Managing Global Transitions

EDITOR
Maja Meško

International Research Journal

Credit Risk Scoring in Entrepreneurship: Feature Selection

Mirjana Pejić Bach
Ekonomski fakultet Zagreb, Croatia
mpejic@efzg.hr

Nataša Šarlija
Ekonomski fakultet Zagreb, Croatia
natasa@efos.hr

Jovana Zoroja
Ekonomski fakultet Zagreb, Croatia
jzoroja@efzg.hr

Božidar Jaković
Ekonomski fakultet Zagreb, Croatia
bjakovic@efzg.hr

Dijana Ćosić
Wealthengine, Washington DC, USA
dijana.cosic@gmail.com

The goal of this research is to investigate the impact of different algorithms for the feature selection for the purpose of credit risk scoring for the entrepreneurial funding by the Croatian financial institution. We use demographic and behavioral data, and apply various algorithms for the development of classification model. In addition, we evaluate several algorithms for the variable selection, which are additionally based on the classification accuracy. Sequential Minimal Optimization algorithm in combination with the Class CfcSubsetEval and ConsistencySubsetEval algorithms for variable selection was the most accurate in predicting credit default, and therefore the most useful for the credit risk scoring.

Key Words: data mining, credit scoring, variable selection, decision tress, classification

Introduction

Data mining methods are used to find undiscovered valuable information from large databases. In other words, the main goal of data mining

techniques is to extract knowledge in order to make successful management decisions (Wu et al. 2012). Applications of data mining methods are used in almost every industry: banking, marketing, finance, manufacturing, medicine, education, trade, supply (Wei et al. 2013; Lejeune 2001; Choudhary, Harding, and Tiwari 2008). Each industry has its own characteristics, which implies the usage of different data mining methodologies. Therefore, in the banking industry, characterized with a high level of fraud and risks, which requires successful prediction of credit default, scoring, and applicants, usage of data mining techniques is very common (Ngai et al. 2011).

Data mining is one of the most common techniques used in financial analysis, especially in the banking industry. Prediction of credit risk, mostly prediction of credit default, presents an important activity of the banking industry (Thomas et al. 2005). There are several different data mining techniques that can be used for financial data analysis because of their high level of success. However, their success also depends on the data available, its cleaning, and transformation. Decision trees are one of the most commonly used methods (Quinlan 1992; Breiman et al. 1984). Decision trees are one of the classification methods which group variables into one or more categories of the target variables (Yap, Ong, and Husain 2011). When using the decision trees process it is important to follow three main steps: (i) determine the sample, (ii) choose variables, and (iii) select an appropriate algorithm.

In this paper we use ten data mining algorithms in order to develop a credit scoring system for the classification of banking clients according to the credit default, using different sets of the variables: Entrepreneurial idea; Growth plan; Marketing plan, Personal characteristics of entrepreneurs, Characteristics of SME, Characteristics of credit program, and Relationship between the entrepreneur and a financial institution.

The variables are selected by the usage of three different algorithms, provided in the Weka software: Class CfsSubsetEva algorithm, ChiSquaredVariableEval algorithm, and ConsistencySubsetEval algorithm. Previous research that tested the efficiency of algorithms for the selection of variables was mostly conducted on the retail credit risk datasets (Oreski, Oreski, and Oreski 2012). The scientific contribution of our paper is that the algorithms for the selection of variables are tested on the real-world dataset of Croatian financial institution's business clients (entrepreneurs from Eastern Croatia).

The paper consists of six sections. After the Introduction, as the first section, there is a literature review. In Literature review, data mining methodology and its usage for predicting credit default are presented. Decision trees, as one of the data mining methods, are described as well as variables and techniques selection approaches used in this research. In the third section named Methodology, data, decision trees techniques and the variable selection process are discussed. Research results are provided in the fourth and fifth section. The fourth section elaborates on results of the different variable selection strategies, while the fifth section of the paper discusses results regarding classification efficiency measures, classification matrices and falsely predicted good and bad debtors with different variable selection approaches. The last section is Conclusion.

Literature Review

DATA MINING METHODOLOGY

The amount of data has been constantly increasing, which creates difficulties for managers and successful decision making. Fast increase of valuable as well as invaluable data in databases has created a need for the use of different methodologies which can help to find, extract and analyze the data important for decision makers (Priya and Ghosh 2013).

Data mining technology combines different approaches, e.g. machine learning, statistics and database management, which are used for finding valuable patterns in data for further prediction and decision making. In addition, data mining techniques can also be used for determining relationships among data in order to create knowledge (Ngai, Xiu, and Chau 2009). The main purpose of data mining is to find and analyze disorganized information with the goal of improving business knowledge and activities.

The most commonly used data mining methods are classification, regression, clustering, visualization, decision trees, association rules, neu-

ral networks, support vector machine (Ngai et al. 2011; Strohmeier and Piazza 2013; Patel and Sarvakar 2014).

PREDICTING CREDIT DEFAULT WITH THE DATA MINING APPROACH

Countries, especially their economies and financial institutions, have been facing a strong financial crisis in the last years. Therefore, in many countries, governments have brought many saving measures in order to decrease costs and to restart economy development. In addition, credit default has increased, and nowadays banks pay much more attention to credit risk assessment and to prediction of credit default with the goal to reduce risk (Marinakis et al. 2008).

