptual variables of self-efficacy, knowing an entrepreneur, and perceiving good opportunities are the most important predictors in the logistic regression model. This is followed by the persons age, and fear of failure. The same pattern emerges in the recursive partitioning model. The GBM exhibits a similar pattern, with self-efficacy and knowing an entrepreneur the two most important predictors. This is followed by age, perceiving good opportunities and having discontinued a business. Figure two shows a pruned version of the final decision tree, with the full tree structure shown in appendix 2. The tree splits first on self-efficacy, with individuals who do not have the skills to start a business being less likely to do so. The second split on both sides of the tree is on whether or not the person knows an entrepreneur, and in both cases a higher proportion of people that know an entrepreneur are engaged in TEA compared with those who do not. Overall, the GBM is most accurate, with an AUC-ROC of 0.7827.

Table 5: TEA variable importance and model performance

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dependent Variable: TEA | | | | | |
| Logistic Regression | | Recursive Partitioning | | Gradient Boosted Machine | |
| Variable | Importance | Variable | Importance | Variable | Importance |
| Self-efficacy | 100 | Self-efficacy | 100 | Self-efficacy | 100 |
| Network | 56.139 | Network | 83.3201 | Network | 25.821 |
| Opportunity | 32.087 | Opportunity | 52.8563 | Age | 16.522 |
| Age | 30.652 | Age | 44.0812 | Opportunity | 9.9728 |
| Fear of failure | 26.945 | Fear of failure | 26.2655 | Business Exit | 6.0819 |
| Business Exit | 24.695 | Business exit | 16.1845 | Fear of failure | 5.5108 |
| Business Angel | 22.295 | Business angel | 14.9444 | Business Angel | 4.9558 |
| Household income upper 1/3 | 12.218 | Household size | 9.1925 | Household size | 4.1782 |
| Household size | 11.852 | Gender | 2.0347 | Easy to start | 3.144 |
| Gender | 11.084 | Educ: post secondary | 1.4702 | Household income upper 1/3 | 2.4947 |
| Educ: lower secondary | 9.162 | Good career | 1.4464 | Social enterprise | 1.9185 |
| Media coverage | 8.901 | Household income upper 1/3 | 1.0219 | High status | 1.7849 |
| Educ: primary | 8.507 | Educ: lower secondary | 0.8057 | Media coverage | 1.7212 |
| Educ: post secondary | 7.578 | Media coverage | 0.633 | Good career | 1.6076 |
| Educ: upper secondary | 6.285 | Educ: secondary | 0.4344 | Gender | 1.0378 |
| Good career | 6.226 | Educ: pre primary | 0.3814 | Household income middle 1/3 | 0.8407 |
| Educ: 1st stage tertiary | 4.631 | Educ: primary | 0.3401 | Educ: lower secondary | 0.7209 |
| High status | 3.444 | Household income middle 1/3 | 0.3068 | Educ: post secondary | 0.541 |
| Educ: pre primary | 3.437 | Easy to start | 0.2857 | Educ: pre-primary | 0.473 |
| Easy to start | 2.568 | Social enterprise | 0.205 | Educ: primary | 0.4303 |
|  |  |  |  |  |  |
| ROC | 0.751 |  | 0.748 |  | 0.7827 |
| Kappa | 0.1807 |  | 0.1844 |  | 0.2125 |
| Accuracy | 0.5910 |  | 0.6558 |  | 0.6933 |

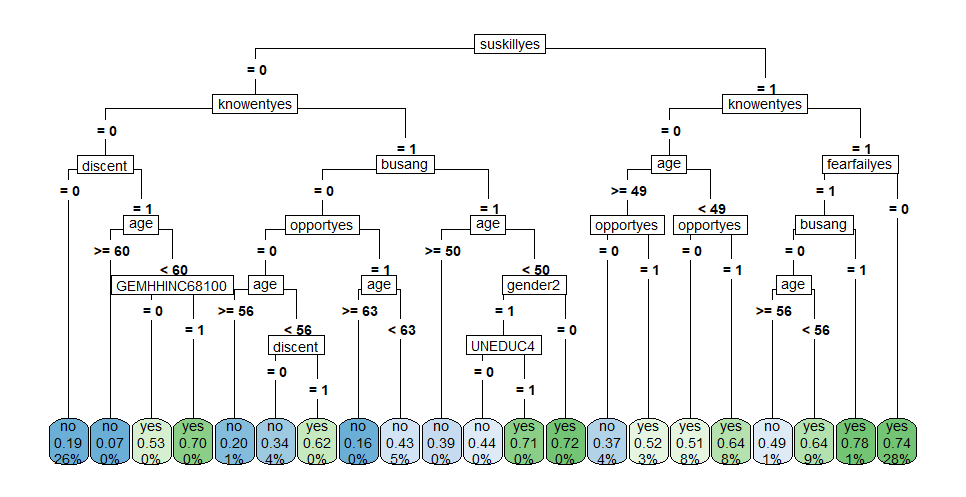


Figure 2: Decision tree model predicting TEA. The figure has been pruned for presentation purposes.

***6.0 Discussion***

The results presented in the previous section highlight the most important predictors of EI and TEA, and although there are some consistencies amongst predictors there are also differences in terms of relative importance and accuracy. We discuss these further here, drawing on insights from the existing literature to explain our findings. Although nearly all of the variables are found to be statistically significant in the regression models, the machine learning approach shows that not all of the variables are important in making actual predictions about EI and TEA. The predictive accuracy of the models are dominated by a smaller subset of variables.

Across all models, the self-perception variables are consistently amongst the most predictive factors of both EI and TEA. The importance of self-perception factors as determinants of entrepreneurship is consistent with the wider literature (Arenius & Minniti, 2005).Having the skills to start a business is the most important predictor of TEA across all three models, and although of lower overall importance in the EI models, is the first split in the EI decision tree, and is of relatively high importance across the three models. This is consistent with theoretical arguments about the importance of self-efficacy (Bandura, 1989, 1997), and with the wider empirical literature (Boudreaux et al., 2019) (SEQUEIRA et al., 2007; Wilson et al., 2007; Zhao et al., 2005).

Knowing an entrepreneur is consistently of relatively higher importance in predicting TEA compared with predicting EI. This could suggest that networks of entrepreneurs are important in actually starting a business, but are of less importance in intending to start a business. In contrast to arguments from the wider literature, fear of failure is found to be of relatively lower importance in predicting EI, but consistent with the literature, is amongst the top predictors of TEA.

Opportunities to start a business are important predictors of EI and TEA, which aligns with arguments from the literature about the fundamental role of opportunities in entrepreneurship (e.g. Shapero, 1984). However, it is perhaps surprising that perceived opportunities are not of even more importance in predicting EI and TEA. One potential explanation for this is that we do not distinguish between opportunity driven and necessity driven entrepreneurship although previous research has suggested that the main dimensions of motivations are similar for entrepreneurs in various sub-groups including deprived communities and those in more affluent regions, the former usually associated with necessity-driven reasons (Stephan, Hart & Drews, 2015). Rather, it may be the case that perception of opportunity in itself differs from opportunity evaluation and development; the latter required for a viable business (Ardichvili, Cardozo & Ray, 2003).

The experiential variables of having discontinued a business, and being a business angel are of relatively high importance in the EI models. This finding is consistent with the wider literature that argues that entrepreneurs learn from past experience, which increases the likelihood of entrepreneurship (Stam et al. 2008). Although important in the EI models, past experience is of relatively less predictive importance in the TEA models. An occupation status of retired/disabled is also relatively important in predicting EI, which is not surprising as people who have left the workforce are less likely to intend to start a business (Kautonen, Down & Minniti, 2014; Lévesque and Minniti, 2006). However, this is less important in predicting TEA.

