QUANTIFYING FAMILY BUSINESS OWNERS' ATTITUDES TOWARDS SUCCESSION: DELVING DEEPER BY RANDOM FORESTS

Kılıç, A., Filizöz, B.

Arzu Kilic / Department of Business Administration, Faculty of Economics and Administrative Sciences, Ardahan University, Ardahan, Turkey. Email: arzukilic@ardahan.edu.tr.

Berrin Filizöz / Department of Business Administration, Faculty of Economics and Administrative Sciences, Cumhuriyet University, Sivas, Turkey. Email: berrinfilizoz@gmail.com.

Abstract

Family businesses are the lifeblood of the economic growth of the nations. However, a large gap exists about the application of machine-learning algorithms such as Random Forests (RF) to the quantification of patterns, drivers, and interactions in the succession process of family businesses. The primary aim and novelty of this study lie in the quantification of variable importance based on machine-learning algorithms, and the differences among the characteristics of family businesses, family employees, and family business owners (FBOs) for multivariate responses. For this reason, a field study was carried out in family businesses in Sivas and Ardahan provinces. The questionnaire form created by the researchers was used in this study. In this research, RF classification model was applied. RF classification models of 17 response variables were constructed as a function of 32 predictors.

Implications for Central European audience: Impacts of characteristics of FBOs, family businesses, and family employees on FBOs' willingness to transfer to successors and preferences about the successor's qualities were modelled. High-dimensional data were collected from 53 family business owners (FBOs) in two cities for a total of 49 variables. As a result, the domain of the FBOs' characteristics was found to have a more profound impact on both FBO's willingness to transfer to a successor and what successor's qualities were most valued than did the domains of the family business and employee characteristics.

Keywords: machine learning; small-sized enterprises; family business

JEL Classification: M10, M59

Introduction

Family businesses, in particular, small-scale family enterprises are the backbone of national economies. Micro-scale (local) socio-economic decisions are the drivers of macro-scale (national and regional) socio-economic changes (Van Praag et al., 2007). Globally, family businesses were on average estimated to constitute 70-to-90% of all business entities and create 60-to-90% of gross domestic product (GDP) annually (Allio, 2004; Blodgett et al., 2011; Tucker, 2011). In Turkey, 95% of all businesses are family businesses with a mean survival rate of 25 years 30, 12 and 3% of which have been transferred to the second, third and fourth generations, respectively (Tuncel, 2011; ASSBA, 2016). The succession process is one of the main driving forces behind the sustainability of the family business. The lack of strategical planning and management in the succession process hinders the long-term existence of family businesses (Venter et al., 2005).

Many factors influential on the successful transition of management and ownership to heirs or successors in family businesses have been studied extensively in related literature (Le Breton-Miller et al., 2004). In the literature of family business, succession has become one of the most significant themes (De Massis et al., 2008; Nordqvist & Melin, 2010; Sharma, 2004). Chua et al. (2003) pointed out that the sustenance and transition of family business were what the family business leaders were most concerned with. The reason for this relates not only to the necessity for every family firm to be eventually transferred to an heir, or a successor as an integral part of its life cycle but also to the challenges of the complex process of “emotional and financial adaptation and management and/or ownership transfer” for family firms (Laakkonen & Kansikas, 2011, pp. 984).

The most influential factors on the succession process were reported as the owners’ attitudes towards losing their authority, control over their successors and individuality identified with firms, their senses of potential conflicts with their successors, their fear of death, and their indecision about the choice of successors (Hacıbayramoğlu, 2014). The succession as a delicate process that requires special attention is usually characterized by uncertainty, tense mood, conflict, and reorientation. For example, new roles, and their restructuring are needed to be defined for all the stakeholders involved, such as family members, transferees, potential heirs, and non-family managers and employees. Once a successor is chosen and operationally active, the power structure needs to be restructured due to its change (Zehe, 2016). The success of the succession primarily depends on successors’ characteristics and attitudes, in particular, on their commitment and willingness (Chrisman et al., 1998; Sharma & Irving, 2005). Also, heirs’ or successors’ related education and training, on- and off-job experience and duration, entrance level, internal position, breadth of knowledge, communication ability, trustworthiness and age also play an essential role in the selection (Hacıbayramoğlu, 2014; Kılıç & Filizöz, 2019).

There exist various studies about the transfer of management and its management, heir or successor selection, and most sought heir or successor characteristics in related literature such as Cesaroni & Sentuti (2017), Chua et al. (1999), Devany (2006), Gomba & Kele (2016), Güller (2010), and Sharma & Rao (2000). An extensive literature review (e.g., Yu et al., 2012; Bau et al., 2013) indicated that the majority of the related studies had used questionnaire data, convenience sampling, small sample sizes, conceptual analyses and descriptions, and qualitative data analyses. In addition to the above common features, interactions among multiple factors, non-linear relationships, non-parametric nature of variables measured, high-dimensional data, a large number of missing values, and the presence of outliers and noise have been an integral part of such studies and research data (Grimmer, 2015). All these challenges may render machine-learning algorithms useful in accounting for the drivers and patterns of the succession process since machine learning is resistant to these challenges (Lazer et al., 2009). Machine learning can be described as the process of deriving decision rules from training by examples of known input-output behaviours and their application to the problems of prediction, classification, optimization, change detection, and pattern recognition (Jordan & Mitchell, 2015).

The primary aim and novelty of this study lie in the quantification of variable importance based on machine-learning algorithms, and the differences among the characteristics of family businesses, family employees, and family business owners (FBOs) for multivariate responses. To the best of our knowledge, there exists no study about the application of machine-learning approaches to the questionnaire data about FBOs' attitudes towards his/her heirs and successors. In particular, Random Forests (RF) is among the best of today's most powerful machine-learning methods and has not been implemented in this field thus far. To contribute to filling this gap in related literature, the objectives of this exploratory study were to (1) quantify the relative importance and sensitivity of explanatory variables in elucidating FBOs' attitudes in succession by using RF classification models; and (2) explore the benefits of adopting RF in the computational social studies.

1 Data and methodology

In the structured questionnaire surveys, 28 firms from the Ardahan city, and 25 firms from the Sivas city participated. The main fractions of the participants from both cities were as follows: 5.7% women, 35.9% in the age of 41-50, 45.2% in the age of ≥ 51, 37.7% high school graduates, 76% in favour of transferring to heir/successor, 55% those who did not determine the transfer time, and 40.8% those who planned on transferring within 1-10 years. In response to the questions about FBOs' attitudes in the process of succession preferred on-site training, monitoring, supervision, transfer of rights before and after death, and actively working after the transfer, respectively (Table A1). The participants were asked to reveal their preferences for the dichotomous questions about the successor qualities sought (as indicated by QMS in Table A1). Their most frequent responses were thus: preferred honesty, commitment to a family business, entrepreneurship, and ability to solve problems (Kılıç & Filizöz, 2019).

A total of 17 response variables concerning FBOs' attitudes in succession and a total of 32 predictors (9 business characteristics + 8 family employee characteristics + 15 FBO characteristics) were measured on the nominal and ordinal scales (Table A1). The ordinal scale used was binary with the codes 1 and 2 referring to the least and most preferred, respectively. Quality control and quality assurance (QCQA) of the gathered data were implemented by employing the procedures of identifying erroneous values and data gaps, statistical distributions, descriptive statistics, and multicollinearity. Following the QCQA procedure, the resultant data matrix of 49 columns (variables) × 53 rows (N = sample size) was analyzed. Anderson-Darling (AD) statistic was performed to test the normality assumption of error terms and how well the data follow any of 14 parametric distributions and 2 transformations. The non-parametric tests of Spearman's rank-order and Kendall Tau-b rank-order correlation matrices were used to capture the strength (the magnitude of correlation coefficient—r) and direction (the sign of r) of pairwise linear relationships among the variables.

Since the response variables were categorical with binary values, RF classification model, a non-parametric method, was performed. The RF algorithm developed by Breiman (2001) creates an ensemble of decision trees, instead of one tree (model), based on a bootstrapping-based sampling of *N* rows of the original data with replacement. About one-third (37%) of the original data not included in the training dataset of a given decision tree is referred to as out-of-bag (OOB) observations and treated as a validation dataset (Breiman, 2001). The bootstrapping procedure is what most makes RF distinct and among the best of today's available machine-learning algorithms with the strongest predictive and classification power and accuracy (Luo et al., 2004). The number of decision trees to be grown was tried using 200, 500 and 1000 trees in the exploratory data analysis and was determined finally as 200. The number of predictors tested at each node in RF was 32 (Table A1). The minimum number of a non-terminal node was selected as two since this number was reported to enable decision trees to grow to their full sizes, thus yielding the best results with RF (Luo et al., 2004).

The artificial intelligence and machine-learning algorithms are referred to as the black-box models whose inputs and outputs are well observed. But their internal structures are either not well or not at all understood. Therefore, the interpretability of the RF models was introduced using their following two outputs based on the OBB validation dataset: (1) predictive performance, and (2) normalized relative variable importance. Area under a receiver operation characteristic (ROC) curve (AUC), lift, and overall misclassification rate were used to evaluate the performance of the 17 RF classification models. The scale- and classification-threshold-invariant AUC value varies between 0 (100% wrong predictions) and 1 (100% correct predictions). Lift measures how frequently two independent events are associated with a focus on the most critical x^{th} percentile of the samples, and thus, indicates the classification efficiency. In this study, the 10^{th} percentile of classification accuracy was adopted to compare against the entire dataset. The lower misclassification rate is also known as generalization error rate shows that the more dependencies or better discriminations between output and inputs take place.