Financial institutions and banks are using different intelligent techniques, e.g. mathematical models, statistics analysis, data mining methods with the goal to make efficient credit decisions. A detailed analysis of data on the characteristics of current and previous credit users plays an important factor in forecasting the future credit default of new clients (Thomas 2000).

VARIABLES AND TECHNIQUES SELECTION APPROACHES

In order to predict credit default, financial institutions and banks mostly use behavioral and demographic variables of previous and current clients, e. g. monthly income, marital status, real-estate owner, employment, age, gender (Lucas 2001).

The main purpose of our research is to classify banking clients regarding credit default with a decision tree analysis, using different variables related to entrepreneurship activity. In addition, financial institutions and banks, when approving credits to clients, strive to select those clients who will be able to repay it in the given period of time (Wu et al. 2008). In other words, they are focused on good clients.

There are also studies about methods used in credit scoring. One of the examples is the research which used demographic and behavioral data and three data mining methods: credit scorecard, logistic regression and a decision tree model (Yap, Ong, and Husain 2011). The results of the research showed that all three methods are appropriate for use, but the scorecards method is the easiest to apply.

There is also a study in which authors have investigated recent researches conducted in the field of credit risk assessment regarding clients and their ability to repay credits (Crook, Edelman, and Thomas 2007).

Research results showed that logistic regression is the most commonly used method to group clients into good or bad debtors.

Recent studies showed that intelligence methods used for discovering credit scoring are mostly non-parametric methods and computational intelligence techniques, e.g. decision trees, artificial neural networks, support vector machines and evolutionary algorithms (Zhang, Leung, and Ye 2008; Pourzandi and Babaei 2010; Lucas 2001).

Methodology

DATA

Data used in this research was collected from an entrepreneurship credit dataset. Data were collected randomly from the database of clients (entrepreneurs from Eastern Croatia) of the financial institution that is focused on financing small and medium enterprises, mostly start-ups. There are 200 applicants in the sample.

There are two main reasons for a small dataset: (i) a quite low level of business activity regarding entrepreneurship in Croatia (Total Early State Entrepreneurial Activity (TEA) index for the year 2015 = 7.69; TEA index for the year 2004 = 7.97) which means that a low number of people is taking a credit to start a business, and (ii) financial institutions are rejecting too risky start-ups applications for a credit. Therefore, collecting a larger sample will be possible in the next few years, when the entrepreneurial climate and perception of entrepreneurship activity improves.

The following variables were used for the development of the credit scoring model: Entrepreneurial idea; Growth plan; Marketing plan, Personal characteristics of entrepreneurs, Characteristics of SME, Characteristics of credit program, Relationship between the entrepreneur and a financial institution, and Creditworthiness. Most of the variables are nominal, while two of them are numeric (Entrepreneurs' age, Number of employees, and Credit amount).

Variables related to the future plans for the SME were estimated by a banking clerk (table 1). First, it was estimated whether the entrepreneur has a clear vision of business development (for newly established SMEs), or it is already established business (Variable Vision). Second, the variable Better estimated what the main competitive advantage of the SME (better quality, technology, price, or expertise of employees) is. Third, it was estimated what the main market for SME's products/services is: local, national, wider region, or the narrowly targeted customers (Variable Mar-

TABLE 1 Variables Related to the Future Plans For the SME

Category	Variable	Question asked	Answers
Entrepre-neurial idea	Vision	Does the entrepreneur have a clear vision of the business?	1 – no clear vision (for newly established SMES) 2 – clear vision (for newly established SMES) 3 – established business
	Better	Advantages of products/services	1 – better quality 2 – better technology 3 – good price 4 – expertise of employees 100 – no answer
	Market	Market for products/services	1 – local 2 – narrow targeted customers 3 – wider region 4 – Croatia 100 – no answer
Growth plan	Reinvest	Projected percentage of the invested profit (reinvested profit/profit*100)	1 – 0 to 30% 2 – 30.01 to 50% 3 – 50.01 to 70% 4 – 70.01 to 100% 100 – missing value
Marketing plan	Ad	Promotion of products/services	1 – without promotion 2 – no need for promotion 3 – all media 5 – personal selling, presentation 6 – posters, leaflets, internet 100 – missing value
	Compet	Can the entrepreneur identify competition?	1 – no competition 2 – not defined 3 – defined competition 100 – no answer

ket). Entrepreneurs stated which percentage of the profit is planned to be reinvested in the business operations (Variable Reinvest), and what the plans for the promotion of products/services are (Variable Ad). Also, it was estimated whether the entrepreneur could identify who SME's main competitors are.