In terms of the demographics, age is consistently of relatively high importance across the EI and TEA models, which is consistent with the wider literature (Arenius & Minniti, 2005). However, other studies have argued that the relationship between age and entrepreneurship is non-linear (Levesque and Minniti, 2006). It is therefore useful to refer to the decision tree structure in figures 1 and 2, which also highlight a more complex relationship between age and entrepreneurship. However, gender is found to be less important in predicting both EI and TEA. One potential reason for this could be due to interrelationships between gender and other concepts such as self-efficacy (Chowdhury & Endres, 2005; Gatewood, Shaver, Powers, & Gartner, 2002; Kourilsky & Walstad, 1998), which account for the majority of the predictive ability. Perhaps surprisingly, household income is relatively less important in both the EI and TEA models. Although the wider literature does point to conflicting arguments about the effect of income on entrepreneurship, with some arguing that a high income encourages people to remain in paid employment, whereas others argue that income provides the necessary resources to start a business. These conflicting perspectives could both hold true in different scenarios, decreasing the overall predictive ability of this variable. Similarly, education is also less important in predicting both EI and TEA, which backs up arguments from the literature where no clear relationship emerges (Carter et al., 2001). Other authors also highlight a more general indirect relationship between personal factors and entrepreneurship (Krueger et al., 2000). Although statistically significant in the regression model, the cultural variables are not found to be important predictors of EI or TEA. This is consistent with the wider literature that has found these variables to have low odds ratios (Liñán, Urbano & Guerrero., 2011).

In terms of overall model accuracy, when focusing on the AUC-ROC, the GBMs are most accurate across both EI and TEA. Overall, the models are more accurate in predicting TEA compared to EI. We posit that this is because TEA is a more concrete and specific measure than EI. This is in line with wider arguments from the literature (Autio et al., 2001) that due to its indeterminate nature EI should be measured using more detailed instruments that include a few items and response categories rather than as a dichotomous variable (Schlaegel & Koenig, 2014). Another key point from our analysis is that it is difficult to predict both EI and TEA. All models have quite a high level of predictive error. This is consistent with the wider literature that entrepreneurship is a difficult phenomenon to predict, with multiple determinants (Kuckertz, Berger and Allmendinger, 2015). It also leaves open the possibility of future studies identifying more variables that can enhance the predictive accuracy of the models.

***7.0 Conclusions***

In this study we have implemented a machine learning methodology to examine the relative importance of predictors of EI and TEA, as well as highlighting the complex interrelationships amongst variables. The decision tree highlights the complex interrelationships between factors predicting EI and TEA, whilst the variable importance measures allow us to determine the relative importance of the variables. Self-perceptions are found to be the most important predictors of EI and TEA. Age is an important demographic predictor, but other demographic factors are relatively less important. Cultural perceptions are also found to be relatively less important. Our approach has also allowed us to highlight differences in the relative importance of the predictors of EI and TEA. These findings contribute to our understanding of the determinants of EI and TEA. We also make a methodological contribution by highlighting the differences between the traditional approach and the machine learning approach, showcasing how the machine learning approach can complement the traditional approach in drawing out new theoretical insights and contributions.

Although this study has identified the usefulness of the machine learning approach in studying entrepreneurship, there are also limitations. One limitation is that there is a substantial amount of missing data across some of the GEM variables. Despite this, we still have a large dataset to work with. The cross-sectional nature of the data does not allow us to include intentions as a determinant of TEA. Future studies could potentially enhance the predictive accuracy by including this, and other variables such as wider institutional, regional or country level determinants.

Appendix 1: EI Full Recursive Partitioning Model

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| n=180994 | | | | | | | |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| node), split, n, loss, yval, (yprob) | | | | |  |  |  |  |  |  |
| \* denotes terminal node | | | |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| 1) root 180994 90497 no (0.5000000 0.5000000) | | | | | | |  |  |  |  |
| 2) suskillyes< 0.5 79527 26454 no (0.6673583 0.3326417) | | | | | | | |  |  |  |
| 4) age>=40.5 36775 7907 no (0.7849898 0.2150102) \* | | | | | | | |  |  |  |
| 5) age< 40.5 42752 18547 no (0.5661723 0.4338277) | | | | | | | |  |  |  |
| 10) knowentyes< 0.5 29361 11252 no (0.6167705 0.3832295) | | | | | | | | |  |  |
| 20) nbgoodcyes< 0.5 9717 2858 no (0.7058763 0.2941237) \* | | | | | | | | |  |  |
| 21) nbgoodcyes>=0.5 19644 8394 no (0.5726940 0.4273060) | | | | | | | | |  |  |
| 42) discent< 0.5 19017 7882 no (0.5855287 0.4144713) | | | | | | | | |  |  |
| 84) hhsize< 4.5 13213 4866 no (0.6317263 0.3682737) | | | | | | | | |  |  |
| 168) busang< 0.5 12851 4605 no (0.6416621 0.3583379) \* | | | | | | | | |  |  |
| 169) busang>=0.5 362 101 yes (0.2790055 0.7209945) \* | | | | | | | | |  |  |
| 85) hhsize>=4.5 5804 2788 yes (0.4803584 0.5196416) | | | | | | | | |  |  |
| 170) easystartyes< 0.5 3397 1551 no (0.5434207 0.4565793) | | | | | | | | | |  |
| 340) opportyes< 0.5 2448 1006 no (0.5890523 0.4109477) \* | | | | | | | | | |  |
| 341) opportyes>=0.5 949 404 yes (0.4257113 0.5742887) \* | | | | | | | | | |  |
| 171) easystartyes>=0.5 2407 942 yes (0.3913585 0.6086415) \* | | | | | | | | | |  |
| 43) discent>=0.5 627 115 yes (0.1834131 0.8165869) \* | | | | | | | | |  |  |
| 11) knowentyes>=0.5 13391 6096 yes (0.4552311 0.5447689) | | | | | | | | |  |  |
| 22) busang< 0.5 12227 5841 yes (0.4777133 0.5222867) | | | | | | | | |  |  |
| 44) ownmgeyes< 0.5 10683 5325 no (0.5015445 0.4984555) | | | | | | | | |  |  |
| 88) discent< 0.5 10220 4957 no (0.5149706 0.4850294) | | | | | | | | |  |  |
| 176) hhsize< 3.5 4449 1890 no (0.5751854 0.4248146) \* | | | | | | | | |  |  |
| 177) hhsize>=3.5 5771 2704 yes (0.4685496 0.5314504) | | | | | | | | |  |  |
| 354) gender2>=0.5 2890 1388 no (0.5197232 0.4802768) | | | | | | | | | |  |
| 708) nbgoodcyes< 0.5 817 322 no (0.6058752 0.3941248) \* | | | | | | | | | |  |
| 709) nbgoodcyes>=0.5 2073 1007 yes (0.4857694 0.5142306) \* | | | | | | | | | |  |
| 355) gender2< 0.5 2881 1202 yes (0.4172162 0.5827838) \* | | | | | | | | | |  |
| 89) discent>=0.5 463 95 yes (0.2051836 0.7948164) \* | | | | | | | | |  |  |
| 45) ownmgeyes>=0.5 1544 483 yes (0.3128238 0.6871762) \* | | | | | | | | |  |  |
| 23) busang>=0.5 1164 255 yes (0.2190722 0.7809278) \* | | | | | | | | |  |  |
| 3) suskillyes>=0.5 101467 37424 yes (0.3688293 0.6311707) | | | | | | | | |  |  |
| 6) age>=49.5 24516 11502 no (0.5308370 0.4691630) | | | | | | | |  |  |  |
| 12) GEMOCCU3>=0.5 3972 858 no (0.7839879 0.2160121) \* | | | | | | | |  |  |  |
| 13) GEMOCCU3< 0.5 20544 9900 yes (0.4818925 0.5181075) | | | | | | | | |  |  |
| 26) opportyes< 0.5 9680 4189 no (0.5672521 0.4327479) | | | | | | | | |  |  |
| 52) busang< 0.5 8843 3628 no (0.5897320 0.4102680) | | | | | | | | |  |  |
| 104) hhsize< 4.5 7154 2721 no (0.6196533 0.3803467) \* | | | | | | | | |  |  |
| 105) hhsize>=4.5 1689 782 yes (0.4629959 0.5370041) | | | | | | | | |  |  |
| 210) easystartyes< 0.5 1033 503 no (0.5130687 0.4869313) | | | | | | | | | |  |
| 420) hhsize< 5.5 565 237 no (0.5805310 0.4194690) \* | | | | | | | | |  |  |
| 421) hhsize>=5.5 468 202 yes (0.4316239 0.5683761) \* | | | | | | | | |  |  |
| 211) easystartyes>=0.5 656 252 yes (0.3841463 0.6158537) \* | | | | | | | | | |  |
| 53) busang>=0.5 837 276 yes (0.3297491 0.6702509) \* | | | | | | | | |  |  |
| 27) opportyes>=0.5 10864 4409 yes (0.4058358 0.5941642) \* | | | | | | | | |  |  |
| 7) age< 49.5 76951 24410 yes (0.3172149 0.6827851) | | | | | | | |  |  |  |
| 14) opportyes< 0.5 30937 11974 yes (0.3870446 0.6129554) | | | | | | | | |  |  |
| 28) age>=32.5 16518 7255 yes (0.4392178 0.5607822) | | | | | | | |  |  |  |
| 56) discent< 0.5 15188 6930 yes (0.4562813 0.5437187) | | | | | | | | |  |  |
| 112) knowentyes< 0.5 8945 4466 no (0.5007267 0.4992733) | | | | | | | | |  |  |
| 224) busang< 0.5 8463 4122 no (0.5129387 0.4870613) | | | | | | | | |  |  |
| 448) hhsize< 4.5 6234 2852 no (0.5425088 0.4574912) | | | | | | | | |  |  |
| 896) age>=40.5 3097 1291 no (0.5831450 0.4168550) \* | | | | | | | | |  |  |
| 897) age< 40.5 3137 1561 no (0.5023908 0.4976092) | | | | | | | | |  |  |
| 1794) nbmediayes< 0.5 1332 596 no (0.5525526 0.4474474) \* | | | | | | | | | |  |
| 1795) nbmediayes>=0.5 1805 840 yes (0.4653740 0.5346260) \* | | | | | | | | | | |
| 449) hhsize>=4.5 2229 959 yes (0.4302378 0.5697622) \* | | | | | | | | | |  |
| 225) busang>=0.5 482 138 yes (0.2863071 0.7136929) \* | | | | | | | | |  |  |
| 113) knowentyes>=0.5 6243 2451 yes (0.3925997 0.6074003) \* | | | | | | | | | |  |
| 57) discent>=0.5 1330 325 yes (0.2443609 0.7556391) \* | | | | | | | | |  |  |
| 29) age< 32.5 14419 4719 yes (0.3272765 0.6727235) \* | | | | | | | |  |  |  |
| 15) opportyes>=0.5 46014 12436 yes (0.2702656 0.7297344) \* |  |  |  |  |  |  |  |  |  |  |