Normalized relative variable importance was ranked using the permutation and Gini methods. The permutation variable importance metric was determined monitoring changes in the predictive power of the RF models in response to the shuffling of within-column values of each predictor at a time, regardless of what predictors were internally used in the model. The (internal) Gini method describes the role of predictors in the model in the way that the model was constructed. The 17 RF binary classification models were built using Salford Predictive Modeler® RandomForests® 8.3. The alternative hypotheses posed to capture the differences among the characteristics of family businesses, family employees, and FBOs for the 17 responses were tested using Chi-square goodness-of-fit test based on the normalized variable importance scores. Minitab 17 was used in this analysis.

2 Results

A total of 49 variables (17 responses and 32 predictors), and their observed levels, frequencies and maximum entropy values are described in Table A1 (N = 53). Entropy measures the degree of uncertainty (information availability) in a random variable. It ranges from 0 for a homogenous dataset when no information is needed (predictability peaks) to 1 for a perfectly random dataset when unpredictability (information needed) peaks. In our dataset, the maximum entropy value varied between 0.23 for honesty (QMS1) as the successor quality that FBOs seek and 1.0 for the fields of business activity (BAC). The response rate for the 49 questions was between 100% (N = 53) and 30.2% (N = 16).

Since the dataset did not meet the parametric test assumptions based on Anderson-Darling test, the two non-parametric correlation matrices of Spearman and Kendall Tau-b were used as the more robust measures of linear relationships among the categorical variables. These correlation matrices use different logics to measure the correlation. Although the two matrices had very similar r values, the Spearman's r values (Figure A1) were slightly higher than the Kendall Tau-b ones (Figure A2).

Therefore, only Spearman's r values were reported for the remaining response variables other than FBOs' willingness to transfer to successors (TFM). FBOs' TFM differed significantly between the two cities and was higher in Ardahan than Sivas (r = 0.50; P< 0.05 for both tests) (Figures A1 and A2). FBO's TFM was most strongly correlated with the FBOs' attitudes of sharing experiences (ATF2) and supervision (TAT3) during succession negatively (r = -0.79) and with the partner availability (PAV) and the presence of non-family managers (NFM) positively (r = 0.63) (P< 0.05) (Table 2).

Among the 16 qualities that FBOs sought in successors (QMS1 to 16), the strongest negative correlation was found between successor's commitment to the family business (QMS2), entrepreneurship (QMS3) and past performance (QMS11) and the FBOs' attitude of on-site training during succession (TAT1) (r = -1.0; P < 0.05) (Table 2). Likewise, the strongest positive correlation existed between successor's gender (QMS14) and on-site training during succession (TAT1) (r = 1.0; P < 0.05) (Table 2). In the strongest negative correlations with the QMS1 to 16, the most frequently used predictor was the FBO attitudes in the succession of sharing experiences (ATF2) and supervision (TAT3) (29.4%) followed by on-site training by FBOs (TAT1) and FBOs' actively working (TAT4) (23.5%) (Table 2). For the strongest positive correlations, the most frequent predictor was the job task definition of family employees (FME8), partner availability (PAV), and transfer of rights before death (TAT5) (23.5%) (Table 2).

Table 2 | Maximum correlation coefficient (*r*) values of Spearman's rank-order test between 17 response variables and the remaining variables

Response variable	Max *r* values	
	Negative	**Positive**
TFM	-0.790	0.632
	TAT3-ATF2	PAV-NFM
QMS1	-0.693	0.480
	FME1	FME6
QMS2	-1.0	0.661
	TAT1	ATF4
QMS3	-1.0	0.661
	TAT1	ATF4
QMS4	-0.395	0.395
	BLS-TAT3-4-6-ATF2-3	TAT5
QMS5	-0.500	0.632
	TAT2	GRA
QMS6	-0.395	0.316
	GRA-TAT3-4-ATF2	BLS-PAV-NFM-FME8-TAT6-ATF3
QMS7	-0.395	0.395
	BLS-TAT3-4-6-ATF2-3	TAT5
QMS8	-0.395	0.316
	GRA-TAT3-4-ATF2	BLS-PAV-NFM-FME8-TAT6-ATF3
QMS9	-0.755	0.755
	TAT2	NFMC-TIM
QMS10	-0.790	0.632
	BLS-TAT1-ATF3	PAV-NFM-FME8
QMS11	-1.0	0.661
	TAT1	ATF4
QMS12	-0.800	0.800
	FME7	PAV-NFM-FME8
QMS13	-0.395	0.557
	ATF4	TAT1-GRA-ATF3
QMS14	-0.661	1.0
	ATF4	TAT1
QMS15	-0.294	0.412
	ATF1	BEY-FME1
QMS16	-0.632	0.755
	PAV-NFM-FME8	TAT1

Source: authors

Based on the OOB validation dataset, the ROC-AUC metric varied between 0.47 (0.13 to 0.82 at 95% confidence limit) for birth order of heirs (QMS15) and 0.80 (0.68 to 0.92 at 95% confidence limit) for sub-share size of ownership (QMS16) (Table A3). As the ROC-AUC value reaches unity, the RF classification models have a better discrimination ability. The lift value of the most critical 10th percentile ranged from 0.0 for successor honesty (QMS1) to 4.4 for the successor's ability to solve problems (QMS6) among the 17 models (Table A3). The higher the lift value is, the better the model performance is. The overall misclassification rate of the 17 RF models changed between a minimum of 0.26 for both FBOs' willingness to transfer to a successor (TFM) and successor seniority (QMS13) and a maximum of 0.57 for successor's trustworthiness (QMS7) (Table A3).

The overall classification accuracy according to the confusion matrices of the 17 model outcomes varied between 11.3% for successor's birth order (QMS15) and 64.2% for successor's respectability (QMS10) (Table A4). The second-best classification accuracy belonged to FBO's willingness to transfer to a successor (TFM) (62%). Out of the 16 successor qualities, the successor's creativity (QMS8) had the second-worst classification accuracy (15.1%) (Table A4).

The normalized scores (between 0 and 100%) of relative predictor importance are provided for the 17 RF classification models in Table A5 based on the two methods of the RF algorithm: (1) permutation, and (2) Gini. The Gini method shows the relative importance of the internal predictors of the models in the way that the models were constructed. However, the permutation method quantifies the relative importance by monitoring the impact of randomly shuffling values of each predictor in itself on the model performance, regardless of the internal model predictors and coefficients used. If a certain variable is important, then randomly permuting the order of its values reduces the model accuracy. Hence, the permutation method may as well be considered a sensitivity analysis since the more sensitive the model performance is to a given predictor, the higher its relative importance is. Furthermore, it captures not only main (individual predictor) effects but also interaction effects when the non-linear mechanisms are involved in the prediction of events. It is because if an interaction between two or more predictors exists, then permuting one of the variables in the interaction decreases the predictive power of the interaction, and thus, the model which in turn shows up as important in RF.

Based on the permutation method, the most important and sensitive predictors that were most likely to interact are presented in Table A5 for each of the 17 RF classification models. The most frequent predictor of the relatively most important predictors across the 17 RF classification models was found as FBO's education level (EDU, 25.3%) and successor's on-site training by FBO (TAT1, 17.7%) for the Gini and permutation methods, respectively (Table A5). There were the following six RF classification models out of 17 that had the same most important variables according to both approaches: FBO's education (EDU) for successor's trustworthiness (QMS7) and creativity (QMS8); successor's on-site training by FBO (TAT1) for successor's past performance (QMS11); family employee's job task definition (FME8) for successor's past performance (QMS11); business establishment year (BEY) for successor's birth order (QMS15); and business location (CITY) for successor's sub-share size of ownership (QMS16). The most frequent predictors of the least important predictors among the 17 RF classification models were successor's (out-of-)job experience (FME7), successor's gender (GEN), the number of non-family managers (NFMC), and transfer of rights to successors before FBO's death (TAT5) (11.8%) according to the permutation method, and successor's gender (GEN, 70.6%) according to the Gini method.

As for the testing of the alternative hypotheses (H_{a1}) for the 17 responses, Chi-square goodness-of-fit test was performed to compare the observed distributions of the variable importance of the three predictor classes (1 = the FBO attributes; 2 = the family employee attributes; 3 = the family business attributes) for each RF model to their expected uniform distribution under the assumption that the three classes had equal probability. Based on the extent to which the observed distribution of the normalized variable importance scores varied from the hypothesized distribution, the higher proportions of the predictor class of the FBO's attributes were significantly observed in the 14 RF models (TFM, and QMS1 to QMS13) out of 17 ($P = 0.001$). In the remaining three RF models, the higher proportion belonged to the predictor class of the business attributes in shaping how much FBOs valued successor's gender (QMS14) (P = 0.002) and birth order (QMS15) (P = 0.001), and to that of the family employee's attributes in determining how much FBOs valued successor's share of ownership (QMS16) (P = 0.001).