Table 2 presents the variables related to the characteristics of the entrepreneur and SME. Entrepreneurs' occupations are grouped into 5 main groups: 1 – farmer, veterinarian; 2 – trader, restaurateur; 3 – construction worker; 4 – engineer, physician, and pharmacist, 5 – Technologist,

TABLE 2   Variables Related to the Characteristics of Entrepreneur and SME

Category	Variable	Question asked	Answers
Personal character- istics of en- trepreneurs	Occup	Entrepreneurs' occupations	1 – farmer, veterinarian 2 – trader, restaurateur 3 – construction worker 4 – engineer, physician, pharmacist 5 – technologist, chemist
	Age	Entrepreneurs' ages	numeric
	Location	Entrepreneurs' locations	1 – Baranja, Osijek 2 – S. Brod, Požega 3 – Đakovo, Našice 4 – Vinkovci, Vukovar
Charac- teristics of SME	Ind	Industry	1 – plastics, textiles 2 – car service 3 – food production 4 – health and intellectual services 5 – agriculture 6 – construction 13 – tourism
	Start	Is this a new busi- ness venture	1 – yes 2 – no
	Equip	Does the en- trepreneur have some equipment?	1 – yes 0 – no
	Emp	No. of employees	numeric

chemist. Entrepreneurs' ages are expressed as a numeric variable. Entrepreneurs' locations refer to 4 geographic areas in Croatia. Table 3 represents the variables related to the credit program and the bank: interest repayment frequency (monthly, quarterly, half-yearly), grace period, principal repayment, repayment period (expressed in months), interest rates, and amount of credit (expressed in local currency). Also, the variable Client measures whether an entrepreneur has applied for a credit before. Table 4 represents the classification variable that was used for the credit scoring (variable Default), and groups clients as 'bad' or 'good' based on the regularity of their payment.

ALGORITHMS FOR VARIABLE SELECTION

Three approaches to the variable selection were applied: (1) selection of the variables using the Class CfsSubsetEval algorithm (searching ap-

TABLE 3   Variables Related to the Credit Program and the Bank

Category	Variable	Question asked	Answers
Characteristics of credit program	Int	Interest repayment frequency	1 – monthly 2 – quarterly 4 – half-yearly
	Grace	Grace period	1 – yes 0 – no
	Prin	Principal repayment	1 – monthly 5 – yearly
	Period	Repayment period (months)	numeric
	I_rate	Interest rate	4.9% 6.9% 8.9%
	Amount	Credit amount (local currency)	numeric
Relationship between the entrepreneur and a financial institution	Client	Is this the first time the entrepreneur is applying for a credit?	1 – yes 2 – no

TABLE 4   Goal Variable Used for the Credit Scoring

Variable	Question asked	Answers
Default	'Bad' clients are defined as those who have been late with their payments for more than 45 days at least once. Other clients who have not been late for more than 45 days are labeled as 'good.'	1 – bad 0 – good

proach BestFirst), (2) selection of the variables using the ChiSquaredVariableEval algorithm (searching approach Ranker), and (3) selection of the variables using the ConsistencySubsetEval (searching approach Greedy Stepwise).

There are differences in definition and usage of the three mentioned approaches to the variable selection, which were applied. The Class CfsSubsetEval algorithm is based on the individual estimation of the variables that are highly correlated with the class variables but are not highly mutually correlated.

The ChiSquaredVariableEval calculates the value of a variable regarding the value of the chi-squared statistic with respect to the class. The ConsistencySubsetEval calculates the value of a subset of variables by the level of reliability in the class values (Hall 1998).

The following algorithms were used for the development of credit scoring in this paper: Bayesian Network Classifier, Sequential Minimal Optimization, Logistic regression, K-nearest neighbor's classifier (Lazy IB.K), decision trees C4.5, Gaussian radial basis function network, propositional rule learner, bootstrap aggregating algorithm, Random forest, and Adaboost. These algorithms were applied using Weka software and will be described from that perspective.

Bayesian Network Classifier is a fairly simple algorithm used in data mining for classification and prediction. The probabilities of class attribute values on all the given independent variables based on a simple normal distribution are calculated with the goal of developing a network of parenting and ancestor nodes that would be easy to understand. It assumes conditional independency as only the parent nodes can provide the information and not the ancestors, which can be written as Pr[node ancestors] = Pr[node parents] (Witten and Frank 2005). Once it learns the structure, it performs a calculation of posterior probability for each class on a vector of observed attribute values (Baesensl et al. 2003). Probabilities of discrete variables are based on their frequency and the probabilities of continuous variables are calculated by a normal or kernel density-based method (John et al. 1995). Bayes Network is implemented in software Weka under the name ByesNet (Bouckaert et al. 2013). However, Weka implementation allows only nominal variables, hence continues variables need to be discretized.

Sequential Minimal Optimization (SMO) is an algorithm for training Support Vector Machines which are a mix of linear modeling and instance-based learning. They create a discriminant that separates support vectors (a small number of critical boundary instances) from each class. The further away the data points are from the hyperplane, the better they are classified (Witten and Frank 2005). The SVM maximize the margin between support vectors which becomes a Quadriatic Programming problem (Platt 1998). SMO algorithm solves the QP by breaking it into a number of smaller QP problems which are then solved analytically, which makes it fastest for linear SVM and sparse data sets (Platt 1998).

$$\max_{a} W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j k(\boldsymbol{x}_i, \boldsymbol{x}_j) \alpha_i \alpha_j,$$
$$0 \leq \alpha_i \leq C, \forall i,$$

$$\sum_{i=1}^{l} = y_i\alpha_i = 0. \tag{1}$$

The SMO solves the QR if the $Q_{ij} = y_iy_jk(x_i; x_i)$ positive semi-definite and Karush-Kuhn-Tucker (KKT) conditions are fulfilled for any i (Platt 1998):

$$\alpha_i = 0 \Rightarrow y_if(x_i) \geq 1,$$
$$0 < \alpha_i < X \Rightarrow y_if(x_i) = 1,$$
$$\alpha_i = C \Rightarrow y_if(x_i) \leq 1. \tag{2}$$