Appendix 2: TEA Full Recursive Partitioning Model

n= 217164

node), split, n, loss, yval, (yprob)

\* denotes terminal node

1) root 217164 108582 no (0.50000000 0.50000000)

2) suskillyes< 0.5 81203 20567 no (0.74672118 0.25327882)

4) knowentyes< 0.5 58026 11464 no (0.80243339 0.19756661)

8) discent< 0.5 56473 10593 no (0.81242364 0.18757636)

16) busang< 0.5 54941 9896 no (0.81987951 0.18012049) \*

17) busang>=0.5 1532 697 no (0.54503916 0.45496084)

34) age>=53.5 320 65 no (0.79687500 0.20312500) \*

35) age< 53.5 1212 580 yes (0.47854785 0.52145215)

70) hhsize< 2.5 188 53 no (0.71808511 0.28191489) \*

71) hhsize>=2.5 1024 445 yes (0.43457031 0.56542969) \*

9) discent>=0.5 1553 682 yes (0.43915003 0.56084997)

18) age>=59.5 107 8 no (0.92523364 0.07476636) \*

19) age< 59.5 1446 583 yes (0.40318119 0.59681881)

38) GEMHHINC68100< 0.5 898 420 yes (0.46770601 0.53229399)

76) GEMHHINC3467< 0.5 390 169 no (0.56666667 0.43333333) \*

77) GEMHHINC3467>=0.5 508 199 yes (0.39173228 0.60826772)

154) opportyes< 0.5 282 136 yes (0.48226950 0.51773050)

308) age< 31.5 69 11 no (0.84057971 0.15942029) \*

309) age>=31.5 213 78 yes (0.36619718 0.63380282) \*

155) opportyes>=0.5 226 63 yes (0.27876106 0.72123894) \*

39) GEMHHINC68100>=0.5 548 163 yes (0.29744526 0.70255474) \*

5) knowentyes>=0.5 23177 9103 no (0.60723994 0.39276006)

10) busang< 0.5 21119 7899 no (0.62597661 0.37402339)

20) opportyes< 0.5 10285 3321 no (0.67710258 0.32289742)

40) age>=55.5 1761 361 no (0.79500284 0.20499716) \*

41) age< 55.5 8524 2960 no (0.65274519 0.34725481)

82) discent< 0.5 8181 2749 no (0.66397751 0.33602249) \*

83) discent>=0.5 343 132 yes (0.38483965 0.61516035) \*

21) opportyes>=0.5 10834 4578 no (0.57744139 0.42255861)

42) age>=62.5 368 60 no (0.83695652 0.16304348) \*

43) age< 62.5 10466 4518 no (0.56831645 0.43168355)

86) discent< 0.5 10011 4230 no (0.57746479 0.42253521)

172) gender2>=0.5 4843 1867 no (0.61449515 0.38550485)

344) hhsize< 5.5 4191 1538 no (0.63302314 0.36697686) \*

345) hhsize>=5.5 652 323 yes (0.49539877 0.50460123)

690) age< 22.5 73 5 no (0.93150685 0.06849315) \*

691) age>=22.5 579 255 yes (0.44041451 0.55958549)

1382) age>=51.5 37 0 no (1.00000000 0.00000000) \*

1383) age< 51.5 542 218 yes (0.40221402 0.59778598)

2766) UNEDUC5>=0.5 75 22 no (0.70666667 0.29333333) \*

2767) UNEDUC5< 0.5 467 165 yes (0.35331906 0.64668094) \*

173) gender2< 0.5 5168 2363 no (0.54276316 0.45723684)

346) GEMHHINC68100< 0.5 3036 1283 no (0.57740448 0.42259552) \*

347) GEMHHINC68100>=0.5 2132 1052 yes (0.49343340 0.50656660)

694) age< 22.5 174 54 no (0.68965517 0.31034483) \*

695) age>=22.5 1958 932 yes (0.47599591 0.52400409)

1390) UNEDUC5>=0.5 768 348 no (0.54687500 0.45312500)

2780) age>=23.5 715 308 no (0.56923077 0.43076923) \*

2781) age< 23.5 53 13 yes (0.24528302 0.75471698) \*

1391) UNEDUC5< 0.5 1190 512 yes (0.43025210 0.56974790)

2782) age>=55.5 85 29 no (0.65882353 0.34117647) \*

2783) age< 55.5 1105 456 yes (0.41266968 0.58733032) \*

87) discent>=0.5 455 167 yes (0.36703297 0.63296703) \*

11) busang>=0.5 2058 854 yes (0.41496599 0.58503401)

22) age>=49.5 516 202 no (0.60852713 0.39147287) \*

23) age< 49.5 1542 540 yes (0.35019455 0.64980545)

46) gender2>=0.5 515 251 yes (0.48737864 0.51262136)

