*Journal of
Computational and Theoretical Nanoscience*
Vol. 16, 3483–3488, 2019

An Empirical Study on Loan Default Prediction Models

Uzair Aslam, Hafiz Ilyas Tariq Aziz, Asim Sohail, and Nowshath Kadhar Batcha*

School of Computing, Asia Pacific University of Innovation and Technology, 54200, Malaysia

Loan lending has been playing a significant role in the financial world throughout the years. Altho it is quite profitable and beneficial for both the lenders and the borrowers. It does, however, carry a great risk, which in the domain of loan lending is referred to as Credit risk. Industry experts and Researchers around the world assign individuals with numerical scores known as credit scores to measure the risk and their creditworthiness. Throughout the years machine learning algorithms have been used to calculate and predict credit risk by evaluating an individual's historical data. This study reviews the present literature on models predicting risk assessment that use machine learning algorithms.

Keywords: Artificial Neural Networks, Credit Default, Credit Risk, Credit Scoring, Machine Learning Loan Approval Prediction.

1. INTRODUCTION

Loan lending has been an important part of daily lives for organizations and individuals alike. With the ever-increasing competition in the financial world and due to a significant amount of financial constraints, the activity of taking loan has become more or less inevitable [1]. Individuals around the world depend on the activity of loan lending for reasons such as overcoming their financial constraints in order for them to achieve some personal goals. Similarly, small to large firms depend on the activity of loan lending for the basic purpose of managing their affairs and to function smoothly in times where there are financial constraints [2].

Though loan lending is quite beneficial for both the lenders and the receivers and is considered an essential part of the financial organization, it does carry some great risks [3]. This type of risk represents the inability of the receiver to pay back the loan at the designated time which was decided upon by the lender and the borrower, during the loan origination and is referred to as Credit risk [1]. Credit risk is known to cause major concerns among the financial institutes as it can result in a dire situation known as credit defaulting which can prove to be drastic or the lending party [4]. Even after the defined risk, financial institutes around the world consider the activity of loan lending as a major opportunity for profit-making and essential for the smooth functioning of their businesses [5].

According to a study conducted, although profit and risk have a directly proportional relationship, the risk factor in the domain of loan lending is on the rise [6]. Effective and thorough evaluation of credit risk can result in a scenario where there is a minimalized risk of defaulters for the financial institutes [7]. Throughout the financial world, lenders try to minimize the credit risk by thoroughly evaluating and verifying the ability of a borrower to deliver on their obligation of repaying the loan.

While in the past, the financial institutes focused on employing highly professional individuals whose sole purpose was to evaluate whether a candidate was eligible for receiving a loan based on two factors. For risk evaluation which help in verifying the worthiness of a candidate for loan approval or rejection based on a numerical score [8].

The process of credit scoring, though in the past required experts alongside statistical algorithms to accurately predict the worthiness of a candidate for loan approval or rejection. Quite recently, however, the researchers and institutes have opted for training classifiers based on machine learning algorithms and neural networks to automatically predict the credit score of an individual based on their historical data and sift through credit defaulters from the lot before the loan is approved [1–5].

The purpose of this study is to bifold thoroughly the review of the literature on Loan default prediction and credit scoring: to present future directions to researchers in the domain of Loan default prediction. The rest of this study is structured in the following format, the next section will talk about the different models implemented in the past for predicting loan approval candidates, and the third section will present the discussion and conclusion.

*Author to whom correspondence should be addressed.

2. LITERATURE REVIEW

Credit scoring has become an essential tool in the highly competitive financial world, which has brought more focus towards research on credit risk assessment in the recent years [9]. Due to the high demand and reliability of financial institutes on loan lending, there is a significant demand for further improvements of the models for credit scoring [2]. There have been a multitude of techniques which were used to assign credit scores and much research has been done on the topic throughout the years. Unlike before, where the initial models depended on professional opinions for assessing the loan worthiness of an individual, recently focus has shifted towards applying advanced machine learning algorithms and neural networks for credit scoring and risk assessment. These techniques can be further classified into two categories: the traditional statistical techniques and advanced statistical techniques [10].

2.1. Traditional Statistical Techniques for Loan Default Prediction

Throughout the reviewed literature, researchers have employed a multitude of statistical techniques for loan default prediction such as discriminant analysis, Probit analysis and linear regression [11]. However, the three most accurate and prominent of these are discussed below.