Logistic regression algorithm in Weka is a class for building a multinomial logistic regression model combined with ridge estimators (le Cessie and van Houwelingen 1992). Logistic regression calculates the probability that the input falls into a given class, so the probability is always between 0 and 1. The equation for a $p = P(y = 1)$ is (Baesensl et al. 2003):

$$p = \frac{e^{b_0+b_1x_1+b_2x_2+\cdots}}{1 - e^{b_0+b_1x_1+b_2x_2+\cdots}}, \tag{3}$$

where $b_0 + b_1x_1 + b_2x_2 + \cdots$ is a boundary function (logistic regression creates the discriminator as well as the SMO), x_n are the independent variables and b's are regression coefficients (b_0 is an intercept and b_n are the parameter vectors). In order to improve the parameter estimates and to reduce the error, Weka uses ridge estimators (le Cessie and van Houwelingen 1992).

Lazy.IBK is the K-nearest neighbors' classifier. It is an instant-based learning classifier, which is a type of lazy learning algorithm that is fairly simple (Aha, Kibler, and Albert 1991). It makes a construction hypothesis from the training instances. The output is a class membership. KNN considers only k-most similar data instances from the training set in order to classify an instance. Class label is classified by the majority of the k-nearest neighbors. Lazy.IBK uses Euclidean distance as the similarity measure (Baesens et al. 2003) :

$$d(x_1, x_j) = \|x_1 - x_j\| = [(x_1 - x_j)^T(x_1 - x_j)]^{1/2}, \tag{4}$$

$x_1, x_j \in R^n$ as the input vectors of data instance i and j.

Decision trees C4.5 (J48) is an algorithm which is very popular given that decision trees are very easy to understand. Like it's predecessor ID3, it uses the concept of information entropy (Baesens et al. 2003):

$$Entropy(S) = -p_1\log_2(p_1) - p_0\log_2(p_0), \tag{5}$$

Where $p_1)p_0$ is the proportion of examples in class 1 (0) in sample S. Expected reduction of entropy is (Baesens et al. 2003):

$$Gain(S, x_i) = Entropy(S) - \sum_{v \in values(x_i)} \frac{|S_v|}{|S|} Entropy(S_v), \qquad (6)$$

where S_v is a subsample of S and attribute x_i has one specific value.

So, unlike ID3, it works with numerical attributes as well as with nominal ones. That is why it has built-in mechanisms that suggest 3 types of tests: Standard test on a discrete attribute (same as ID3); binary test where the P threshold and the values will be defined and more complex test based on a discrete attribute where the values will be gathered in a larger number of groups with one branch for each group. Another difference between ID3 and C4.5 is that ID3 uses 'information gain' to compare potential data and C4.5 uses 'Gain ratio' (Hssina et al. 2014) which is a proportion of information generated by the split that is useful in classification as a normalization (Baesens et al. 2003):

$$Gainratio(S, x_i) = \frac{Gain(S, x_i)}{SplitInformation(S, x_i)},$$

$$SplitInformation(S, x_i) = \sum_{k \in values(x_i)} \frac{|S_k|}{|S|} \log_2 \frac{|S_k|}{|S|}, \qquad (7)$$

where S_k is a subsample of S and attribute x_i has one specific value and *SplitInformation* is the entropy of S regarding the values of x_i. C4.5 favors splits with the biggest gain ratio but the information gain must be as large as the average gain over all splits. As a result, the tree can be very complex so C4.5 follows postpruning method.

functions.RBFNetwork is a normalized Gaussian radial basis function network which is actually a neural network (Scholpkopf et al. 1996):

$$g(\boldsymbol{x}) = \sum_{i=1}^{K} w_i G_i(\boldsymbol{x}) + b = \sum_{i=1}^{K} w_i \frac{1}{(2\pi)^{d/2} \sigma_i^d} \exp(-\frac{\|\boldsymbol{x} - \boldsymbol{c}_i\|^2}{2\sigma_i^d} + b, \quad (8)$$

where G_i is the ith Gaussian basis function with center $\boldsymbol{c}_i$ and variance σ_i^d. Variant coefficients are w_i and b is a bias term (Scholkopf et al. 1995).

rules.JRip is a rule builder, based on k-iterations of the optimization (Cohen 1995). In a building stage, this algorithm repeats growing and pruning phase until the discretion length of the rule set and examples is 64 bits greater than the smallest discretion length, or the error rate is $\geq 50\%$ or there are no positive examples. In the growing phase, it grows one rule by adding conditions to the rule until the rule is 100% accurate.

meta.Bagging is a bagging predictor (bootstrap aggregating algorithm) that generates multiple versions of a predictor in order to get an aggregated predictor. If the outcome is numerical (estimation problem) – it averages over the versions of the predictor. For classification problems – it does a plurality votes. It makes bootstrap replicates of the learning set and to use them as new learning sets (Breiman 1996). The voting procedure can be explained as follows (Machová, Barčák, and Bednár 2006):

$$H(d_i) = sign\left(\sum_{m=1}^{M} \alpha_m H_m(d_i)\right),\tag{9}$$

d_i is classified to the class that has a majority of classifiers vote. $\alpha_m, m = 1, \ldots, M$ are set so that more precise classifiers have stronger influence on the final prediction. H_m are weak classifiers as their precision as a base classifiers can only be a little bit higher than the precision of random classification (Machová, Barčák, and Bednár 2006). Bagging was formulated by Leo Breiman in 1994 for improving classification accuracy.