92) UNEDUC4< 0.5 371 162 no (0.56334232 0.43665768) \*

93) UNEDUC4>=0.5 144 42 yes (0.29166667 0.70833333) \*

47) gender2< 0.5 1027 289 yes (0.28140214 0.71859786) \*

3) suskillyes>=0.5 135961 47946 yes (0.35264524 0.64735476)

6) knowentyes< 0.5 50205 23559 yes (0.46925605 0.53074395)

12) age>=48.5 15643 6670 no (0.57361120 0.42638880)

24) opportyes< 0.5 9463 3458 no (0.63457677 0.36542323)

48) age>=58.5 3664 992 no (0.72925764 0.27074236) \*

49) age< 58.5 5799 2466 no (0.57475427 0.42524573)

98) discent< 0.5 5413 2222 no (0.58950674 0.41049326)

196) fearfailyes>=0.5 1816 609 no (0.66464758 0.33535242) \*

197) fearfailyes< 0.5 3597 1613 no (0.55157075 0.44842925)

394) busang< 0.5 3388 1481 no (0.56286895 0.43713105) \*

395) busang>=0.5 209 77 yes (0.36842105 0.63157895) \*

99) discent>=0.5 386 142 yes (0.36787565 0.63212435) \*

25) opportyes>=0.5 6180 2968 yes (0.48025890 0.51974110)

50) age>=71.5 319 58 no (0.81818182 0.18181818) \*

51) age< 71.5 5861 2707 yes (0.46186658 0.53813342)

102) fearfailyes>=0.5 1311 553 no (0.57818459 0.42181541)

204) hhsize< 6.5 1228 493 no (0.59853420 0.40146580) \*

205) hhsize>=6.5 83 23 yes (0.27710843 0.72289157) \*

103) fearfailyes< 0.5 4550 1949 yes (0.42835165 0.57164835)

206) hhsize< 2.5 1856 889 yes (0.47898707 0.52101293)

412) age>=53.5 1429 711 no (0.50244927 0.49755073)

824) UNEDUC4< 0.5 1138 528 no (0.53602812 0.46397188)

1648) age< 70.5 1087 489 no (0.55013799 0.44986201) \*

1649) age>=70.5 51 12 yes (0.23529412 0.76470588) \*

825) UNEDUC4>=0.5 291 108 yes (0.37113402 0.62886598) \*

413) age< 53.5 427 171 yes (0.40046838 0.59953162) \*

207) hhsize>=2.5 2694 1060 yes (0.39346696 0.60653304)

414) age>=67.5 53 10 no (0.81132075 0.18867925) \*

415) age< 67.5 2641 1017 yes (0.38508141 0.61491859) \*

13) age< 48.5 34562 14586 yes (0.42202419 0.57797581)

26) opportyes< 0.5 16685 8113 yes (0.48624513 0.51375487)

52) fearfailyes>=0.5 6675 2959 no (0.55670412 0.44329588)

104) discent< 0.5 6291 2696 no (0.57145128 0.42854872)

208) hhsize< 3.5 2524 936 no (0.62916006 0.37083994) \*

209) hhsize>=3.5 3767 1760 no (0.53278471 0.46721529)

418) age< 20.5 234 64 no (0.72649573 0.27350427) \*

419) age>=20.5 3533 1696 no (0.51995471 0.48004529)

838) age>=38.5 1260 502 no (0.60158730 0.39841270) \*

839) age< 38.5 2273 1079 yes (0.47470304 0.52529696)

1678) gender2>=0.5 1152 547 no (0.52517361 0.47482639)

3356) UNEDUC4>=0.5 152 37 no (0.75657895 0.24342105) \*

3357) UNEDUC4< 0.5 1000 490 yes (0.49000000 0.51000000)

6714) hhsize< 4.5 494 213 no (0.56882591 0.43117409) \*

6715) hhsize>=4.5 506 209 yes (0.41304348 0.58695652)

13430) UNEDUC3>=0.5 160 67 no (0.58125000 0.41875000) \*

13431) UNEDUC3< 0.5 346 116 yes (0.33526012 0.66473988) \*

1679) gender2< 0.5 1121 474 yes (0.42283675 0.57716325) \*

105) discent>=0.5 384 121 yes (0.31510417 0.68489583) \*

53) fearfailyes< 0.5 10010 4397 yes (0.43926074 0.56073926)

106) discent< 0.5 9311 4236 yes (0.45494576 0.54505424)

212) busang< 0.5 8817 4087 yes (0.46353635 0.53646365)

424) age>=40.5 2093 998 no (0.52317248 0.47682752)

848) gender2>=0.5 915 381 no (0.58360656 0.41639344) \*

849) gender2< 0.5 1178 561 yes (0.47623090 0.52376910)

1698) UNEDUC1>=0.5 77 22 no (0.71428571 0.28571429) \*

1699) UNEDUC1< 0.5 1101 506 yes (0.45958220 0.54041780)

3398) nbgoodcyes< 0.5 459 226 no (0.50762527 0.49237473)

6796) age>=42.5 317 129 no (0.59305994 0.40694006) \*

6797) age< 42.5 142 45 yes (0.31690141 0.68309859) \*

3399) nbgoodcyes>=0.5 642 273 yes (0.42523364 0.57476636)

6798) GEMHHINC3467>=0.5 139 54 no (0.61151079 0.38848921) \*

6799) GEMHHINC3467< 0.5 503 188 yes (0.37375746 0.62624254) \*

425) age< 40.5 6724 2992 yes (0.44497323 0.55502677)

850) age< 20.5 629 289 no (0.54054054 0.45945946)

1700) easystartyes< 0.5 396 155 no (0.60858586 0.39141414) \*

1701) easystartyes>=0.5 233 99 yes (0.42489270 0.57510730) \*

851) age>=20.5 6095 2652 yes (0.43511075 0.56488925) \*

213) busang>=0.5 494 149 yes (0.30161943 0.69838057) \*

107) discent>=0.5 699 161 yes (0.23032904 0.76967096) \*

27) opportyes>=0.5 17877 6473 yes (0.36208536 0.63791464) \*

7) knowentyes>=0.5 85756 24387 yes (0.28437660 0.71562340)

14) fearfailyes>=0.5 24831 8688 yes (0.34988522 0.65011478)

28) busang< 0.5 21725 8010 yes (0.36869965 0.63130035)

56) age>=55.5 1913 943 no (0.50705698 0.49294302)

112) opportyes< 0.5 874 375 no (0.57093822 0.42906178)

224) UNEDUC5< 0.5 633 242 no (0.61769352 0.38230648) \*

225) UNEDUC5>=0.5 241 108 yes (0.44813278 0.55186722) \*

113) opportyes>=0.5 1039 471 yes (0.45332050 0.54667950) \*

57) age< 55.5 19812 7040 yes (0.35534020 0.64465980) \*

29) busang>=0.5 3106 678 yes (0.21828719 0.78171281) \*

15) fearfailyes< 0.5 60925 15699 yes (0.25767747 0.74232253) \*

References

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access, 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Ardichvili, A., Cardozo, R. and Ray, S. (2003). A theory of entrepreneurial opportunity identification and development, Journal of Business Venturing, 18:1, pp105-123. https://doi.org/10.1016/S0883-9026(01)00068-4.

Ahmad, S. Z., Xavier, S. R., & Abu Bakar, A. R. (2014). Examining entrepreneurial intention through cognitive approach using Malaysia GEM data. Journal of Organizational Change Management, 27(3), 449–464.

Ajzen, I. (1991). The theory of planned behavior. Orgnizational Behavior and Human Decision Processes, 50, 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Allen, I., Elam, A., Langowitz, N. & Dean, M. (2008). Global Entrepreneurship Monitor, 2007 Report on Women and Entrepreneurship. Babson College & London Business School: Kauffman Foundation

Aldrich, H. E., & Zimmer, C. (1986). Entrepreneurship through social networks. In C. Zimmer (Ed.), The art and science of entrepreneurship (pp. 3–23). Cambridge, MA: Ballinger.