3 Discussion

Single decision trees can also be established using various methods such as classification and regression tree (CART), conditional inference trees (CIT), and boosted regression tree (BRT) (Hothorn et al., 2006; Elith et al., 2008). However, RF consists of hundreds of decision trees using a bootstrapped randomization which renders it resistant to perturbations in the training dataset and able to perform better on the validation (unseen) data (Verikas et al., 2011). Understanding of family business-related insights to be derived from the application of RF is still in its infancy due to the very limited existence of practical and academic studies in related literature. For example, Peltonen (2018) found RF models to outperform BRT and artificial neural networks by being able to correctly classify family firms based on survey data from 7.153 European family businesses. Muller et al. (2016) successfully predicted the business volume of small-to-medium-sized enterprises in Switzerland as a function of characteristics of companies and cities in which businesses operated, and their interactions using RF for binary classification of small and large volume companies. They emphasized RF as the state-of-the-art method in providing a unique combination of prediction accuracy and model interpretability which can serve as the basis for a decision support system.

At the forefront of the FBOs' attitudes in the process of succession were on-site training, supervision, sharing experiences, and actively working. A strong negative correlation of the FBOs' attitude of on-site training was found with the heir or successor qualities of commitment to a family business, entrepreneurship, and past performance (Figures A1 and A2) (Table 2). On-site training was positively associated with male heirs or successors (Figures A1 and A2). Overall, our correlation results in Table 2 and Figures A1 and A2 showed that the FBOs' attitudes of on-site training, supervision, sharing experiences, and actively working in succession lessened with the heir or successor attributes that they considered appropriate.

The permutation method, similar to the logic of the leave-one-out cross-validation, was shown to be a better measure of the relative variable importance than is the Gini method (Strobl et al., 2007). The explanatory variables found to contribute most to segmenting the binary classifications of the 17 response variables can also be considered to lead to the maximum information gain; in other words, the maximum difference in entropies. Strobl et al. (2007) found that the Gini method was biased towards the categorical variables with many distinct values, while the permutation method showed a bias towards correlated variables. In this context, the correlation matrices can also help to mitigate the multicollinearity issue by reducing the possibility of incorporating the highly correlated features in the RF models.

4 Limitations and future research directions

The present study has its limitations in the three main domains. First, the sample size was relatively small when compared to the target population size of 621 and 1562 family businesses officially registered to the Chambers of Commerce and Trade in Ardahan and Sivas, respectively. Therefore, more diverse and larger samples from different geographic regions, sectors, and religious orientations are recommended to obtain more generalizable and transferable results. Second, due to the very private nature of the subject matter, it was complicated to have participators in both cities show a willingness to openly respond to the questions posed in this study. Finally, the (semi-)structured interview and script in qualitative research that takes longer to complete and reveal confidential information generally suffer from validity and reliability. In future studies, other advanced machine-learning algorithms such as gradient boosting machines, and support vector machines remain to be explored in the computational social sciences.

Theoretical and practical conclusions

In practice, the domain of the FBOs' characteristics was found to have a more profound impact on both FBO's willingness to transfer to a successor and what successor's qualities were most valued than did the domains of the family business and employee characteristics. A total of the 17 RF classification models were constructed and validated for the 17 response variables of the FBOs' willingness to transfer to successors, and what successor's qualities FBOs valued most. The relative importance, sensitivity, and interaction effect of the explanatory variables were detected for each of the 17 response variables. What the present study and other similar studies have in common is the high-dimensional data (49 variables in our case) of non-parametric and non-linear nature with small sample sizes, outliers, missing values, and noise. RF was shown to be robust to the complex and dynamic data structures (Bhattacharyya et al., 2011), and thus, how important to maximize the benefit from its use in a better understanding of causal relationships. The theoretical contribution made by this study is related to the data-driven (inductive) nature of the RF algorithm. This machine-learning and the data-mining algorithm used for the classification predictions has the potential to contribute to theory by discovering unknown patterns, drivers and interactions of relationships in the subject matter of the succession process of the family business. The machine-learning algorithms appear to be promising in terms of deriving theoretical and practical insights and strategies from computational social sciences in the era of big data.

References

Allio, M. (2004). Family Businesses: Their Virtues, Vices and Strategic Path. Strategy and Leadership, 32(4), 24–33.

ASSBA. (2016).“Key to Sustainable Success of Family Businesses.” Akbank, Manager Development Unit of Sabanci University EDU, Deloitte and TAIDER Association of Family Businesses (ASSBA), Retrieved December 15, 2018, from https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/risk/aile-sirketlerinde-surdurulebilir-basarinin-anahtarlari.pdf

Bau, M., Hellerstedt, K.,Nordqvist, M., & Wennberg, K. (2013). Succession in Family Firms. In The Landscape of Family Business,167–197. Edward Elgar.

Bhattacharyya, S., Jha, S., Tharakunnel, K.,& Westland, J. C.(2011). Data Mining for Credit Card Fraud: A Comparative Study. Decision Support Systems, 50(3), 602–613.

Blodgett, M. S., Dumas, C., & Zanzi, A. (2011). Emerging Trends in Global Ethics: A Comparative Study of U.S. and International Family Values. Journal of Business Ethics, 99, 29–38.

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.

Cesaroni, F. M., & Sentuti, A. (2017). Family Business Succession and External Advisors. The Relevance of “Soft” Issues. Small Enterprise Research, 24(2),167–188.

Chrisman, J. J., Chua, J. H., & Sharma, P. (1998). Important Attributes of Successors in Family Businesses: An Exploratory Study. Family Business Review,1(11), 19–34.

Chua, J. H., Chrisman, J. J., & Sharma, P. (1999). Defining the Family Business by Behavior. Entrepreneurship Theory and Practice. 23(4), 19–39.

Chua, J. H., Chrisman, J. J., & Sharma, P. (2003). Succession and Nonsuccession Concerns of Family Firms and Agency Relationship with Nonfamily Managers. Family Business Review, 16(2), 89–107.

De Massis, A., Chua, J. H., & Chrisman, J. J. (2008). Factors Preventing Intra-Family Succession. Family Business Review, 2(21), 183–199.

Devany, C. (2006). All in The Family: Succession Planning in A Family Business Is Important for Keeping Family and Business Intact. Journal of Property Management, 71(5), 14–14.

Elith, J., Leathwick, J. R., & Hastie, T.(2008). A Working Guide to Boosted Regression Trees. The Journal of Animal Ecology, 77(4), 802–813.

Grimmer, J. (2015). We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together. Political Science & Politics, 48(1), 80–83.

Gomba, M., & Kele, T. (2016). Succession Planning in Black-Owned Family Businesses: A South African Perspective. International Journal of Business Administration, 7(5), 9–21.

Güller, S. (2010).The Role of Turnover Planning in Institutionalization of Family Businesses and a Research Istanbul University Institute of Social Sciences. [Master thesis]. Department of Business Administration.

Hacıbayramoğlu, M. G. (2014). The Impact of Succession on Family Business Innovation: A Case Study on Machinery Manufacturing Sector in Ankara, M.Sc., Department of Science and Technology Policy Studies, The Graduate School of Social Sciences of Middle East Technical University

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. Journal of Computational and Graphical Statistics, 15(3), 651–674.

Jordan, M. I., & Mitchell, T. M. (2015). Machine Learning: Trends, Perspectives, and Prospects. Science, 349, 255–260.

Laakkonen, A., & Kansikas, J. (2011). Evolutionary Selection and Variation In Family Businesses. Management Research Review, 34(9), 980–995.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A., Brewer, D., & Christakis, N., et al. (2009). Computational Social Science. Science, 323, 721–723.

Le Breton-Miller, I., Miller, D., & Steier, L. (2004). Toward an Integrative Model of Effective FOB Succession. Entrepreneurship Theory and Practice, 28, 305–328.

Luo, T., Kramer, K., Goldgof, D. B., Hall, L. O., Samson, S., Remsen, A., & Hopkins, T. (2004). Recognizing Plankton Images from the Shadow Image Particle Profiling Evaluation Recorder. IEEE Transactions on Systems Man and Cybernetics Part B—Cybernetics, 34(4), 1753–1762.

Muller, D., Te, F., Meyer, F., & Cvijikj, I. P. (2016). Towards Data Driven Decision Support for Financial Institutions: Predicting Small Companies' Business Volume in Switzerland. Paper presented at the annual meeting for the 7th International Conference on Computer Science and Information Technology (CSIT), Amman, Jordan, pp. 1–6.

Nordqvist, M., & Melin, L. (2010). Entrepreneurial families and family firms. Entrepreneurship & Regional Development, 22(3–4), 211–239.

Peltonen, J. (2018). Can Supervised Machine Learning Be Used to Identify Family Firms Using A Sophisticated Definition? Academy of Management Proceedings, Advance online publication. https://doi.org/10.5465/AMBPP.2018.154.

Sharma, P. (2004). An Overview of the Field of Family Business Studies: Current Status and Directions for the Future. Family Business Review,17(1), 1–36.

Sharma, P., & Irving, P. G. (2005). Four Bases of Family Business Successor Commitment: Antecedents and Consequences. Entrepreneurship Theory and Practice, 29(1), 13–33.

Sharma, P., & Rao, A. S. (2000). Successor Attributes in Indian and Canadian Family Firms: A Comparative Study. Family Business Review, 13(4), 313–330.

Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in Random Forest Variable Importance Measures: Illustrations, Sources and A Solution. BMC Bioinformatics, 8, 25.

Tucker, J. (2011). Keeping the Business in the Family and the Family in the Business: What is the Legacy? Journal of Family Business, 1(1), 65–73.

Tuncel, H.T. (2011). *(Not)* Institutionalizing in Family Business. Konya Chamber of Commerce, Survey-Research Service. Retrieved December 15, 2018, from http://www.kto.org.tr/d/file/aile_sirketi_rapor.pdf

Van Praag, C. M., & Versloot, P. H. (2007). What is the Value of Entrepreneurship? A Review of Recent Research. Small Business Economics, 29(4), 351–382.

Venter, E., Boshoff, C., & Maas, G. (2005). The Influence of Successor-Related Factors on the Succession Process in Small and Medium-Sized Family Businesses. Family Business Review, 18(4), 283–303.

Verikas, A., Gelzinis, A., & Bacauskiene, M. (2011). Mining Data with Random Forests: A Survey and Results of New Tests. Pattern Recognition*,* 44(2), 330–349.

Yu, A., Lumpkin, G. T., Sorenson, R. L., & Brigham, K. H. (2012). The Landscape of Family Business Outcomes: A Summary and Numerical Taxonomy of Dependent Variables. Family Business Review*,* 25, 33–57.

Zehe, A. K. (2016). Three Essay On Family Succession. [Dissertation thesis]. Department of Economics at the Technical University of Munich, 1-172.

The research paper passed the review process. | Received: November 27, 2019**; Revised:** January 24, 2020**; Accepted:** February 8, 2020**; Published:** July 10, 2020.

Appendices

Figure A1 | Spearman's rank-order correlation matrix (*P*< 0.05 when *r* ≥ 0.3 and -0.3)

	CITY	GEN	AGE	EDU	GRA	BAC	BLS	BEY	PAV	FAS	BEC	NFM	NFMC	FME1	FME2	FME3	FME4	FME5	FME7	FME8	TFM	TIM	TAT1	TAT2	TAT3	TAT4	TAT5	TAT6	ATF2	ATF3
GEN	-0.19																													
AGE	-0.40	0.74																												
EDU	0.20	0.11	-0.09																											
GRA	-0.39	0.55	0.76	-0.43																										
BAC	0.50	0.19	-0.38	-0.07	-0.36																									
BLS	0.10	0.55	0.06	0.55	-0.30	0.44																								
BEY	-0.05	-0.57	-0.56	-0.22	-0.15	-0.23	-0.23																							
PAV	-0.32	-0.48	-0.05	-0.46	-0.05	-0.40	-0.65	0.19																						
FAS	-0.19	-0.29	-0.23	-0.66	-0.22	0.19	-0.11	0.34	0.60																					
BEC	0.20	0.55	0.38	-0.09	0.26	0.51	0.32	-0.68	-0.60	-0.22																				
NFM	-0.32	-0.48	-0.05	-0.46	-0.05	-0.40	-0.65	0.19	1.00	0.60	-0.60																			
NFMC	0.50	0.38	-0.05	0.54	-0.05	0.50	0.39	-0.45	-0.79	-0.76	0.54	-0.79																		
FME1	-0.25	0.19	0.60	-0.51	0.44	-0.13	-0.37	-0.60	0.32	0.19	0.51	0.32	-0.25																	
FME2	-0.19	0.36	-0.17	0.44	-0.06	0.19	0.55	0.17	-0.48	-0.29	-0.11	-0.48	0.38	-0.66																
FME3	0.35	0.62	0.24	0.41	0.16	0.38	0.67	-0.29	-0.95	-0.57	0.63	-0.95	0.75	-0.23	0.34															
FME4	0.30	0.46	0.33	0.37	0.25	0.23	0.14	-0.72	-0.62	-0.80	0.70	-0.62	0.86	0.23	0.06	0.63														
FME5	0.30	0.46	0.33	0.37	0.25	0.23	0.14	-0.72	-0.62	-0.80	0.70	-0.62	0.86	0.23	0.06	0.63	1.00													
FME7	0.79	-0.06	-0.29	0.46	-0.28	0.40	0.14	-0.19	-0.55	-0.60	0.28	-0.55	0.79	-0.32	-0.06	0.57	0.62	0.62												
FME8	-0.24	-0.55	-0.21	-0.55	0.05	-0.36	-0.75	0.48	0.91	0.55	-0.72	0.91	-0.72	0.07	-0.33	-0.89	-0.67	-0.67	-0.46											
TFM	-0.50	0.38	0.60	-0.39	0.44	-0.25	-0.20	-0.30	0.63	0.38	-0.15	0.63	-0.50	0.50	-0.19	-0.45	-0.25	-0.25	-0.63	0.48										
TIM	0.44	0.37	0.03	0.60	0.03	0.33	0.33	-0.44	-0.78	-0.87	0.47	-0.78	0.98	-0.25	0.37	0.74	0.89	0.89	0.78	-0.71	-0.49									
TAT1	0.38	0.36	0.00	0.44	-0.06	0.19	0.55	0.17	-0.48	-0.29	-0.11	-0.48	0.38	-0.66	0.36	0.62	0.06	0.06	0.48	-0.33	-0.19	0.37								
TAT2	0.00	-0.38	-0.60	-0.24	-0.49	0.25	0.15	0.70	0.32	0.76	-0.54	0.32	-0.50	-0.50	0.19	-0.35	-0.86	-0.86	-0.32	0.43	0.00	-0.60	0.19							
TAT3	0.63	-0.48	-0.67	0.42	-0.51	0.32	-0.09	0.05	-0.35	-0.48	0.05	-0.35	0.63	-0.40	0.06	0.19	0.43	0.43	0.80	-0.23	-0.79	0.62	0.06	-0.16						
TAT4	0.63	0.06	0.10	0.14	0.00	0.32	-0.09	-0.57	-0.35	-0.48	0.60	-0.35	0.63	0.32	-0.48	0.43	0.77	0.77	0.80	-0.41	-0.32	0.62	0.06	-0.63	0.55					
TAT5	0.32	-0.06	0.00	-0.74	0.18	0.40	-0.42	-0.05	0.35	0.48	0.19	0.35	-0.16	0.40	-0.60	-0.19	-0.10	-0.10	0.10	0.41	0.32	-0.26	-0.06	0.16	-0.10	0.35				
TAT6	-0.32	0.60	0.76	0.46	0.32	-0.40	0.42	-0.57	-0.35	-0.48	0.37	-0.35	0.16	0.32	0.06	0.43	0.43	0.43	-0.10	-0.59	0.16	0.26	0.06	-0.63	-0.35	0.10	-0.55			
ATF2	0.63	-0.48	-0.67	0.42	-0.51	0.32	-0.09	0.05	-0.35	-0.48	0.05	-0.35	0.63	-0.40	0.06	0.19	0.43	0.43	0.80	-0.23	-0.79	0.62	0.06	-0.16	1.00	0.55	-0.10	-0.35		
ATF3	0.16	0.06	0.14	0.74	-0.18	-0.40	0.42	0.05	-0.35	-0.48	-0.19	-0.35	0.16	-0.40	0.06	0.43	0.10	0.10	0.35	-0.41	-0.32	0.26	0.60	-0.16	0.10	0.10	-0.55	0.55	0.10	
ATF4	-0.25	0.19	0.60	-0.51	0.44	-0.13	-0.37	-0.60	0.32	0.19	0.51	0.32	-0.25	1.00	-0.66	-0.23	0.23	0.23	-0.32	0.07	0.50	-0.25	-0.66	-0.50	-0.40	0.32	0.40	0.32	-0.40	-0.40

Source: authors

Figure A2 | Kendall Tau-b rank-order correlation matrix (P< 0.05 when r ≥ 0.3 and -0.3)