trees.RandomForest is a class for constructing a forest of random trees. It was introduced by Leo Breiman in 2001. As the name says, randomness is applied both at row and at column level in order to generate many trees. As in bagging, for the classification problems – voting is used and for regression – averaging is used to declare the final result. The margin function can be formulated as (Breiman 2001):

$$mg(X, Y) = av_k I(h_k(X) = Y) - \max_{j \neq Y} av_k I(h_k(X) = j),\tag{10}$$

where $h_1(x), h_2(x), \ldots, h_k(x)$ are a group of classifiers that together with a training set (which is drawn from the distribution of random vector Y, X) define the function, and where $I(.)$ is the indicator function. So the margin calculates how much the average number of votes at X, Y for the given class exceeds the average vote for any other class. The generalization error is as follows (Breiman 2001):

$$\text{PE}^\star = P_{X,Y}(mg(X, Y) < 0).\tag{11}$$

Random forests produce a limiting value of the generalization error which depends on the strength of the individual trees and the correlation between them (Breiman 2001).

meta.AdaBoostM1 is a class for boosting a nominal class classifier. It uses Adaboost M1 method. Just as bagging, boosting is repeatedly running a given weak learning algorithm and combining the classifiers, but it constructs a distribution in a different way. There are two major ef-

TABLE 5   Confusion Matrix

Data class	Classified as positive	Classified as negative
Positive	True positive (TP)	False negative (FN)
Negative	False positive (FP)	True negative (TN)

fects of boosting. First one is that it generates a hypothesis whose error on the training set is small by combining many hypotheses whose error may be large. The second one is a variance reduction (taking a weighted majority over many hypotheses, which were trained on different samples from the same training set – it reduces the random variability of the combined hypothesis). Adaboost M1 calls the WeakLearn learning algorithm repeatedly in a serious of rounds. The goal is to find a hypothesis which minimizes the training error (Freund and Schapire 1996):

$$e_t = Pr_{iD_t}[h_t(x_i) \neq y_i], \tag{12}$$

where x_i is an instance, y_i is the class label associated with x_i, D_t is the distribution provided by weak learner in a round t and h_t is the hypothesis. The combined hypothesis is a weighted linear threshold of the weak hypotheses. AdaBoostM1 gives more weight to examples which seem to be the hardest. So the upper bound on error of the final hypothesis is as follows (Freund and Schapire 1996):

$$\frac{1}{m}|\{i : h_{fin}(x_i) \neq y_i\}| \leq \prod_{t=1}^{T} \sqrt{1 - 4\gamma_t^2} \leq \exp(-2\sum_{t=1}^{T} \gamma_t^2). \tag{13}$$

The major disadvantage of Adaboost M1 is that it cannot handle a weak hypothesis with error greater than 1/2 (Freund and Schapire 1996).

ACCURACY MEASURES

The following measures were used for the comparison of algorithms for credit scoring in combination with various variable selection approaches: % correct, Kappa, True positive rate, True negative rate, False positive rate, False negative rate, IP Precision, IP Recall, F-measure, and ROC area.

The correctness of a classification algorithm can be measured by computing the number of correctly classified class instances: TP (true positive) and TN (true negative) and a number of instances that were incorrectly assigned to the class: FP (false positive) and FN (false negative). Together they form the confusion matrix for binary classification (Sokolova and Lapalme 2009) (table 5):

Percentage of correctly classified instances is defined as a sum of TP

and TN divided by a total number of instances. It is also called accuracy.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}. \tag{14}$$

Precision is defined as TP divided by the number of examples classified as positive (TP + FP). It is a class agreement of the data labels with the positive labels classified.

$$Precision = \frac{TP}{TP + FP}. \tag{15}$$

True positive (TP) rate is the number of correctly classified positive divided by a positive label (a sum of TP and FN). It is also called a sensitivity measure or a recall. It shows the effectiveness of an algorithm to predict positive cases. The higher the TP rate, the better the performance of an algorithm.

$$TP \text{ rate} = \frac{TP}{TP + FN}. \tag{16}$$

In a case of the credit risk data it is a measure that shows how well the algorithms are classifying instances into 'good' clients (variable 'Default'). Both recall and precision do not take into account the number of TN.

True negative (TN) rate is the number of correctly classified negative instances divided by negative labels (sum of TN and FP). It shows how well a classifier identifies negative labels. It is also called a measure of specificity.

$$TN \text{ rate} = \frac{TN}{TN + FN}. \tag{17}$$

False positive (FP) rate is a number of falsely classified positive instances divided by a total number of negative instances.

$$FP \text{ rate} = \frac{FP}{FP + TN}. \tag{18}$$

False negative (FN) rate is a number of falsely predicted negative events divided by positive instances (a total of FN and TP).

$$FM \text{ rate} = \frac{FN}{FM + TP}. \tag{19}$$

F-measure also named balanced F-score is a measurement of precision and recall as it is the approximate average of the two (Zhang, Zhang, and Yang 2003).

$$F = 2\frac{precision \cdot recall}{precision + recall}. \tag{20}$$

It is the same as accuracy but without the effect of TN. The reason to do F-measure is that there can be a very high precision and very low recall

for example. That is why F measurement does the averaging. For ratios it is better to use harmonic mean over the geometric and arithmetic mean. Same as in recall and precision, the higher the value is, the algorithm performs better (Powers 2011). A more general form of F-measure is defined as weighted harmonic mean of recall and precision:

$$F_\beta = \frac{(\beta^2 + 1) \cdot P \cdot R}{\beta^2 P + R}, \tag{21}$$

where β is a relative importance of precision and recall (Zhang 2007).