Amani, F. A., & Fadlalla, A. M. (2017). Data mining applications in accounting: A review of the literature and organizing framework. International Journal of Accounting Information Systems, 24, 32–58. https://doi.org/10.1016/j.accinf.2016.12.004

Amini Sedeh, A., Beck, J. and Forghani Bajestani, M. (2020). Perceptual versus institutional determinants of entrepreneurial entry, Journal of Small Business and Enterprise Development, Vol. 27 No. 2, pp. 329-346. https://doi.org/10.1108/JSBED-07-2019-0245

Antretter, T., Blohm, I., Grichnik, D., & Wincent, J. (2019). Predicting new venture survival: A Twitter-based machine learning approach to measuring online legitimacy. Journal of Business Venturing Insights, 11(September 2018), e00109. <https://doi.org/10.1016/j.jbvi.2018.e00109>

Ardichvili, A., Cardozo, R., & Ray, S. (2003). A theory of entrepreneurial opportunity identification and development. Journal of Business Venturing, 18(1), 105–123

Arenius, P., & Minniti, M. (2005). Perceptual variables and nascent entrepreneurship. Small Business Economics, 24(3), 233–247. https://doi.org/10.1007/s11187-005-1984-x

Arin, K. P., Huang, V. Z., Minniti, M., Nandialath, A. M., & Reich, O. F. M. (2015). Revisiting the Determinants of Entrepreneurship: A Bayesian Approach. Journal of Management, 41(2), 607–631. https://doi.org/10.1177/0149206314558488

Athey, S., & Imbens, G. W. (2019). Machine learning methods economists should know about. ArXiv, (September).

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. Annals of Statistics, 47(2), 1179–1203. https://doi.org/10.1214/18-AOS1709

Audretsch D.B., Erdem D.K. (2005) Factors Affecting Entrepreneurial Activity: Literature Review. In: Alvarez S.A., Agarwal R., Sorenson O. (eds) Handbook of Entrepreneurship Research. International Handbook Series on Entrepreneurship, vol 2. Springer, Boston, MA. https://doi.org/10.1007/0-387-23622-8\_7

Autio, E., Keeley, R H., Klofsten, M., Parker, G. C., & Hay, M. (2001). Entrepreneurial intent among students in Scandinavia and in the USA. Enterprise and Innovation Management Studies, 2, 145-160.

Bandura, A. (1989). Human agency in social-cognitive theory. American Psychologist, 44, 1175–1184.

Bandura, A. (1997). Self-efficacy: The exercise of control. New York, NY: Worth Publishers

Bareinboim, E., & Tian, J. (2015). Recovering causal effects from selection bias. Proceedings of the National Conference on Artificial Intelligence, 5(December 2014), 3475–3481.

Beynon, M. J., Jones, P., & Pickernell, D. (2016). Country-based comparison analysis using fsQCA investigating entrepreneurial attitudes and activity. Journal of Business Research, 69(4), 1271–1276. https://doi.org/10.1016/j.jbusres.2015.10.091

Beynon, M. J., Jones, P., & Pickernell, D. (2018a). Entrepreneurial climate and self-perceptions about entrepreneurship: a country comparison using fsQCA with dual outcomes. Journal of Business Research, 89(June 2017), 418–428. https://doi.org/10.1016/j.jbusres.2017.12.014

Beynon, M. J., Jones, P., & Pickernell, D. (2019). Country-level entrepreneurial attitudes and activity through the years : A panel data analysis using fsQCA. Journal of Business Research, (November), 1–13. https://doi.org/10.1016/j.jbusres.2019.11.021

Bird, B. (1988). Implementing Entrepreneurial Ideas: The Case for Intention. Academy of Management Review, 13(3), 442–453. <https://doi.org/10.5465/amr.1988.4306970>

Blanchflower, D. (2004). Self-employment: more may not be better. Swedish Economic Policy Review, 11(2), 15–73.

Boden, R.J.Jr. (1999), Gender inequality in wage earnings and female self-employment selection, Journal of Socio-Economics, Vol. 28 No. 3, pp. 351-364

Boudreaux, C. J., Nikolaev, B. N., & Klein, P. (2019). Socio-cognitive traits and entrepreneurship: The moderating role of economic institutions. Journal of Business Venturing, 34(1), 178–196. https://doi.org/10.1016/j.jbusvent.2018.08.003

Braun, M. T., Kuljanin, G., & DeShon, R. P. (2018). Special Considerations for the Acquisition and Wrangling of Big Data. Organizational Research Methods, 21(3), 633–659. https://doi.org/10.1177/1094428117690235

Breiman, L. (2001). Statistical Modeling: The Two Cultures. Statistical Science, 16(3), 199–231. <https://doi.org/10.2307/2676681>

Budescu, D. (1993). Dominance Analysis: A new approach to the problem of relative importance of predictors in multiple regression. Psychological Bulletin, 114(3), 542–551.

Cacciotti, G., & Hayton, J. C. (2015). Fear and entrepreneurship: A review and research agenda. International Journal of Management Reviews, 17(2), 165–190. <https://doi.org/10.1111/ijmr.12052>

Carter, S., Anderson, S., & Shaw, E. (2001). Women’s business ownership: a review of the academic, popular and internet literature, Report to the Small Business Service, RR002/01

Casson M. and Giusta, M.D. (2007) Entrepreneurship and Social Capital: Analysing the Impact of Social Networks on Entrepreneurial Activity from a Rational Action Perspective. International Small Business Journal, 25(3):220-244. doi:10.1177/0266242607076524

Celbiş, M. G. (2021). A Machine Learning Approach to Rural Entrepreneurship. Papers in Regional Science. https://doi.org/10.1111/pirs.12595

Chatfield, C. (1995). Model Uncertainty, Data Mining and Statistical Inference. Journal of the Royal Statistical Society. Series A (Statistics in Society), 158(3), 419. <https://doi.org/10.2307/2983440>

Chell, E. (2013). Review of skill and the entrepreneurial process. International Journal of Entrepreneurial Behaviour & Research, 19(1), 6–31

Chen, C.C., Greene, P.G. and Crick, A. (1998), “Does entrepreneurial self-efficacy distinguish entrepreneurs from managers?”, Journal of Business Venturing, Vol. 13 No. 4, pp. 295-316.

Chowdhury, S & Endres, M. (2005). Gender Difference and the Formation of Entrepreneurial Self-Efficacy (2005). University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship

Coad, A., & Srhoj, S. (2019). Catching Gazelles with a Lasso: Big data techniques for the prediction of high-growth firms. Small Business Economics. https://doi.org/10.1007/s11187-019-00203-3

Coduras, A., Clemente, J. A., & Ruiz, J. (2016). A novel application of fuzzy-set qualitative comparative analysis to GEM data. Journal of Business Research, 69(4), 1265–1270. https://doi.org/10.1016/j.jbusres.2015.10.090

Courtney, C., Dutta, S., & Li, Y. (2017). Resolving Information Asymmetry: Signaling, Endorsement, and Crowdfunding Success. Entrepreneurship: Theory and Practice, 41(2), 265–290. https://doi.org/10.1111/etap.12267

Coussement, K., & De Bock, K. W. (2013). Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning. Journal of Business Research, 66(9), 1629–1636. https://doi.org/10.1016/j.jbusres.2012.12.008

Davidsson, P. (2015), “Entrepreneurial opportunities and the entrepreneurship nexus: a re-conceptualization”, Journal of Business Venturing, Vol. 30 No. 5, pp. 674-695

Delen, D., & Zolbanin, H. M. (2018). The analytics paradigm in business research. Journal of Business Research, 90(May), 186–195. <https://doi.org/10.1016/j.jbusres.2018.05.013>

Díaz-García, M. C., & Jiménez-Moreno, J. (2010). Entrepreneurial intention: the role of gender. International Entrepreneurship and Management Journal, 6(3), 261–283.