2.1.1. Support Vector Machine

Support vector machines have emerged as a strong candidate for classification in the domain of neural networks and machine learning. According to the research, support vector machines are supervised learning models which utilize the concepts of most learning algorithms for analyzing binary classes [12]. An SVM model comprises of a hyperplane which is used to separate the different observations for classification purpose [13].

A hybrid model was proposed by the researchers in the study where they combined Naïve Bayesian algorithm along with Support Vector machine to develop a model for predicting Loan defaulting. The researchers collected a German dataset from UCI which comprised of 21 different attributes. Although Naïve Bayesian is considered quite robust and is famous for its simplicity in classification problems, it takes a long time in execution. Whereas Support vector machine is known for its fast execution time and accurate results in classification. The researchers conducted several experiments by using only Naïve Bayes, SVM and then both of them combined. Their result showed that although Naïve Bayes had an accuracy of 77%, it fell a bit short in both accuracy and execution time as compared to SVM. Whereas SVM and the combined NBSVM model had the same accuracy of 79%. While the accuracy was the same, NBSVM executed twice as fast as the SVM model. The researchers concluded that though each of these classifiers is great separately, and each has its own pros, using both in combination enhanced the model accuracy and time 2-fold [14].

The Researchers in the study proposed a model where they analyzed the results for predicting loan defaults by using Linear kernel Support vector machine and Quadratic Discriminant Analysis. The researchers concluded that when it comes to the overall efficiency LSVM had the most success rate out of the two models. However, in terms of computational execution time, LSVM took much longer than QDA [15]. Experimenting with a large number of explanatory variables in a dataset of about 5000 instances, the researchers noticed that LSVM was performing much better and showed a higher rate of success in predicting loan defaults. Whereas with a smaller number of explanatory variables there was a little difference among the performance of both models. However, the difference was significantly unnoticeable. In a later study, performed multiple experiments using a range of explanatory variables from 1 to 250. They built models using different classifiers such as QDA, LSVM, generalized linear model with a logit link (GLMLOGIT) and neural networks. Each classifier was tested for its efficiency and effectiveness, and their results concluded that LSVM and GLMLogit outperformed all others when it came to computational efficiency and execution time [16].

2.1.2. Logistic Regression

Logistic regression is one of the most popular statistical techniques among others in the financial world for credit risk assessment models. The major strengths of a logistic regression model lie in its simple understanding, sturdy performance and easy implementation [17]. Moreover, logistic regression outperforms linear regression as it overcomes multiple issues, such as in linear regression the output of the regression can be a negative value or greater than the value 1, which is not possible for probability. Logistic regression solves this by providing a continuous range of grades between 0 and 1 and keeping the output limited to values between 0 and 1 [18].

Researchers have proposed a model for predicting loan defaulters for the loan lending program "Korean Student Aid Foundation" (KOSAF). They employed two different classifiers namely logistic regression and Cox proportional hazard algorithm to accurately predict students who were going to default on their loans. The study was aimed at finding the major factors that can influence a loan default among students. They collected the data directly from KOSAF from 2012 till 2014. The data used in the study comprised of a training set of 127,432 loans out of which 125,291 were labeled good loans and 2141 were labeled bad. Similarly, the test set contained 83,560 good loans and 1480 bad ones. Their findings indicated that major influencing factors that led to students defaulting on their debt included age, household income, monthly repayable amount and field of study. Using logistic regression, the suggested that those students that are in their mid to late 20's are less likely to default on a loan as compared to those in the late teens or early 20's. Not only were they able to develop a strong and accurate model using logistic regression, rather their model attained an AUC of 0.697 for the testing data [19].

In a similar study proposed a model based on logistic regression which accurately predicted loan defaulters for a bank in Ghana where 30% of the population lives under the poverty line. Their study was aimed to identify the risk factor which influences individuals living in Ghana to default on their loans. The research utilized the data from a microfinance company in Ghana which comprised of a historical data regarding their customers' profiles. In particular, the dataset contained information on about 500+ customers within the interval of 2013 till 2014. The researcher chose to employ logistic regression for the prediction of loan defaulting as linear regression was violating the normality among the attributes. Their logistic regression-based model perfectly predicted loan defaulters with an accuracy of 86% [20].