ROC (Receiver operating characteristic) graph is a fundamental tool for diagnostic test evaluation. ROC shows the FP rate (specificity) on the x axis and TP rate (sensitivity) on the y axis. It shows relations between benefits (TP) and costs (FP). The diagonal line on ROC graph presents a 50:50 chance of guessing TP on account of FP. There are 5 discrete classifiers. D presents a perfect classifier with specificity = 0 and sensitivity = 1 and C (0,7,0,7) represents a random chance of guessing TP 70% of the time. The classifier is better the further northwest it is from the classifier it compares to. Those that have low TP rate and also make a small mistake on FP are called 'conservative.' The ones that are placed in an opposite corner on the left side of ROC are called 'liberal' as they predict more TP but also do a fair share of FP. Another extreme is a point E which represents a negation of a point B and performs worse than guessing (Fawcett 2005). That classifier has the useful information but it does not apply them correctly.

Many classifiers produce a class decision on each instance and in the end produce only one confusion matrix which gives only one point in ROC space. Such are decision trees or rule sets. But rather than having only one point, scoring and voting can be used in order to create a curve. In order to compare classifiers AUC (Area under an ROC curve) can be used. It shows 'the probability that the classifier will rank a randomly chosen positive instance higher than a negative one' (Fawcett 2005).

Cohen's kappa coefficient (κ) is a classification measurement for categorical variables which measures inter-rate agreement. In other words, it is a measure of how well the algorithm performed to how well it would have performed by chance. If κ is high – there is a big difference between the accuracy and the null error rate.

$$\kappa \equiv \frac{p_o - p_e}{1 - p_e}, \text{ where } p_o = p_{11} + p_{12} \text{ and } p_e = f_1 g_1 + f_2 g_2, \tag{22}$$

where p_o is the proportion of observed agreement among the raters and

TABLE 6   Variables Selected by Different Algorithms

Category	Variable	(1)	(2)	(3)
Entrepreneurial idea	Vision	✓	✓	✓
	Better	×	✓	✓
	Market	×	✓	✓
Growth plan	Reinvest	✓	✓	✓
Marketing plan	Ad	✓	✓	✓
	Compet	×	✓	✓
Personal characteristics of entrepreneurs	Occup	×	✓	✓
	Age	×	✓	×
	Location	×	×	×
Characteristics of SME	Ind	×	✓	✓
	Start	×	✓	×
	Equip	×	✓	×
	Emp	×	✓	×
Characteristics of credit program	Int	×	✓	✓
	Grace	×	✓	✓
	Prin	×	✓	×
	Period	×	✓	×
	I_rate	×	✓	×
	Amount	×	✓	×
Relationship between the entrepreneur and a financial institution	Client	×	✓	×

NOTES   Column headings are as follows: (1) Class CfsSubsetEval, (2) ChiSquaredVariableEval, (3) ConsistencySubsetEval.

p_e is the hypothetical probability of a chance agreement (Byrt, Bishop, and Carlin 1993). Kappa statistic varies from 0 to 1, with the 1 as perfect agreement, and 0 as the 0 = agreement equivalent to chance.

Results

Table 6 presents the variables used for a different approach to the variable selection. The Class CfsSubsetEval algorithm selected only three variables: Variable Vision – does the entrepreneur have a clear vision of the business? Variable Ad – promotion of products/services. Variable Reinvest – projected percentage of the invested profit.

The ChiSquaredVariableEval algorithm selected all of the variables ex-

cept the Variable Location. The ConsistencySubsetEval selected all of the variables related to the Entrepreneurial idea, Growth plan, and Marketing plan. However, only a few algorithms were selected that were related to the Personal characteristics of entrepreneurs, Characteristics of SME, Characteristics of a credit program, and Relationship between the entrepreneur and a financial institution.

Research results showed that variables selected by the algorithm Class CfsSubsetEval have the best results regarding the percentage of correctly classified instances. On the other hand, according to the percentage of bad debtors falsely predicted as the good ones, the decision tree generated using the variables selected by the ChiSquaredVariableEval is the worse. According to the criteria of the minimal percentage of falsely predicted bad debtors as good, the best approach was to use the decision tree generated using the variables selected by the Class CfsSubsetEval or the decision tree generated using the variables selected by the ConsistencySubsetEval. In addition, for financial institutions, especially for banks, the most valuable data are the data on prediction of bad debtors, and in our case two mentioned algorithms should be used. However, since the Class CfsSubsetEval generates a decision tree that is based only on the Variable Vision, it is prone to subjective mistakes, since this variable was estimated by a banking clerk. The ConsistencySubsetEval could be considered as more reliable, since it produces similar results as the Class CfsSubsetEval, and it is based on a larger number of variables. Most of the variables are related to 'what has been done' instead of 'who is doing it.' In other words, variables related to Entrepreneurial idea, Growth plan, and Marketing plan were more relevant than variables related to Personal characteristics of entrepreneurs, Characteristics of SME, Characteristics of credit program, and Relationship between the entrepreneur and a financial institution.