Dincer, H., Hacioglu, U., Tatoglu, E., & Delen, D. (2019). Developing a hybrid analytics approach to measure the efficiency of deposit banks. Journal of Business Research, 104(July), 131–145. https://doi.org/10.1016/j.jbusres.2019.06.035

Doran, D., Schulz, S., & Besold, T. R. (2018). What does explainable AI really mean? A new conceptualization of perspectives. CEUR Workshop Proceedings, 2071.

Duchessi, P., & Lauría, E. J. M. (2013). Decision tree models for profiling ski resorts’ promotional and advertising strategies and the impact on sales. Expert Systems with Applications, 40(15), 5822–5829. https://doi.org/10.1016/j.eswa.2013.05.017

Dybowski, R. (2020). Interpretable machine learning as a tool for scientific discovery in chemistry. New Journal of Chemistry, 44(48), 20914–20920. https://doi.org/10.1039/d0nj02592e

Efron, B. (2020). Prediction, Estimation, and Attribution. Journal of the American Statistical Association, 115(530), 636–655. https://doi.org/10.1080/01621459.2020.1762613

Engle, R. L., Schlaegel, C., & Dimitriadi, N. (2011). Institutions and entrepreneurial intent: A cross-country study. Journal of Developmental Entrepreneurship (Vol. 16). https://doi.org/10.1142/S1084946711001811

Fayolle, A., & Liñán, F. (2014). The future of research on entrepreneurial intentions. Journal of Business Research, 67(5), 663–666. <https://doi.org/10.1016/j.jbusres.2013.11.024>

Fernández-Serrano, J., Liñán, F., & Sántos, F. J. (2009). Cognitive aspects of potential entrepreneurs in southern and northern Europe: an analysis using GEM-data. Revista De Economica Mundial, 23(1), 151– 178.

Gatewood, E., Shaver, K., Powers, J., & Gartner, W. (2002). Entrepreneurial expectancy, task, effort and performance. Entrepreneurship Theory and Practice, 27(Winter), 187–206

GEM. (2018). Global Entrepreneurship Monitor 2018/2019 Global Report. Global Entrepreneurship Monitor. Retrieved from http://www.gemconsortium.org/report/50012

George, N. M., Parida, N., V., Lahti, T. & Wincent, J. (2016) A systematic literature review of entrepreneurial opportunity recognition: insights on influencing factors. International Entrepreneurship and Management Journal, 12, 309–350. https://doi.org/10.1007/s11365-014-0347-y

Gepp, A., Kumar, K., & Bhattacharya, S. (2010). Business failure prediction using decision trees. Journal of Forecasting, 29(6), 536–555. https://doi.org/10.1002/for.1153

Graham, B., Bond, R., Quinn, M., & Mulvenna, M. (2018). Using Data Mining to Predict Hospital Admissions from the Emergency Department. IEEE Access, 6. https://doi.org/10.1109/ACCESS.2018.2808843

Guo, R., Cheng, L., Li, J., Hahn, P. R., & Liu, H. (2020). A Survey of Learning Causality with Data. ACM Computing Surveys, 53(4), 1–37. https://doi.org/10.1145/3397269

Hair, J. F., & Sarstedt, M. (2021). Data, measurement, and causal inferences in machine learning: opportunities and challenges for marketing. Journal of Marketing Theory and Practice, 00(00), 1–13. https://doi.org/10.1080/10696679.2020.1860683

Hakanen, J. J., Bakker, A. B., & Turunen, J. (2021). The relative importance of various job resources for work engagement: A concurrent and follow-up dominance analysis. BRQ Business Research Quarterly. <https://doi.org/10.1177/23409444211012419>

Haus, I., Steinmetz, H., Isidor, R., & Kabst, R. (2013). Gender effects and entrepreneurial intention: a metaanalytical structural equation model. International Journal of Gender and Entrepreneurship, 5(2), 130– 156.

Haworth, J. M., Brearley, S. A., & Chell, E. (1991). Categorization of business owners using neural networks. Entrepreneurship and Regional Development, 3(3), 221–235. https://doi.org/10.1080/08985629100000015

He, H., Bai, Y., & Xiao, X. (2020). HOW PAST FAILURE PREDICTS SUBSEQUENT ENTREPRENEURIAL INTENTION: A COMPARATIVE STUDY OF MAINLAND CHINA AND TAIWAN.

Hoang, H., & Yi, A. (2015). Network-based research in entrepreneurship: A decade in review. Foundations and Trends in Entrepreneurship, 11(1), 1–54

Hünermund, P., Kaminski, J., & Schmitt, C. (2021). Causal Machine Learning and Business Decision Making \*, 1–76.

Huselid, M. A. (2018). The science and practice of workforce analytics: Introduction to the HRM special issue. Human Resource Management, 57(3), 679–684. https://doi.org/10.1002/hrm.21916

Issue, S., Bux, S. R., & Honglin, Y. (2016). Examining entrepreneurial intention through A cognitive approach in China by using gem data. Science International, 28(2), 1681–1687.

Kacperczyk, A. J. (2013). Social influence and entrepreneurship: The effect of university peers on entrepreneurial entry. Organization Science, 24(3), 664–683. doi:10.1287/orsc.1120.0773

[Karimi](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Karimi%2C+Saeid), S., [Biemans](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Biemans%2C+Harm+J+A), H., [Lans](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Lans%2C+Thomas), T., [Chizari](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Chizari%2C+Mohammad), M., & Mulder, M. (2016). The Impact of Entrepreneurship Education: A Study of Iranian Students' Entrepreneurial Intentions and Opportunity Identification. Journal of Small Business Management, 54(1):187–20

Kautonen, T., Down, S., & Minniti, M. (2014). Ageing and entrepreneurial preferences. Small Business Economics. Volume 42, Issue 3. 579-594. DOI: 10.1007/s11187-013-9489-5.

Kautonen, T., van Gelderen, M., & Fink, M. (2015). Robustness of the theory of planned behavior in predicting entrepreneurial intentions and actions. Entrepreneurship: Theory and Practice, 39(3), 655–674. https://doi.org/10.1111/etap.12056

Khan, A.M., Arafat, M.Y., Raushan, M.A., Khan, M.S., Dwivedi, A.K., Khan, M.M., and Saleem, I. (2019) Role of intellectual capital in augmenting the start-up intentions of Indians—an analysis using GEM data, Journal of Global Entrepreneurship Research 9:25 https://doi.org/10.1186/s40497-019-0146-6

Kim, H. J., Kim, T. S., & Sohn, S. Y. (2020). Recommendation of startups as technology cooperation candidates from the perspectives of similarity and potential: A deep learning approach. Decision Support Systems, 130(December 2019), 113229. https://doi.org/10.1016/j.dss.2019.113229

Koellinger, P., Minniti, M., & Schade, C. (2013). Gender Differences in Entrepreneurial Propensity. Oxford Bulletin of Economics and Statistics, 75(2), 213–234. https://doi.org/10.1111/j.1468-0084.2011.00689.x

Kolvereid, L., & Isaksen, E. (2006). New business start-up and subsequent entry into self-employment. Journal of Business Venturing, 21(6), 866–885. https://doi.org/10.1016/j.jbusvent.2005.06.008

Kourilsky, M. & Walstad, M. (1998). Entrepreneurship and female youth: Knowledge, attitudes, gender differences and educational practices. Journal of Business Venturing, 13, 77–88.