2.1.3. Decision Trees

One of the most popular techniques used for classification used in the domain of credit scoring is known as decision trees which comprise several branches, root nodes, and leaf nodes. As the name implies, the technique generates a structure similar to a tree by classifying the instances and utilizing an (RPA) or recursive portioning algorithm [18]. Each leaf node in the tree represents a class label and all the branches in the tree represent the outcomes for the test, these tests are represented by internal nodes for an attribute [2].

Researchers in the study proposed a two-step risk assessment model, where decision trees-based classifier is used to predict whether the customer is worthy for loan approval or not. The study employed a cross-industry standard process as a methodology for building the model. Using a dataset from various midsized banks in India such as IDBI, syndicate banks and AXIS the researchers optimized their data using support vector machine. The dataset was of a moderate size, having only 1140 instances which comprised multiple dependent variables such as age, sex, marital status, loan type etc. and only a single dependent variable which was whether the credit was approved or not. The researchers used a C4.5 algorithm for decision trees in Weka software to predict the loan approval and individual default risk. Their classifier takes customer credit scoring and behavior as an input and made a decision on whether the customer should be accepted for a lending loan or not [21].

In a similar study, the researchers proposed a model where they applied and compared the results from both logistic regression and decision trees in order to predict the repayment risk for small businesses. The study collected a dataset containing 11 million observations from Equifax which included information from unique organizations. After random sampling and dimensionality reduction of the dataset using clustering, the researchers were able to extract a random sample of 100,000 instances where they split it 60% and 40% for training and testing purposes respectively. Two models were built using logistic regression and decision trees in SAS on the training set. After experimentation and testing it was noted that due to the smaller number of attributes in a large enough sample, decision trees outperformed logistic regression. For the univariate splitting at each node, an impurity-based criterion known as Gini index was used. For the stopping criteria, the researcher set the maximum tree depth value to 6. After comparison, it was noted that having only 10 predictors of decision trees had an accuracy of 96.6% with the lowest percentage of false negative when compared to logistic regression [22].

2.2. Advanced Techniques for Loan Default Prediction

Apart from the traditional statistical techniques, the present research has opted for employing advanced techniques using Artificial intelligence based on neural nets and genetic programming [3–24]. These techniques have proved to be highly accurate and robust candidates in the domain of credit risk assessment and loan defaulter prediction [10].

2.2.1. Neural Nets

Neural networks have been the focus of research in multiple scientific domains across the globe due to their enhanced capability of easily approximating complex function [2]. A neural network generally consists of small processor units, referred to as neurons or nodes. A neural network comprises of three different types of layers, Input layer receives independent variables as inputs and feeds them to a hidden layer where they are processed and then forwarded to the output layer. In the domain of loan lending, Artificial neural networks have received significant attention for credit scoring and loan default prediction [5–23].

A recent study proposed a novel method of loan default prediction using artificial neural networks as classifiers. The researcher collected a dataset from one of India's leading bank known as "lending club bank." Due to the large unstructured form of the data where originally the dataset contained about 9500+ rows and 14 columns, the researchers used Principle component analysis (PCA) for dimensionality reduction and converted the number of columns to 19. The sampling of the data was done in such a way that 90% of the dataset was kept for training purposes while only 10% was kept for testing the model. The study then employed Multi-layered perceptron neural network for the purpose of predicting loan defaulters within the dataset. By feeding each attribute to the input layer of the multi-layered perceptron the neural network is structured as such that the 18 initial perceptron's are directly connected to 20 immediate perceptron's. The researcher added 2 hidden layers with 20 perceptron connected to each layer where the final layer is connected directly to the output. This resulted in a model that was able to predict loan defaulters with 93% accuracy. The researchers concluded that not only did the model outperform existing and traditional model, but it had significant room for improvement and the performance could be enhanced by adding more hidden layers [25].