Table 7 presents the impact of different variable selections to performance of classification algorithms using selected accuracy measures. According to the IP Recall, the most superior algorithm is C4.5 in combination of Class CfcSubsetEval, with IP Recall of 0,9829. However, according to other accuracy measures, the most superior algorithm is SVO, in combination with the Class CfcSubsetEval and ConsistencySubsetEval.

Table 8 presents the performance of various algorithms with the usage of various approaches for the variable selection, namely true positive and true negative rate, as well as false positive and false negative rate. Again, according to these accuracy measures, the most superior algorithm is

TABLE 7   Performance of Various Algorithms with the usage of Various Approaches for the Variable Selection, According to Accuracy Measures

Measure	Dataset	(1)	(2)	(3)	(4)	(5)	(6)
AdaBoost	Original dataset	68.5763	0.1778	0.7046	0.9120	0.7924	0.6016
	Class CfcSubsetEval	70.7526	0.2148	0.7100	0.9494	0.8107	0.6768
	ChiSquaredAttributeEval	70.4605	0.2182	0.7139	0.9326	0.8068	0.6512
	ConsistencySubsetEval	68.8842	0.1844	0.7060	0.9158	0.7948	0.6049
Bagging	Original dataset	64.8237	0.1079	0.6904	0.8576	0.7622	0.6406
	Class CfcSubsetEval	68.4632	0.1816	0.7064	0.9038	0.7906	0.7125
	ChiSquaredAttributeEval	66.2947	0.1421	0.6958	0.8761	0.7727	0.6575
	ConsistencySubsetEval	64.8289	0.1121	0.6918	0.8537	0.7613	0.6384
Bayesian Network	Original dataset	67.5658	0.2641	0.7539	0.7682	0.7556	0.6977
	Class CfcSubsetEval	72.9289	0.2764	0.7243	0.9612	0.8249	0.7474
	ChiSquaredAttributeEval	69.7526	0.3011	0.7580	0.8054	0.7762	0.7142
	ConsistencySubsetEval	68.2974	0.2794	0.7561	0.7775	0.7617	0.7024
c4.5	Original dataset	67.6079	0.1588	0.7010	0.8995	0.7839	0.6039
	Class CfcSubsetEval	71.1421	0.1960	0.7029	0.9829	0.8188	0.6183
	ChiSquaredAttributeEval	69.0921	0.1765	0.7015	0.9335	0.7985	0.6136
	ConsistencySubsetEval	67.3421	0.1521	0.6990	0.8979	0.7820	0.5990
K-nearest neighbor	Original dataset	64.3211	0.1466	0.7041	0.7986	0.7454	0.5994
	Class CfcSubsetEval	70.1974	0.2630	0.7344	0.8699	0.7935	0.6863
	ChiSquaredAttributeEval	65.9026	0.1545	0.7035	0.8432	0.7641	0.6394
	ConsistencySubsetEval	62.7711	0.0924	0.6863	0.8071	0.7386	0.6074
Linear logistic	Original dataset	69.6263	0.3189	0.7795	0.7705	0.7674	0.7281
	Class CfcSubsetEval	71.8921	0.2924	0.7379	0.9023	0.8089	0.7460
	ChiSquaredAttributeEval	72.2211	0.3672	0.7855	0.8092	0.7922	0.7212
	ConsistencySubsetEval	71.7263	0.3621	0.7916	0.7912	0.7848	0.7416
Random Forest	Original dataset	66.1263	0.1399	0.6974	0.8683	0.7712	0.6833
	Class CfcSubsetEval	70.6184	0.2786	0.7405	0.8660	0.7953	0.7077
	ChiSquaredAttributeEval	66.8105	0.1872	0.7099	0.8453	0.7688	0.6832
	ConsistencySubsetEval	66.6342	0.1509	0.7000	0.8746	0.7754	0.6710
rbf Network	Original dataset	66.7658	0.2297	0.7358	0.7860	0.7550	0.6693
	Class CfcSubsetEval	72.6474	0.2995	0.7372	0.9209	0.8166	0.7179
	ChiSquaredAttributeEval	69.9711	0.2952	0.7532	0.8226	0.7818	0.6963
	ConsistencySubsetEval	67.5368	0.2513	0.7418	0.7878	0.7600	0.6723
Ripper rule induction	Original dataset	65.8237	0.1393	0.6969	0.8582	0.7673	0.5673
	Class CfcSubsetEval	68.9105	0.2058	0.7140	0.8939	0.7904	0.6004
	ChiSquaredAttributeEval	68.6105	0.2129	0.7168	0.8751	0.7852	0.6017
	ConsistencySubsetEval	68.6368	0.2090	0.7152	0.8807	0.7864	0.5976
svo	Original dataset	74.4132	0.4168	0.8043	0.8209	0.8088	0.7081
	Class CfcSubsetEval	73.1289	0.2824	0.7259	0.9612	0.8259	0.6217
	ChiSquaredAttributeEval	73.7237	0.3815	0.7789	0.8521	0.8102	0.6835
	ConsistencySubsetEval	75.7921	0.4445	0.8095	0.8403	0.8204	0.7195

NOTES   Column headings are as follows: (1) % correct, (2) kappa, (3) ip precission, (4) ip recall, (5) F-measure, (6) roc area.