	CITY	GEN	AGE	EDU	GRA	BAC	BLS	BEY	PAV	FAS	BEC	NFM	NFMC	FME1	FME2	FME3	FME4	FME5	FME7	FME8	TFM	TIM	TAT1	TAT2	TAT3	TAT4	TAT5	TAT6	ATF2	ATF3
GEN	-0.19																													
AGE	-0.38	0.69																												
EDU	0.18	0.10	-0.04																											
GRA	-0.36	0.51	0.69	-0.34																										
BAC	0.50	0.19	-0.35	-0.07	-0.34																									
BLS	0.09	0.52	0.04	0.46	-0.30	0.42																								
BEY	-0.05	-0.55	-0.53	-0.24	-0.16	-0.22	-0.20																							
PAV	-0.32	-0.48	-0.04	-0.44	-0.04	-0.40	-0.61	0.18																						
FAS	-0.19	-0.29	-0.21	-0.63	-0.21	0.19	-0.10	0.33	0.60																					
BEC	0.18	0.52	0.35	-0.08	0.26	0.49	0.31	-0.64	-0.57	-0.21																				
NFM	-0.32	-0.48	-0.04	-0.44	-0.04	-0.40	-0.61	0.18	1.00	0.60	-0.57																			
NFMC	0.50	0.38	-0.05	0.51	-0.05	0.50	0.37	-0.43	-0.79	-0.76	0.51	-0.79																		
FME1	-0.25	0.19	0.57	-0.49	0.41	-0.13	-0.35	-0.58	0.32	0.19	0.49	0.32	-0.25																	
FME2	-0.19	0.36	-0.16	0.42	-0.05	0.19	0.52	0.16	-0.48	-0.29	-0.10	-0.48	0.38	-0.66																
FME3	0.34	0.60	0.20	0.36	0.12	0.36	0.60	-0.25	-0.91	-0.55	0.56	-0.91	0.72	-0.22	0.33															
FME4	0.29	0.44	0.29	0.32	0.20	0.22	0.12	-0.63	-0.59	-0.76	0.56	-0.59	0.82	0.22	0.05	0.58														
FME5	0.29	0.44	0.29	0.32	0.20	0.22	0.12	-0.63	-0.59	-0.76	0.56	-0.59	0.82	0.22	0.05	0.58	1.00													
FME7	0.79	-0.06	-0.27	0.44	-0.26	0.40	0.13	-0.18	-0.55	-0.60	0.26	-0.55	0.79	-0.32	-0.06	0.55	0.59	0.59												
FME8	-0.22	-0.51	-0.15	-0.48	0.04	-0.33	-0.63	0.39	0.85	0.51	-0.59	0.85	-0.67	0.07	-0.30	-0.81	-0.62	-0.62	-0.42											
TFM	-0.50	0.38	0.57	-0.37	0.41	-0.25	-0.18	-0.29	0.63	0.38	-0.14	0.63	-0.50	0.50	-0.19	-0.43	-0.24	-0.24	-0.63	0.45										
TIM	0.42	0.36	0.04	0.57	0.04	0.32	0.26	-0.41	-0.75	-0.84	0.39	-0.75	0.95	-0.24	0.36	0.68	0.87	0.87	0.75	-0.63	-0.47									
TAT1	0.38	0.36	0.00	0.42	-0.05	0.19	0.52	0.16	-0.48	-0.29	-0.10	-0.48	0.38	-0.66	0.36	0.60	0.05	0.05	0.48	-0.30	-0.19	0.36								
TAT2	0.00	-0.38	-0.57	-0.23	-0.45	0.25	0.14	0.67	0.32	0.76	-0.51	0.32	-0.50	-0.50	0.19	-0.34	-0.82	-0.82	-0.32	0.40	0.00	-0.58	0.19							
TAT3	0.63	-0.48	-0.63	0.39	-0.47	0.32	-0.09	0.05	-0.35	-0.48	0.04	-0.35	0.63	-0.40	0.06	0.18	0.41	0.41	0.80	-0.21	-0.79	0.60	0.06	-0.16						
TAT4	0.63	0.06	0.09	0.13	0.00	0.32	-0.09	-0.55	-0.35	-0.48	0.57	-0.35	0.63	0.32	-0.48	0.41	0.73	0.73	0.80	-0.38	-0.32	0.60	0.06	-0.63	0.55					
TAT5	0.32	-0.06	0.00	-0.70	0.17	0.40	-0.39	-0.05	0.35	0.48	0.18	0.35	-0.16	0.40	-0.60	-0.18	-0.09	-0.09	0.10	0.38	0.32	-0.25	-0.06	0.16	-0.10	0.35				
TAT6	-0.32	0.60	0.72	0.44	0.30	-0.40	0.39	-0.55	-0.35	-0.48	0.35	-0.35	0.16	0.32	0.06	0.41	0.41	0.41	-0.10	-0.55	0.16	0.25	0.06	-0.63	-0.35	0.10	-0.55			
ATF2	0.63	-0.48	-0.63	0.39	-0.47	0.32	-0.09	0.05	-0.35	-0.48	0.04	-0.35	0.63	-0.40	0.06	0.18	0.41	0.41	0.80	-0.21	-0.79	0.60	0.06	-0.16	1.00	0.55	-0.10	-0.35		
ATF3	0.16	0.06	0.13	0.70	-0.17	-0.40	0.39	0.05	-0.35	-0.48	-0.18	-0.35	0.16	-0.40	0.06	0.41	0.09	0.09	0.35	-0.38	-0.32	0.25	0.60	-0.16	0.10	0.10	-0.55	0.55	0.10	
ATF4	-0.25	0.19	0.57	-0.49	0.41	-0.13	-0.35	-0.58	0.32	0.19	0.49	0.32	-0.25	1.00	-0.66	-0.22	0.22	0.22	-0.32	0.07	0.50	-0.24	-0.66	-0.50	-0.40	0.32	0.40	0.32	-0.40	-0.40

Source: authors

Volume **9** | Issue **3** | 2020
https://doi.org/10.18267/j.cebr.235

Table A1| Description of response and explanatory variables used in this study (N = 53)

No	Symbol	Description	*N*	Observed level		Level frequency (%)	Variable type	Information classification	Max entropy
1	GEN	Gender	53	1	Female	5.7	Predictor	FBO attribute	0.31
				2	Male	94.3			
2	AGE	Age	53	1	18-30	7.6	Predictor	FBO attribute	0.84
				2	31-40	11.3			
				3	41-50	35.9			
				4	> 50	45.2			
3	EDU	Education level	53	1	Primary	15.1	Predictor	FBO attribute	0.94
				2	Secondary	17.0			
				3	High	37.7			
				4	University	18.9			
				5	Masters	11.3			
4	GRA	Graduation	16	1	Business/Economics	43.8	Predictor	FBO attribute	0.77
				2	Engineering	37.5			
				3	Education	6.3			
				4	Veterinary	6.3			
				5	Others	6.1			
5	BAC	Activity field	51	1	Commerce	33.3	Predictor	Business attribute	1.0
				2	Service	33.3			
				3	Manufacturing	33.4			
6	CITY	City	53	1	Ardahan	52.8	Predictor	Business attribute	0.99
				2	Sivas	47.2			
7	BLS	Legal status	45	1	Private Co.	73.3	Predictor	Business attribute	0.63
				2	Incorporated Co.	4.4			
				3	Limited Co.	22.3			
8	BEY	Establishment year	51	1	1920-1950	2.0	Predictor	Business attribute	0.83
				2	1951-1980	27.5			
				3	1981-2000	39.2			
				4	2001-2018	31.3			
9	PAV	Partner availability	53	1	Yes	20.8	Predictor	Business attribute	0.73
				2	No	79.2			
10	FAS	Family ownership share	51	1	≤ 50%	2.0	Predictor	Business attribute	0.37
				2	> 50% ≤ 99%	9.8			
				3	100%	88.2			
11	BEC	Employee count	47	1	Non	4.3	Predictor	Business attribute	0.80
				2	1-5	53.2			
				3	6-10	14.9			
				4	> 10	27.6			
12	NFM	Non-family manager availability	53	1	Yes	26.4	Predictor	Business attribute	0.83
				2	No	73.6			
13	NFMC	Non-family manager count	52	1	1	84.6	Predictor	Business attribute	0.61
				2	> 1	15.4			
14	FME1	Employee count	43	1	Non	7.0	Predictor	Family employee attribute	0.39
				2	1-5	86.1			
				3	6-10	4.7			
				4	> 10	2.2			
15	FME2	Gender	45	1	Female	15.6	Predictor	Family employee attribute	0.62
				2	Male	84.6			
16	FME3	Education	45	1	Primary	2.2	Predictor	Family employee attribute	0.77
				2	Secondary	13.3			
				3	High	40.0			
				4	University	44.5			
17	FME4	Work period in the place	45	1	< 1 year	4.4	Predictor	Family employee attribute	0.87
				2	1-5 years	22.2			
				3	6-10 years	20.0			
				4	11-20 years	42.2			
				5	> 20 years	11.2			
18	FME5	Work period outside	43	1	< 1 year	25.6	Predictor	Family employee attribute	0.94
				2	1-5 years	16.3			