In a similar study, researchers proposed a model which employs neural networks for classifying borrowers into two categories i.e., defaulters and non-defaulters, based on some explanatory variables. The study collected a dataset from a loan lending company in Europe by the name of "Bandora." The dataset had both default and non-default loans with approximately 16037 observations. 15% of these observations were those of defaulted loans while the other 85% were of non-defaulted loans. After sampling the data into 70 to 30 ratios for training and testing set respectively, the researcher fed the 15 attributes into the input layer of the neural network. The input layer had 14 neurons which were connected to 5 neurons in the hidden layer, while the output layer had only 1 neuron respectively. The model predicted defaulters on the training and testing set with an accuracy of 74%. Their results further indicated that attributes such as marital status have little to no influence on the probability of loan default, whereas, the attribute "new offer made" had the most significant importance and most influence on the result. The researchers then benchmarked their model using AUC and accuracy against logistic regression. They noted that in terms of predicting non-default loans, logistic regression showed promising results and had high success rates. While in terms of loan defaults neural network outperformed logistic regression and achieved an accuracy of 74% [23].

2.2.2. Genetic Programming

One of the more recent techniques in predicting loan default is called genetic programming and is considered to be an extension of the genetic algorithms [10]. These algorithms apply genetic operations and based on the fitness values, transform a dataset. The idea of genetic programming originated with Darwin's theory of natural selection and evolution. Genetic programming utilizes various processes of crossovers to randomly generate a set of competing programs [26].

The researchers in the study applied a novel approach for credit scoring by combining genetic programming with deep neural networks. The study utilized two different datasets from UCI namely, Australian credit data which had about 690 observations out of these about 307 were non-loan defaults and the rest were loan defaults and a German credit dataset. The data consisted of 14 attributes in total. Their model was structured in such a way that the genetic algorithm was used to discover hidden patterns among the various attributes and use them as classification rules. These extracted rules would play the role of a filtering mechanism in the testing phase in the first layer of the neural network. They employed the stacked Autoencoder neural network by using a technique called greedy layer approach which automatically constructs a deep neural network with multiple hidden layers based on the outcome. The model once benchmarked against traditional statistical techniques proved to be quite robust and outperformed all other models [27].

3. DISCUSSION

The ever-increasing demand for loan lending has made credit scoring as one of the most significant tools in the financial world. This significance has paved the way for a lot of research in the domain of loan default prediction [9]. Since banks and financial institutes rely heavily on financial lending for self-sustenance and profit-making, there is an ever-growing demand for the improvement of base models for credit scoring [2].

Previously lenders and financial institutes employed highly professional individuals to evaluate a candidate's worthiness before approving or rejecting a loan. However, recently these institutes have started employing various models for loan evaluation in order to decide whether to reject or approve a loan to a borrower based on their credit score and ability to repay [8]. The models that they have opted for are based on machine learning algorithms along with artificial neural networks for accurately predicting credit defaulters among the borrowers [1–5]. These models typically assign a numerical score, which represents the creditworthiness of the borrowers based on historical data [17].

As per Support vector machine is quite famous among researchers in the domain for loan lending due to its ability of classification with highly accurate results [13]. When compared to other classifiers it was noted in study that support vector machine outperformed most of them when it came to effectiveness and computational execution where large datasets with numerous explanatory variables were involved [16]. However, the major disadvantage highlighted in the mentioned studies that when there are less than 10 explanatory attributes the accuracy of SVM is reduced, and other classifiers such as QDA, and Logistic Regression outperformed SVM.

The main advantage of using Logistic regression according to the study is the simplicity in interpreting results by utilizing odds ratios [19]. Moreover, Logistic Regression is a powerful candidate for models which rely on binomial outcomes using multiple explanatory attributes. Similarly, researchers in another indicated another advantage of using logistic regression [20]. They concluded that since normal regression violates the assumption of normality where there is a categorical output variable, logistic regression

Table I. Comparison of techniques and their limitations.

Paper	Classifier	Performance metrics	Score	Limitations
Perez Martin and Vaca Lamata (2017)	LSVM	RMSE	–	Loosses performance where there are less than 10 explanatory variables
Pérez-Martín, Pérez-Torregrosa and Vaca (2018)	LSVM + LR		–	
Vimala and Sharmili (2018)	SVM + Naïve Bayes	Accuracy	79%	
Han, Choi and Kim (2018)	Logistic regression		0.697	Looses perfomance in datasets with large number of variables due to on linearity relation between them
Agbemava et al. (2016)	Logistic regression		86%	
Sudhakar, Reddy and Pradesh (2016)	Decision trees		89%	Unstable as a small change in the data can output a moajor change in the structure of the tree
Wang et al. (2017)	Decision trees		96.60%	
Kumar et al. (2018)	Neural network		93%	Requires larger training sets to give accurate results
Byanjankar, Heikkila and Mezei (2015)	Neural network		74%	
Tran, Duong and Ho (2016)	Genetic programming + NN	RMSE	0.13 + −0.0003	Over-fitting issue + There's no guarantee of finding global maxima