Krueger, N. F. (2003). The cognitive psychology of entrepreneurship. In Z. J. Acs & D. B. Audretsch (Eds.), Handbook of entrepreneurship research: An interdisciplinary survey and introduction (pp. 105–140). London: Kluwer

Krueger, N. (1993). The Impact of Prior Entrepreneurial Exposure on Perceptions of New Venture Feasibility and Desirability. Entrepreneurship Theory and Practice, 18(1)5-21

Krueger, N. F., & Carsrud, A. L. (1993). Entrepreneurial intentions: Applying the theory of planned behaviour. Entrepreneurship and Regional Development, 5(4), 315–330. <https://doi.org/10.1080/08985629300000020>

Krueger, N., & Dickson, P. R. (1994). How believing in ourselves increases risk taking: perceived self-efficacy and opportunity recognition. Decision Sciences, 25(3), 385–400

Krueger, N. F., Reilly, M. D., & Carsrud, A. L. (2000). Competing models of entrepreneurial intentions. Journal of Business Venturing, 15(5), 411–432. https://doi.org/10.1016/S0883-9026(98)00033-0

Kuckertz, A., Berger, E.S.C. & Allmendinger, M. (2015). What Drives Entrepreneurship? A Configurational Analysis of the Determinants of Entrepreneurship in Innovation-Driven Economies. Die Betriebswirtschaft/Business Administration Review, 75,4, pp. 273-288, Available at SSRN: https://ssrn.com/abstract=2554944

Kuhn, M., & Johnson, K. (2013). Applied predictive modelling. London: Springer.

Kuhn, M. (2017). The Caret Package. Retrieved October 1, 2017, from https://topepo.github.io/caret/

Kumar, M., Kee, F. T., & Charles, V. (2010). Comparative evaluation of critical factors in delivering service quality of banks: An application of dominance analysis in modified SERVQUAL model. International Journal of Quality and Reliability Management, 27(3), 352–378. https://doi.org/10.1108/02656711011023320

Leazar, E. P. (2005). Entrepreneurship. Journal of Labor Economics, University of Chicago Press, 23(4), 649–680.

Lévesque, M., & Minniti, M. (2006). The effect of aging on entrepreneurial behavior. Journal of Business Venturing, 21(2), 177–194

Lévesque, M., Obschonka, M., & Nambisan, S. (2020). Pursuing Impactful Entrepreneurship Research Using Artificial Intelligence. Entrepreneurship Theory and Practice, (April 2020), forthcoming.

Levie, J., & Autio, E. (2008). A theoretical grounding and test of the GEM model. Small Business Economics, 31(3), 235-263

Liñán, F., & Fayolle, A. (2015). A systematic literature review on entrepreneurial intentions: citation, thematic analyses, and research agenda. International Entrepreneurship and Management Journal, 11(4), 907–933. https://doi.org/10.1007/s11365-015-0356-5

Liñán, F., Urbano, D., & Guerrero, M. (2011). Regional variations in entrepreneurial cognitions: Start-up intentions of university students in Spain. Entrepreneurship and Regional Development, 23(3–4), 187–215. https://doi.org/10.1080/08985620903233929

McKenzie, D., & Sansone, D. (2019). Predicting entrepreneurial success is hard: Evidence from a business plan competition in Nigeria. Journal of Development Economics, 141(April 2018), 102369. https://doi.org/10.1016/j.jdeveco.2019.07.002

Montebruno, P., Bennett, R. J., Smith, H., & Lieshout, C. van. (2020). Machine learning classification of entrepreneurs in British historical census data. Information Processing and Management, 57(3), 102210. https://doi.org/10.1016/j.ipm.2020.102210

Moriano, J. A., Gorgievski, M., Laguna, M., Stephan, U., & Zarafshani, K. (2012). A Cross-Cultural Approach to Understanding Entrepreneurial Intention. Journal of Career Development, 39(2), 162–185. https://doi.org/10.1177/0894845310384481

Murphy, K., Shleifer, A., & Vishny, R. (1991). The allocation of talent: implications for growth. Quarterly Journal of Economics, 106, 503–530.

Obschonka, M., & Audretsch, D. B. (2020). Artificial intelligence and big data in entrepreneurship: a new era has begun. Small Business Economics, 55, 529–539. https://doi.org/10.1007/s11187-019-00202-4

Obschonka, M., Fisch, C., & Boyd, R. (2017). Using digital footprints in entrepreneurship research: A Twitter-based personality analysis of superstar entrepreneurs and managers. Journal of Business Venturing Insights, 8(February), 13–23. https://doi.org/10.1016/j.jbvi.2017.05.005

Olanrewaju, A. S. T., Hossain, M. A., Whiteside, N., & Mercieca, P. (2020). Social media and entrepreneurship research: A literature review. International Journal of Information Management, 50(November 2018), 90–110. https://doi.org/10.1016/j.ijinfomgt.2019.05.011

Ozaralli, N. and Rivenburgh, N.K. (2016). Entrepreneurial intention: antecedents to entrepreneurial behavior in the U.S.A. and Turkey. Journal of Global Entrepreneuship Research 6, 3. <https://doi.org/10.1186/s40497-016-0047-x>

Parker, S. C. (2009). The Economics of Entrepreneurship. Cambridge: Cambridge University Press.

Passaro, R., Quinto, I. and Thomas, A. (2018). The impact of higher education on entrepreneurial intention and human capital, Journal of Intellectual Capital, Vol. 19 Issue: 1, pp.135-156, https://doi.org/10.1108/JIC-04-2017-0056

Pathak, S., & Muralidharan, E. (2021). Consequences of cross-cultural differences in perceived well-being for entrepreneurship. Journal of Business Research, 122(October 2020), 582–596. https://doi.org/10.1016/j.jbusres.2020.09.034

Prüfer, J., & Prüfer, P. (2020). Data science for entrepreneurship research: studying demand dynamics for entrepreneurial skills in the Netherlands. Small Business Economics, 55(3), 651–672. https://doi.org/10.1007/s11187-019-00208-y

Puriwat, W., & Tripopsakul, S. (2015). The comparative study of attitudes toward entrepreneurial intention between ASEAN and EUROPE : An analysis. International Journal of Applied Business and Economic Research, 13(6), 4485–4494.

Reynolds, P. D., Bygrave, B., & Hay, M. (2003). Global entrepreneurship monitor report. Kansas: E. M. Kauffmann Foundation.

Roper, S., & Scott, J. (2009). Perceived financial barriers and the start up decision. An econometric analysis of gender differences using GEM data. International Small Business Journal, 27(2), 149–171.

Sabahi, S., & Parast, M. M. (2020). The impact of entrepreneurship orientation on project performance: A machine learning approach. International Journal of Production Economics, 226(December 2019), 107621. <https://doi.org/10.1016/j.ijpe.2020.107621>

Sández-Escobedo, M., Díaz-Casero, J. C., Díaz-Aunión, A. M., & Hernández-Mogollón, R. (2014). Gender analysis of entrepreneurial intentions as a function of economic development across three groups of countries. International Entrepreneurship and Management Journal, 10(4), 747–765

Santos, S. C., Caetano, A., Costa, S. F., Rueff Lopes, R., Silva, A. J., & Neumeyer, X. (2020). Uncovering the affective turmoil during opportunity recognition and exploitation: A nonlinear approach. Journal of Business Venturing Insights, 14(June). https://doi.org/10.1016/j.jbvi.2020.e00184

Saradhi, V. V., & Palshikar, G. K. (2011). Employee churn prediction. Expert Systems With Applications, 38(3), 1999–2006. https://doi.org/10.1016/j.eswa.2010.07.134

Scherer, R.F., Adams, J.S., Carley, S.S. and Wiebe, F.A. (1989). Role model performance effects on development of entrepreneurial career preference, Entrepreneurship Theory & Practice, Vol. 13 No. 3, pp. 53-71.