				3	6-10 years	41.9			
				4	11-20 years	16.2			
19	FME6	Full vs part time status	45	1	Full time	73.3	Predictor	Family employee attribute	0.83
				2	Part time	26.7			
20	FME7	(out-of-)job experience	45	1	Yes	46.7	Predictor	Family employee attribute	0.99
				2	No	53.3			
21	FME8	Job task definition	45	1	Manager	48.9	Predictor	Family employee attribute	0.83
				2	Vice manager	13.3			
				3	Technical staff	22.2			
				4	Office staff	8.9			
				5	Others	6.7			
22	TIM	Time to leave to heir/successor	49	1	No decision	55.1	Predictor	FBO attitude	0.75
				2	In 1-10 years	40.8			
				3	In > 10 years	4.1			
23	TAT1	On-site training in succession	53	1	Least preferred	43.4	Predictor	FBO attitude	0.98
				2	Most preferred	56.6			
24	TAT2	Monitoring in succession	53	1	Least preferred	52.8	Predictor	FBO attitude	0.99
				2	Most preferred	47.2			
25	TAT3	Supervision in succession	53	1	Least preferred	56.6	Predictor	FBO attitude	0.98
				2	Most preferred	43.4			
26	TAT4	Actively working in succession	53	1	Least preferred	66.0	Predictor	FBO attitude	0.92
				2	Most preferred	34.0			
27	TAT5	Transfer of rights before death	53	1	Least preferred	56.6	Predictor	FBO attitude	0.98
				2	Most preferred	43.4			
28	TAT6	Transfer of rights after death	53	1	Least preferred	66.0	Predictor	FBO attitude	0.92
				2	Most preferred	34.0			
29	ATF1	Encouraging heir to take initiatives	53	1	Least preferred	20.8	Predictor	FBO attitude	0.73
				2	Most preferred	79.2			
30	ATF2	Sharing experiences with heir	53	1	Least preferred	26.4	Predictor	FBO attitude	0.83
				2	Most preferred	73.6			
31	ATF3	Criticizing heir	53	1	Least preferred	58.5	Predictor	FBO attitude	0.97
				2	Most preferred	41.5			
32	ATF4	Remaining neutral	53	1	Least preferred	84.9	Predictor	FBO attitude	0.61
				2	Most preferred	15.1			
33	TFM	Willingness to transfer to successor	50	1	Yes	76.0	Response	FBO attitude	0.79
				2	No	24.0			
34	QMS1	Honesty	53	1	Least preferred	3.8	Response	Successor quality FBO seeks	0.23
				2	Most preferred	96.2			
35	QMS2	Commitment to family business	53	1	Least preferred	54.7	Response	Successor quality FBO seeks	0.99
				2	Most preferred	45.3			
36	QMS3	Entrepreneurship	53	1	Least preferred	30.2	Response	Successor quality FBO seeks	0.88
				2	Most preferred	69.8			
37	QMS4	Education level	53	1	Least preferred	37.7	Response	Successor quality FBO seeks	0.95
				2	Most preferred	62.3			
38	QMS5	Ability to socialize	53	1	Least preferred	43.4	Response	Successor quality FBO seeks	0.98
				2	Most preferred	56.6			
39	QMS6	Ability to solve problems	53	1	Least preferred	17.0	Response	Successor quality FBO seeks	0.65
				2	Most preferred	83.0			
40	QMS7	Trustworthiness	53	1	Least preferred	22.6	Response	Successor quality FBO seeks	0.77
				2	Most preferred	79.4			
41	QMS8	Creativity	53	1	Least preferred	13.2	Response	Successor quality FBO seeks	0.56
				2	Most preferred	86.8			
42	QMS9	Ability to strategically plan	53	1	Least preferred	22.6	Response	Successor quality FBO seeks	0.77
				2	Most preferred	77.4			
43	QMS10	Respectability	53	1	Least preferred	52.8	Response	Successor quality FBO seeks	0.99
				2	Most preferred	47.2			
44	QMS11	Past performance	53	1	Least preferred	34.0	Response	Successor quality FBO seeks	0.92
				2	Most preferred	66.0			
45	QMS12	Know-how	53	1	Least preferred	45.3	Response	Successor quality FBO seeks	0.99
				2	Most preferred	54.7			
46	QMS13	Seniority	53	1	Least preferred	71.7	Response	Successor quality FBO seeks	0.85

Volume **9** | Issue **3** | 2020
https://doi.org/10.18267/j.cebr.235

				2	Most preferred	28.3			
47	QMS14	Gender	53	1	Least preferred	67.9	Response	Successor quality FBO seeks	0.90
				2	Most preferred	32.1			
48	QMS15	Birth order	53	1	Least preferred	88.7	Response	Successor quality FBO seeks	0.50
				2	Most preferred	11.3			
49	QMS16	Subshare size	53	1	Least preferred	75.5	Response	Successor quality FBO seeks	0.80
				2	Most preferred	24.5			

Source: authors

Table A3 | Error measures of 17 RF classification models

Name	TFM	QMS1	QMS2	QMS3	QMS4	QMS5	QMS6	QMS7	QMS8	QMS9	QMS10	QMS11	QMS12	QMS13	QMS14	QMS15	QMS16
ROC-AUC	0.60	0.76	0.69	0.61	0.66	0.66	0.78	0.47	0.66	0.64	0.74	0.59	0.59	0.74	0.66	0.47	0.80
Lower confidence limit	0.42	0.53	0.55	0.45	0.52	0.51	0.62	0.29	0.47	0.43	0.60	0.44	0.44	0.59	0.50	0.13	0.68
Upper confidence limit	0.77	0.98	0.84	0.78	0.81	0.81	0.94	0.65	0.84	0.86	0.88	0.75	0.74	0.89	0.81	0.82	0.92
Lift	1.32	0.00	1.14	1.88	1.00	1.87	4.44	1.67	1.43	3.33	1.43	1.11	1.38	1.13	1.47	0.85	1.33
Overall misclassification rate	0.26	0.38	0.36	0.49	0.40	0.40	0.30	0.57	0.47	0.40	0.28	0.45	0.42	0.26	0.38	0.34	0.28

Source: authors

Table A4 | Confusion matrices of 17 RF classification models

	TFM				**QMS1**				**QMS2**			
Observed classes	*N*	Percent correct	Predicted classes		*N*	Percent correct	Predicted classes		*N*	Percent correct	Predicted classes	
			1	2			1	2			1	2
1	38	65.8	25	13	2	100.0	2	0	29	48.3	14	15
2	12	50.0	6	6	51	25.5	38	13	24	70.8	7	17
Total	50		31	19	53		40	13	53		21	32
Average		57.9				62.8				59.6		
Overall accuracy		62.0				28.3				58.5		
	QMS9				QMS10				QMS11			
1	12	100.0	12	0	28	57.1	16	12	18	88.9	16	2
2	41	7.3	38	3	25	72.0	7	18	35	28.6	25	10
Total	53		50	3	53		23	30	53		41	12
Average		53.7				64.6				58.7		
Overall accuracy		28.3				64.2				49.1		

	QMS3				**QMS4**				**QMS5**			
Observed classes	*N*	Percent correct	Predicted classes		*N*	Percent correct	Predicted classes		*N*	Percent correct	Predicted classes	
			1	2			1	2			1	2
1	16	93.8	15	1	20	90.0	18	2	23	87.0	20	3
2	37	18.9	30	7	33	36.4	21	12	30	36.7	19	11
Total	53		45	8	53		39	14	53		39	14
Average		56.3				63.2				61.8		
Overall accuracy		41.5				56.6				58.5		
	QMS12				QMS13				QMS14			
1	24	79.2	19	5	38	15.8	6	32	36	33.3	12	24
2	29	41.4	17	12	15	93.3	1	14	17	88.2	2	15
Total	53		36	17	53		7	46	53		14	39
Average		60.3				54.6				60.8		
Overall accuracy		58.5				37.7				50.9		

	QMS6				**QMS7**				**QMS8**			
Observed classes	*N*	Percent correct	Predicted classes		*N*	Percent correct	Predicted classes		*N*	Percent correct	Predicted classes	
			1	2			1	2			1	2
1	9	100.0	9	0	12	100.0	12	0	7	100.0	7	0
2	44	9.1	40	4	41	14.6	35	6	46	2.2	45	1
Total	53		49	4	53		47	6	53		52	1
Average		54.6				57.3				51.1		
Overall accuracy		24.5				34.0				15.1		
	QMS15				QMS16							
1	47	0.0	0	47	40	22.5	9	31				
2	6	100.0	0	6	13	100.0	0	13				
Total	53		0	53	53		9	44				
Average		50.0				61.3						
Overall accuracy		11.3				41.5						

Source: authors

Table A5 | Normalized scores (%) of relative predictor importance based on (a) permutation and (b) Gini methods of 17 RF classification models

(a) TFM		(b) TFM		(a) QMS1		(b) QMS1		(a) QMS2		(b) QMS2	
GRA	100.0	FME8	100.0	TAT1	100.0	FME6	100.0	TAT2	100.0	BEC	100.0
FME8	72.2	GRA	98.9	FME3	72.1	TAT1	66.5	FME4	29.1	TAT2	92.0
CITY	58.4	EDU	80.0	FME6	64.2	FME4	60.6	GRA	27.9	GRA	82.7
BAC	52.5	BAC	56.5	AGE	57.6	FME5	49.8	BEC	26.7	FME4	71.9
ATF2	51.2	BEY	50.4	EDU	45.1	AGE	34.3	ATF3	26.3	TIM	61.8
EDU	38.3	AGE	48.7	FME1	39.4	FME3	31.9	TAT3	24.3	FME8	51.0
TIM	30.9	CITY	48.0	PAV	23.3	TIM	30.4	BAC	22.1	ATF3	51.0
FME1	30.4	FME3	45.7	TAT3	22.2	ATF3	20.5	FME5	17.5	FME5	49.8
TAT3	23.1	ATF2	39.2	BAC	13.5	FME1	20.0	TIM	16.1	EDU	49.2
NFM	19.8	FME4	37.7	BEC	13.1	EDU	20.0	FME3	14.0	AGE	47.4
BEY	19.5	FME5	36.7	FME8	7.4	BEC	18.1	ATF2	13.2	BEY	38.1
FME3	19.3	FME1	36.1	TAT6	2.8	TAT2	15.8	TAT6	13.1	FME3	35.9
NFMC	18.4	TAT3	23.6	TAT5	2.3	CITY	14.8	TAT1	9.3	TAT3	35.7
BLS	11.6	ATF3	17.7			BAC	12.3	FME1	8.6	BAC	32.6
AGE	11.5	TIM	16.9			ATF1	10.7	PAV	8.2	TAT1	30.1
TAT5	11.0	BEC	16.9			BLS	10.1	TAT4	7.4	ATF1	24.6
FME5	10.8	ATF4	12.9			TAT3	9.0	FME7	5.9	FME1	23.6
BEC	9.6	TAT4	12.3			FME7	8.9	FME2	3.4	TAT5	21.7
ATF3	9.5	NFMC	11.9			FME8	8.0	ATF1	3.4	ATF2	19.8
FME4	9.3	NFM	10.9			BEY	7.2	TAT5	3.1	NFM	19.5
TAT4	8.8	TAT5	10.5			TAT4	6.2	NFMC	2.3	TAT6	18.6
FAS	7.2	TAT2	10.2			ATF2	6.1	FAS	1.8	NFMC	16.0
PAV	6.4	PAV	10.2			TAT6	3.3	NFM	0.4	FME6	13.9
GEN	5.0	BLS	10.0			NFM	2.9			CITY	13.6
TAT1	3.3	FAS	9.7			TAT5	2.8			PAV	13.2
FME7	3.2	TAT6	9.5			FME2	2.5			TAT4	12.0
TAT6	2.5	ATF1	9.0			FAS	1.4			BLS	11.5
FME6	2.3	TAT1	8.8			PAV	0.7			FAS	11.1