tackles the situation by allowing the model to portray the nonlinear output in a linear way within the bounds of 0 and 1. As per the previously reviewed study, the major benefits of using decision trees for credit risk assessment is that by employing them, one doesn't need much knowledge of the domain as the least amount of human intervention is needed [22]. Moreover, decision trees perform best for binary classifications on a larger dataset with only a few predictors. Whereas logistic regression loses its power when the dataset contains many variables, this is due to the non-linearity of relations between the variables.

Throughout the literature, it was clear that neural networks outperform traditional statistical techniques in cases where there was a large amount of training data available [23–25]. Due to their robust ability to approximate complex functions and detect hidden relations among attributes, neural networks outperformed traditional techniques. However as concluded by the researchers in study, in cases where there was a small amount of training data available neural networks lagged in comparison to the performance of traditional techniques such as SVM, logistic regression and Naïve Bayes [16]. Whereas, the main advantage of using genetic programming according to the study is its ability to use probabilistic selection rules, whereas other traditional methods use deterministic rules. Furthermore, their concept is easy to understand, and they support multiple objectives [28]. Without using any other auxiliary information, genetic programming uses fitness scores which they get directly from the objective function [27].

Throughout the literature it was noticed that most of the studies tuned and optimized their models to attain maximum accuracy for predicting credit scores of borrowers, however only a small number of the reviewed studies discussed the repercussions of misclassifications of false negatives which is summarized in Table I. As discussed before, misclassification of false negatives in the domain of Credit risk assessment can prove to be fatal for the lending company and the financial institutes [29].

4. CONCLUSIONS

Throughout the year's loan lending has played an important role in the world of finance. And in that domain credit scoring is an essential tool for managing credit risk. Industry experts and researchers around the globe utilize machine learning to predict credit risk using credit scoring. The study reviewed the extensive literature on the subject. The past studies employed multiple machine learning technique such as Decision trees, Neural Networks, logistic regression and support vector machine. Each machine learning algorithm proved to be accurate and useful in different settings. And although most studies focused on accuracy for the prediction of loan default detection, it was noticed that few studies focused on the repercussions of false negatives which can be quite devastating to the lending companies. It is therefore advised to the future researchers to focus more towards false negatives when approaching the loan lending prediction problem.

References

1. Perera, H.A.P.L. and Premaratne, S.C., **2016**. An Artificial Neural Network Approach for the Predictive Accuracy of Payments of Leasing Customers in Sri Lanka.
2. Marqués, A.I., García, V. and Sánchez, J.S., **2012**. Exploring the behaviour of base classifiers in credit scoring ensembles. *Expert Systems with Applications, 39*(11), pp.10244–10250.

3. Adewusi, A.O., Oyedokun, T.B. and Bello, M.O., **2016**. Application of artificial neural network to loan recovery prediction. *International Journal of Housing Markets and Analysis, 9*(2), pp.222–238.

4. Choudhary, G., Garud, Y., Shetty, A., Kadakia, R. and Borase, S., **2019**. Loan Default Identification and Its Effect.

5. Atiya, A.F., **2001**. Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on Neural Networks, 12*(4), pp.929–935.

6. Li, H., McCarthy, J. and Pantalone, C., **2014**. High-yield versus investment-grade bonds: Less risk and greater returns? *Applied Financial Economics, 24*(20), pp.1303–1312.

7. Pandit, A., **2016**. Data Mining On Loan Approved Datset For Predicting Defaulters (Doctoral dissertation, Rochester Institute of Technology).

8. Calcagnini, G., Cole, R., Giombini, G. and Grandicelli, G., **2018**. Hierarchy of bank loan approval and loan performance. *Economia Politica, 35*(3), pp.935–954.

9. Sarma, K.S., **2013**. *Predictive Modeling with SAS Enterprise Miner: Practical Solutions for Business Applications*. SAS Institute.

10. Abdou, H.A. and Pointon, J., **2011**. Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance and Management, 18*(2–3), pp.59–88.