Schlaegel C, Koenig M. (2014). Determinants of Entrepreneurial Intent: A Meta–Analytic Test and Integration of Competing Models. Entrepreneurship Theory and Practice, 38(2):291-332. doi:10.1111/etap.12087

Schmutzler J, Andonova V, Diaz-Serrano L. (2019). How Context Shapes Entrepreneurial Self-Efficacy as a Driver of Entrepreneurial Intentions: A Multilevel Approach. Entrepreneurship Theory and Practice,. 43(5):880-920. doi:10.1177/1042258717753142

Schuelke-Leech BA., Barry B.L. (2016) Complexity of Textual Data in Entrepreneurship and Innovation Research. In: Berger E., Kuckertz A. (eds) Complexity in Entrepreneurship, Innovation and Technology Research. FGF Studies in Small Business and Entrepreneurship. Springer, Cham. https://doi.org/10.1007/978-3-319-27108-8\_22

Schwab, A., & Zhang, Z. (2019). A New Methodological Frontier in Entrepreneurship Research: Big Data Studies. Entrepreneurship: Theory and Practice, 43(5), 843–854. https://doi.org/10.1177/1042258718760841

SEQUEIRA, J., MUELLER, S. L., & MCGEE, J. E. (2007). the Influence of Social Ties and Self-Efficacy in Forming Entrepreneurial Intentions and Motivating Nascent Behavior. Journal of Developmental Entrepreneurship, 12(03), 275–293. <https://doi.org/10.1142/s108494670700068x>

Shane, S., Locke, E. A., & Collins, C. J. (2003). Entrepreneurial motivation. Human Resource Management Review, 12, 257–279

Shapero, A. and Sokol, L., (1982). Social dimensions of entrepreneurship. In C. . A. Kent, D. L. SEXTON, & K. H. Vesper (Eds.), C. Kent, D. Sexton and K. Vesper, eds., The Encyclopedia of EntrepreneurshipThe Encyclopedia of Entrepreneurship (pp. 72–90). Prentice Hall.

Shapero, A. (1984). The entrepreneurial event. In C. A. Kent (Ed.), The environment for entrepreneurship (pp. 21–40). Lexington, KY: Lexington Books.

Shinnar, Rachel & Giacomin, Olivier & Janssen, Frank. (2012). Entrepreneurial Perceptions and Intentions: The Role of Gender and Culture. Entrepreneurship Theory and Practice. 36. 465-494. 10.1111/j.1540-6520.2012.00509.x.

Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems resh. MIS Quarterly, 35(3), 553–572.

Simon, M., Houghton, S.M. and Aquino, K. (2000), “Cognitive biases, risk perception, and venture formation: how individuals decide to start companies”, Journal of Business Venturing, Vol. 15 No. 2, pp. 113-134

Singh, J., Tanev, S., & Bailetti, T. (2020). *Using text analytics to discover business scaling research gaps*. Manchester: The International Society for Professional Innovation Management (ISPIM).

Sohn, S. Y., & Lee, A. S. (2013). Bayesian network analysis for the dynamic prediction of early stage entrepreneurial activity index. Expert Systems with Applications, 40(10), 4003–4009. https://doi.org/10.1016/j.eswa.2013.01.009

Stam, E.; Audretsch, D.; Meijaard, J. (2008). Renascent entrepreneurship. Journal of Evolutionary Economics, 18: 493–507.

Stel, A. Van, Carree, M., & Thurik, R. (2005). The effect of entrepreneurial activity on national economic growth. Small Business Economics, 24(3), 311–321. https://doi.org/10.1007/s11187-005-1996-6

Stephan, U., Hart, M. and Drews, C (2015). Understanding Motivations for Entrepreneurship: A Review of Recent Research Evidence, ERC Rapid Evidence Assessment paper February 2015

Tata, A., Martinez, D. L., Garcia, D., Oesch, A., & Brusoni, S. (2017). The psycholinguistics of entrepreneurship. Journal of Business Venturing Insights, 7(January), 38–44. <https://doi.org/10.1016/j.jbvi.2017.02.001>

Thereneau, T. M. and Atkinson, E. J. (2015). An introduction to recursive partitioning using the RPART routines. Mayo Foundation. Retrieved from https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf

Tsai, K.-H., Chang, H.-C., & Peng, C.-Y. (2016). Refining the linkage between perceived capability and entrepreneurial intention: Roles of perceived opportunity, fear of failure, and gender. International Entrepreneurship and Management Journal. https://doi.org/10.1007/s11365-016-0383-x

Thornton PH, Ribeiro-Soriano D, Urbano D. Socio-cultural factors and entrepreneurial activity: An overview. International Small Business Journal. 2011;29(2):105-118. doi:10.1177/0266242610391930

Tugade, M.M.; Fredrickson, B.L. (2004) Resilient individuals use positive emotions to bounce back from negative emotional experiences. Journal of Personality and Social Psychology, 86:320–333.

van Auken, H., Stephens, P., Fry, F., & Silva-Castan, J. (2006). Role model influences on entrepreneurial intentions: A comparison between USA and Mexico, The International Entrepreneurship and Management Journal, 11(2): 157-167

van der Sluis, J., van Praag, M., Vijverberg, W. (2005). Entrepreneurship selection and performance: A meta-analysis of the impact of education in developing economies. World Bank Economic Review, 2: 225-261.

Varamaki, E., Joensuu, S., Tornikoski, E., & Viljammaa, A. (2013). The development of entrepreneurial potential among higher education students. Journal of Small Business and Enterprise Development, 22(3), 563-589.

Verheul, I., Thurik, R., Grilo, I., & van der Zwan, P. (2012). Explaining preferences and actual involvement in self-employment: gender and the entrepreneurial personality. Journal of Economic Psychology, 33(2), 325–341.

Vossenberg, S. (2013). Women Entrepreneurship Promotion in Developing Countries: What explains the gender gap in entrepreneurship and how to close it? Working paper 2013/08. Maastricht School of Management

Walter, S. G., & Heinrichs, S. (2015). Who becomes an entrepreneur? A 30-years-review of individual-level research. Journal of Small Business and Enterprise Development, 22(2), 225-248. doi:http://dx.doi.org/10.1108/JSBED-09-2012-0106

Wang, F., Mack, E. A., & Maciewjewski, R. (2017). Analyzing Entrepreneurial Social Networks with Big Data. Annals of the American Association of Geographers, 107(1), 130–150. https://doi.org/10.1080/24694452.2016.1222263

Wang, W., Chen, W., Zhu, K., & Wang, H. (2020). Emphasizing the entrepreneur or the idea? The impact of text content emphasis on investment decisions in crowdfunding. Decision Support Systems, 136(September 2019), 113341. https://doi.org/10.1016/j.dss.2020.113341

Wei, Y., Lv, H., Chen, M., Wang, M., Heidari, A. A., Chen, H., & Li, C. (2020). Predicting Entrepreneurial Intention of Students: An Extreme Learning Machine with Gaussian Barebone Harris Hawks Optimizer. IEEE Access, 8, 76841–76855. https://doi.org/10.1109/ACCESS.2020.2982796

Williamson, A. J., Drencheva, A., & Battisti, M. (2020). Entrepreneurial disappointment: Let down and breaking down, a machine-learning study. Entrepreneurship Theory and Practice. https://doi.org/10.1177/1042258720964447

Wilson, F., Kickul, J., & Marlino, D. (2007). Gender, Entrepreneurial Self-Efficacy, and Entrepreneurial Career Intentions: Implications for Entrepreneurship Education. Education, 31(617), 387–407. Retrieved from <http://doi.wiley.com/10.1111/j.1540-6520.2007.00179.x>

Zellweger, T., Sieger, P., & Halter, F. (2011). Should I stay or should I go? Career choice intentions of students with family business background. Journal of Business Venturing, 26(5), 521–536.

Zhao, H., Hills, G. E., & Seibert, S. E. (2005). The mediating role of self-efficacy in the development of entrepreneurial intentions. Journal of Applied Psychology, 90(6), 1265–1272. <https://doi.org/10.1037/0021-9010.90.6.1265>

1. \* Corresponding author. E-Mail address: byron.graham@qub.ac.uk [↑](#footnote-ref-1)