Volume 9 | Issue 3 | 2020
https://doi.org/10.18267/j.cebr.235

ATF4 0.9	FME7 8.3		GRA 0.2		ATF4 6.2
	FME6 8.1		ATF4 0.1		FME2 6.0
	FME2 6.0				FME7 5.8
	GEN 2.4				GEN 4.7

(a) QMS3		**(b) QMS3**		**(a) QMS4**		**(b) QMS4**		**(a) QMS5**		**(b) QMS5**	
TIM	100.0	EDU	100.0	TAT5	100.0	EDU	100.0	AGE	100.0	BAC	100.0
TAT1	78.9	BEC	79.7	AGE	71.2	FME8	82.6	BEC	85.5	AGE	85.3
FME4	77.1	TIM	69.5	NFMC	59.3	BEY	72.8	FME5	85.1	FME8	84.8
TAT5	69.3	TAT1	65.4	BEY	56.8	AGE	67.0	EDU	81.6	EDU	84.3
BEC	60.0	FME5	64.1	BAC	56.6	BEC	64.7	TAT2	70.0	FME5	81.5
AGE	48.2	FME8	63.1	PAV	51.9	FME4	62.3	BAC	58.9	BEC	78.8
FME5	45.3	FME4	58.7	ATF3	51.6	NFMC	54.7	BEY	46.0	TAT2	69.4
TAT4	40.7	FME1	50.0	FME6	50.7	FME5	53.9	FME8	34.0	BEY	62.6
FME3	37.6	AGE	48.8	EDU	43.8	TAT5	52.6	FME7	32.2	FME4	56.1
FME1	35.5	BEY	43.9	FME5	37.9	PAV	49.4	TAT6	30.3	GRA	46.8
BAC	33.7	FME3	35.4	FME8	34.8	GRA	46.6	ATF3	25.4	FME7	40.7
EDU	32.9	ATF3	33.5	BEC	34.6	ATF3	39.3	TAT1	23.5	ATF3	37.5
TAT2	27.6	BAC	33.1	FME4	30.0	FME6	35.4	NFM	21.0	FME3	37.4
BLS	27.0	GRA	31.0	GRA	25.7	BAC	35.1	FME3	17.7	TAT1	35.9
TAT3	21.8	BLS	28.9	TAT6	25.3	FME3	32.9	TIM	17.3	NFM	31.6
ATF2	21.4	ATF2	24.5	TAT4	24.1	TAT4	27.7	TAT4	14.3	FME1	28.3
GRA	14.5	GEN	20.3	FME1	15.2	TAT1	25.7	TAT5	14.2	BLS	27.4
TAT6	13.4	TAT5	17.3	FME3	12.8	TAT6	22.1	GRA	13.5	TIM	25.2
ATF1	11.5	CITY	15.5	TAT3	12.4	FME7	22.1	PAV	11.8	TAT6	22.9
FME6	7.8	TAT3	15.3	TAT1	12.0	TIM	21.1	FME6	10.4	CITY	19.7
PAV	7.0	ATF1	14.3	FME7	11.9	CITY	20.4	BLS	9.2	FAS	19.2
FME8	3.6	NFMC	14.2	FAS	6.9	FME1	19.7	NFMC	8.8	TAT5	17.7
FME2	2.0	FME6	14.0	GEN	6.3	ATF2	19.3	FAS	7.5	PAV	16.9

GEN	1.8	TAT4	13.5	NFM	4.9	NFM	18.7	FME4	4.2	TAT4	16.8
NFM	1.1	TAT2	13.4	ATF2	4.4	TAT2	18.3	TAT3	2.6	TAT3	16.8
NFMC	1.0	FME7	12.0	TAT2	2.4	TAT3	16.7	ATF2	0.9	ATF1	16.4
ATF3	0.8	TAT6	11.2			BLS	13.0			NFMC	13.3
FAS	0.6	NFM	10.2			GEN	12.3			FME6	12.6
		FME2	9.0			ATF1	11.2			ATF2	12.6
		ATF4	8.5			FAS	10.4			ATF4	10.6
		PAV	8.4			FME2	10.1			FME2	7.7
		FAS	7.6			ATF4	5.3			GEN	3.5
(a) QMS6		**(b) QMS6**		**(a) QMS7**		**(b) QMS7**		**(a) QMS8**		**(b) QMS8**	
TAT6	100.0	FME8	100.0	EDU	100.0	EDU	100.0	EDU	100.0	EDU	100.0
BAC	98.9	TAT4	87.5	NFM	82.2	NFM	72.0	TAT3	98.7	FME8	63.2
BLS	94.6	BLS	82.7	TAT1	38.7	FME8	70.6	BEC	82.6	BEY	53.2
TAT4	92.3	ATF1	76.7	FME8	33.0	AGE	59.1	FME8	77.6	ATF2	42.0
ATF1	66.1	TAT6	76.6	AGE	30.9	BAC	42.8	ATF2	61.5	TAT5	40.8
FME8	61.7	TAT3	65.7	FME6	28.3	TAT1	40.7	ATF4	58.5	TIM	40.6
TAT3	60.4	FME5	60.0	ATF2	26.2	BEY	40.0	NFM	56.5	BEC	38.1
FME5	59.8	EDU	54.0	TAT5	18.0	BEC	29.9	TIM	53.9	NFM	37.4
FME4	57.1	TAT2	49.7	NFMC	14.7	FME5	28.4	BEY	51.9	TAT3	33.2
FAS	35.6	FME4	47.8	BEY	13.2	FME4	27.5	TAT5	46.6	FME4	32.1
FME7	27.8	BAC	46.9	ATF4	12.8	FME6	26.2	BLS	39.4	TAT1	31.5
PAV	27.4	FAS	43.4	BEC	6.0	GRA	25.2	ATF3	38.5	AGE	28.3
TAT2	23.8	FME3	33.3	BAC	3.6	TAT5	25.1	PAV	38.1	ATF4	28.1
FME1	23.4	AGE	30.7	FME2	3.6	ATF3	23.4	FME4	31.3	FME3	26.2
ATF2	22.6	FME1	26.7	TAT4	2.8	FME3	22.5	TAT1	30.6	TAT6	21.8
FME3	16.5	BEC	24.4	GEN	2.7	FME7	19.9	FME3	30.2	ATF1	21.4
NFM	15.1	ATF2	22.3			NFMC	19.0	TAT6	26.6	BAC	20.8
GRA	14.9	GRA	21.9			TAT2	16.7	FME1	15.5	ATF3	19.6
EDU	14.5	PAV	20.8			TIM	15.8	ATF	10.5	FME	19.6

Volume 9 | Issue 3 | 2020
https://doi.org/10.18267/j.cebr.235

								1		5	
BEC	14.3	BEY	20.6			TAT4	15.4	FAS	10.3	PAV	14.7
FME6	14.1	TAT5	17.3			TAT6	14.6	AGE	7.3	BLS	12.8
BEY	8.7	NFM	16.5			ATF2	14.4	TAT4	6.5	FME6	11.4
TIM	7.5	NFMC	16.4			ATF4	11.6	FME5	5.0	GRA	11.2
TAT5	6.0	TAT1	14.3			PAV	11.5	GEN	4.8	CITY	10.5
NFMC	5.5	ATF3	12.2			TAT3	11.0	FME2	4.0	FME2	10.1
FME2	4.7	CITY	11.8			BLS	10.3	FME6	2.6	TAT2	8.2
ATF3	4.2	TIM	10.1			CITY	10.3	GRA	2.5	TAT4	8.1
ATF4	4.2	FME7	8.6			FME2	10.2	NFMC	0.1	FME1	7.4
GEN	2.3	FME2	7.9			FME1	10.0			NFMC	7.1
		ATF4	6.9			ATF1	9.2			FME7	6.1
		GEN	6.6			GEN	5.4			FAS	5.6
		FME6	5.3			FAS	5.3			GEN	1.8