11. Schneider, A., **2018**. Studies on the impact of accounting information and assurance on commercial lending judgments. *Journal of Accounting Literature, 41*, pp.63–74.

12. Arun, K., Ishan, G. and Sanmeet, K., **2016**. Loan Approval Prediction based on Machine Learning Approach. *National Conference on Recent Trends in Computer Science and Information Technology (NCRTCSIT-2016)*, pp.18–21.

13. Nehrebecka, N., **2018**. Predicting the default risk of companies. Comparison of credit scoring models: LOGIT versus support vector machines. *Econometrics, 22*(2), pp.54–73.

14. Vimala, S. and Sharmili, K.C., **2018**. Prediction of Loan Risk Using Naive Bayes and Support Vector Machine. *Int. Conf. Adv. Comput. Technol. (ICACT)*, Vol. 4, pp.110–113.

15. Pérez-Martín, A. and Vaca, M., **2017**. Computational experiment to compare techniques in large datasets to measure credit banking risk in home equity loans. *International Journal of Computational Methods and Experimental Measurements, 5*(5), pp.771–779.

16. Pérez-Martín, A., Pérez-Torregrosa, A. and Vaca, M., **2018**. Big data techniques to measure credit banking risk in home equity loans. *Journal of Business Research, 89*, pp.448–454.

17. Nalić, J. and Švraka, A., **2018**. Using Data Mining Approaches to Build Credit Scoring Model: Case Study—Implementation of Credit Scoring Model in Microfinance Institution. *2018 17th International Symposium Infoteh-Jahorina (INFOTEH)*, IEEE. pp.1–5.

18. Baesens, B., Roesch, D. and Scheule, H., **2016**. *Credit Risk Analytics: Measurement Techniques, Applications, and Examples in SAS*. United States, John Wiley & Sons.

19. Han, J.T., Choi, J.S., Kim, M.J. and Jeong, J., **2018**. Developing a risk group predictive model for korean students falling into bad debt. *Asian Economic Journal, 32*(1), pp.3–14.

20. Agbemava, E., Nyarko, I.K., Adade, T.C. and Bediako, A.K., **2016**. Logistic regression analysis of predictors of loan defaults by customers of non-traditional banks in Ghana. *European Scientific Journal*, *12*(1), pp.175–189.

21. Sudhakar, M. and Reddy, C.V.K., **2016**. Two step credit risk assessment model for retail bank loan applications using decision tree data mining technique. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 5*(3), pp.705–718.

22. Wang, Y. and Priestley, J.L., **2017**. Binary classification on past due of service accounts using logistic regression and decision tree, Grey Literature from PhD Candidates.

23. Byanjankar, A., Heikkilä, M. and Mezei, J., **2015**, Predicting Credit Risk in Peer-to-Peer Lending: A Neural Network Approach. *2015 IEEE Symposium Series on Computational Intelligence*, December; IEEE. pp.719–725.

24. Jin, Y. and Zhu, Y., **2015**. A Data-Driven Approach to Predict Default Risk of Loan for Online Peer-to-Peer (P2P) Lending. *2015 Fifth International Conference on Communication Systems and Network Technologies*, IEEE. pp.609–613.

25. Angelini, E., di Tollo, G. and Roli, A., **2008**. A neural network approach for credit risk evaluation. *The Quarterly Review of Economics and Finance, 48*(4), pp.733–755.

26. Metawa, N., Elhoseny, M., Hassan, M.K. and Hassanien, A.E., **2016**. Loan Portfolio Optimization Using Genetic Algorithm: A Case of Credit Constraints. *2016 12th International Computer Engineering Conference (ICENCO)*, December; IEEE. pp.59–64.

27. Tran, K., Duong, T. and Ho, Q., **2016**. Credit Scoring Model: A Combination of Genetic Programming and Deep Learning. *2016 Future Technologies Conference (FTC)*, IEEE. pp.145–149.

28. Metawa, N., Hassan, M.K. and Elhoseny, M., **2017**. Genetic algorithm based model for optimizing bank lending decisions. *Expert Systems with Applications, 80*, pp.75–82.

29. Amit, R. and Zott, C., **2015**. Crafting business architecture: The antecedents of business model design. *Strategic Entrepreneurship Journal, 9*(4), pp.331–350.

Received: 24 October 2018. Accepted: 19 February 2019.