a) QMS9		**(b) QMS9**		**(a) QMS10**		**(b) QMS10**		**(a) QMS11**		**(b) QMS11**	
TAT2	100.0	FME5	100.0	TAT1	100.0	BEY	100.0	TAT1	100.0	TAT1	100.0
FME5	89.7	FME8	88.2	ATF3	65.1	ATF3	68.5	TAT3	74.9	FME8	82.5
FME7	70.9	TAT2	80.7	BEY	63.0	TAT1	67.8	TAT2	51.0	EDU	71.6
FME4	43.9	EDU	74.1	FME8	35.6	FME8	61.1	GRA	26.2	FME4	57.8
FAS	41.7	AGE	66.8	GRA	33.0	EDU	46.3	NFM	25.7	BEY	55.6
TAT5	38.9	FAS	66.4	TAT4	31.0	FME5	43.0	ATF2	25.0	AGE	54.1
TIM	35.4	BEY	61.3	NFMC	29.5	TAT4	41.8	CITY	24.9	TAT2	52.0
ATF4	33.3	FME4	53.7	ATF2	27.1	GRA	41.5	BAC	22.2	TAT3	51.2
AGE	29.6	FME3	36.9	TAT2	21.1	FME4	34.0	TIM	21.8	BEC	47.2
EDU	26.1	TAT6	36.6	TAT3	20.5	NFMC	30.4	BEC	18.5	GRA	46.8
FME3	23.8	BEC	36.4	FME3	18.2	ATF2	26.3	FME4	17.1	TIM	39.4
TAT4	20.9	TIM	36.1	BEC	11.5	TAT2	24.6	FME6	15.5	BAC	39.2
NFMC	18.5	NFMC	32.1	TAT6	7.5	AGE	24.3	TAT4	11.0	NFM	39.2
ATF2	18.1	BAC	30.1	BAC	7.3	BEC	22.1	ATF1	9.9	FME3	38.4
TAT6	18.0	FME7	29.6	FME1	3.7	BAC	20.7	BEY	9.7	FME5	38.0

BEY	17.6	TAT5	29.5	PAV	2.8	TAT3	20.5	FME2	8.1	ATF2	36.8
NFM	17.4	BLS	28.8	NFM	2.6	FME3	19.7	FME3	6.9	CITY	34.6
TAT3	13.6	ATF4	24.4	TAT5	2.2	FME1	18.1	FME8	5.6	ATF1	25.7
BEC	12.5	GRA	24.1	FME7	2.0	TIM	13.8	FME7	4.9	FME6	20.1
TAT1	12.3	ATF2	23.2	FME2	1.6	TAT5	11.6	TAT5	4.0	FME1	18.7
BLS	12.0	ATF1	21.8	FME5	0.1	TAT6	11.2	ATF4	1.1	TAT6	18.4
FME6	10.3	FME6	21.7			ATF1	10.3	FME1	0.6	TAT5	17.0
FME2	10.2	FME2	19.4			FME6	9.5			TAT4	14.7
FME8	9.4	FME1	17.9			FME2	8.7			BLS	13.2
BAC	4.7	NFM	15.7			CITY	8.6			ATF3	12.8
GRA	0.8	TAT4	15.6			BLS	7.6			FME2	12.5
		TAT3	13.3			NFM	7.4			ATF4	11.3
		ATF3	12.8			PAV	7.0			FME7	11.2
		TAT1	11.9			FME7	6.6			NFMC	9.3
		CITY	10.9			FAS	6.5			FAS	9.3
		PAV	9.1			ATF4	5.1			PAV	9.0
		GEN	4.5			GEN	2.7			GEN	5.5

(a) QMS12		**(b) QMS12**		**(a) QMS13**		**(b) QMS13**		**(a) QMS14**		**(b) QMS14**	
FME8	100.0	FME8	100.0	AGE	100.0	EDU	100.0	ATF3	100.0	EDU	100.0
TAT1	91.6	FME5	78.3	EDU	83.2	AGE	81.8	BEC	65.9	BEY	97.6
ATF3	52.6	EDU	68.9	FME8	51.7	FME8	70.6	BEY	61.7	ATF3	83.4
GRA	33.3	TAT1	67.5	BEC	47.5	CITY	58.4	FME5	51.0	BEC	76.5
FME4	30.7	FME4	56.8	ATF3	45.8	FME5	48.1	BAC	50.7	BAC	75.9
TAT3	29.8	GRA	51.6	TAT5	34.5	BEC	43.5	FME4	40.4	FME5	70.9
TAT5	25.1	BEY	45.8	FME5	29.0	FME4	38.6	FAS	40.1	FME6	58.6
FME5	24.2	ATF3	45.7	FME4	26.3	BEY	26.1	FME6	37.9	FME8	58.2
ATF2	24.2	AGE	34.8	BLS	23.0	BAC	22.5	FME8	27.5	FME4	57.9
TAT4	15.7	NFM	31.5	TIM	22.8	BLS	20.3	TAT3	24.2	PAV	57.8
NFM	15.1	FME3	31.0	GRA	18.4	FME3	19.5	ATF4	23.2	AGE	53.2
FME6	10.7	BEC	28.2	FME6	14.8	ATF3	16.9	PAV	18.1	FME3	38.6
EDU	7.4	ATF2	28.0	NFM	14.2	FME1	16.4	TAT1	15.2	GRA	35.6
BAC	7.3	CITY	25.7	CITY	13.7	TAT5	15.8	ATF1	14.1	FAS	34.2
CITY	5.7	BAC	22.2	TAT6	12.7	GRA	15.7	AGE	11.0	TAT2	30.6
TAT6	5.5	TAT3	22.2	TAT4	11.4	ATF2	15.6	TAT2	9.8	FME1	22.1
FME2	4.2	TAT5	19.6	FME2	10.7	TIM	14.9	FME2	9.3	FME2	21.9
FME3	3.6	TIM	18.2	ATF2	10.3	FME6	14.2	GRA	8.0	TAT6	20.4

NFMC	0.2

BLS	17.5
FME1	17.1
TAT4	15.8
FME6	15.3
ATF1	14.0
PAV	13.4
TAT2	11.9
TAT6	11.8
FME7	11.6
NFMC	11.4
ATF4	8.9
FME2	8.4
FAS	7.0
GEN	5.1

BEY	9.4
ATF1	7.7
TAT3	7.2
NFMC	6.2
TAT1	6.1
ATF4	4.9
FME3	3.8
PAV	1.6
FAS	1.5
FME7	0.5

TAT3	13.0
NFM	12.7
FME2	12.6
TAT6	11.1
ATF1	10.7
FME7	10.3
TAT2	10.2
ATF4	9.1
NFMC	8.9
TAT4	8.9
PAV	7.6
TAT1	7.5
FAS	7.5
GEN	3.5

EDU	4.1
TIM	3.1
TAT5	2.3
FME7	0.5

BLS	19.7
ATF4	19.1
TIM	18.0
TAT3	16.9
ATF2	15.7
NFMC	15.4
CITY	14.3
FME7	14.1
ATF1	13.1
NFM	12.5
TAT4	12.4
TAT1	11.3
TAT5	7.7
GEN	4.1

(a) QMS15 **(b) QMS15** **(a) QMS16** **(b) QMS16**

BEY	100.0
BEC	99.6
FME8	82.2
TAT2	50.5
FME3	47.3
NFM	35.7
FME4	21.4
TIM	19.5
PAV	15.2
NFMC	14.1
BLS	11.9
ATF1	11.7
ATF4	11.2
TAT4	7.3
GRA	6.5
TAT6	6.3
TAT3	6.2
FME2	4.9
FME6	4.4
FAS	3.8
GEN	2.8
AGE	2.5
EDU	1.8
ATF3	1.6
TAT5	1.3

BEY	100.0
FME8	78.0
BEC	70.5
FME3	61.4
EDU	58.8
TAT2	44.2
AGE	35.3
FME4	28.5
FME5	28.5
TIM	27.7
BAC	27.3
ATF1	26.1
NFM	24.0
BLS	20.5
TAT4	16.5
PAV	16.0
CITY	13.2
ATF4	12.0
FME2	11.4
FME6	10.7
ATF3	10.6
TAT3	9.7
ATF2	9.4
TAT5	9.3
FME1	9.3
TAT1	9.0
TAT6	8.4
FME7	8.2
GRA	7.8
NFMC	7.1
FAS	6.4
GEN	3.7

CITY	100.0
FME5	88.0
AGE	85.7
FME4	82.5
FME7	49.2
EDU	38.7
BEY	34.3
ATF3	25.5
FME3	24.6
NFM	22.0
BAC	14.3
TAT5	10.1
FME6	9.6
TAT3	6.2
FME8	5.1
FAS	4.8
TAT4	4.4
ATF1	4.0
PAV	3.7
BEC	3.2
TAT2	3.0
TIM	2.9
BLS	2.8
TAT1	2.5

CITY	100.0
FME5	86.0
FME4	82.4
AGE	80.1
EDU	74.7
FME7	52.0
FME8	44.7
FME3	44.4
BEC	36.2
BEY	33.6
BAC	31.6
ATF3	29.2
GRA	15.9
NFM	15.0
PAV	14.6
FME1	11.9
FAS	10.7
TAT3	10.6
FME6	10.4
TIM	10.4
BLS	8.9
TAT5	8.8
TAT2	8.7
TAT6	8.7
TAT1	7.9
TAT4	7.3
ATF1	6.6
ATF4	6.5
NFMC	5.2
ATF2	4.8
FME2	4.7
GEN	4.6

Source: authors