TABLE 8   Performance of Various Algorithms with the Usage of Various Approaches
for the Variable Selection, According to Forecasting Rate

Measure	Dataset	(1)	(2)	(3)	(4)
AdaBoost	Original dataset	0.9120	0.2448	0.7552	0.0880
	Class CfcSubsetEval	0.9494	0.2345	0.7655	0.0506
	ChiSquaredAttributeEval	0.9326	0.2590	0.7410	0.0674
	ConsistencySubsetEval	0.9158	0.2464	0.7536	0.0842
Bagging	Original dataset	0.8576	0.2407	0.7593	0.1424
	Class CfcSubsetEval	0.9038	0.2571	0.7429	0.0962
	ChiSquaredAttributeEval	0.8761	0.2488	0.7512	0.1239
	ConsistencySubsetEval	0.8537	0.2488	0.7512	0.1463
Bayesian Network	Original dataset	0.7682	0.4957	0.5043	0.2318
	Class CfcSubsetEval	0.9612	0.2767	0.7233	0.0388
	ChiSquaredAttributeEval	0.8054	0.4871	0.5129	0.1946
	ConsistencySubsetEval	0.7775	0.4993	0.5007	0.2225
C4.5	Original dataset	0.8995	0.2407	0.7593	0.1005
	Class CfcSubsetEval	0.9829	0.1805	0.8195	0.0171
	ChiSquaredAttributeEval	0.9335	0.2176	0.7824	0.0665
	ConsistencySubsetEval	0.8979	0.2357	0.7643	0.1021
K-nearest neighbor	Original dataset	0.7986	0.3393	0.6607	0.2014
	Class CfcSubsetEval	0.8699	0.3738	0.6262	0.1301
	ChiSquaredAttributeEval	0.8432	0.2981	0.7019	0.1568
	ConsistencySubsetEval	0.8071	0.2774	0.7226	0.1929
Linear logistic	Original dataset	0.7705	0.5521	0.4479	0.2295
	Class CfcSubsetEval	0.9023	0.3631	0.6369	0.0977
	ChiSquaredAttributeEval	0.8092	0.5543	0.4457	0.1908
	ConsistencySubsetEval	0.7912	0.5726	0.4274	0.2088
Random Forest	Original dataset	0.8683	0.2581	0.7419	0.1317
	Class CfcSubsetEval	0.8660	0.3945	0.6055	0.1340
	ChiSquaredAttributeEval	0.8453	0.3238	0.6762	0.1547
	ConsistencySubsetEval	0.8746	0.2610	0.7390	0.1254
RBF Network	Original dataset	0.7860	0.4383	0.5617	0.2140
	Class CfcSubsetEval	0.9209	0.3476	0.6524	0.0791
	ChiSquaredAttributeEval	0.8226	0.4617	0.5383	0.1774
	ConsistencySubsetEval	0.7878	0.4567	0.5433	0.2122
Ripper rule induction	Original dataset	0.8582	0.2669	0.7331	0.1418
	Class CfcSubsetEval	0.8939	0.2905	0.7095	0.1061
	ChiSquaredAttributeEval	0.8751	0.3174	0.6826	0.1249
	ConsistencySubsetEval	0.8807	0.3074	0.6926	0.1193
SVO	Original dataset	0.8209	0.5952	0.4048	0.1791
	Class CfcSubsetEval	0.9612	0.2821	0.7179	0.0388
	ChiSquaredAttributeEval	0.8521	0.5150	0.4850	0.1479
	ConsistencySubsetEval	0.8403	0.5988	0.4012	0.1597

NOTES   Column headings are as follows: (1) true positive rate, (2) true negative rate, (3) false positive rate, (4) false negative rate.

svo, in combination with the Class CfcSubsetEval and ConsistencySubsetEval.

Conclusion

The goal of the research was to investigate which data mining algorithm would reveal the best accuracy results for the purpose of developing credit scoring for entrepreneurs financing in one Croatian banking institution. The novelty of this work is based on the data as well as algorithm selection process. We use the data from one Croatian banking institution, which provided the dataset of entrepreneurial credit, with non-standard data, such as behavioral and demographic data, as well as data which are specific for entrepreneurial projects as well as entrepreneurial attitudes. On this dataset, we tested the accuracy of various data mining algorithms, combined with various algorithms for variable selection. Based on most of the accuracy measures, the most accurate algorithm was svo in combination with the Class CfcSubsetEval and ConsistencySubsetEval algorithms for variable selection.

References

Aha, D., D. Kibler, and M. Albert. 1991. 'Instance-Based Learning Algorithms.' *Machine Learning* 6 (1): 37–66.

Baesensl, B., T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen. 2003. 'Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring.' *Journal of the Operational Research Society* 54 (6): 627–35.

Bouckaert, R. R., E. Frank, M. Hall, R. Kirkby, P. Reutemann, A. Seewald, A., and D. Scuse. 2013. *weka Manual for Version 3-7-8.* Hamilton: University of Waikato.

Breiman, L. 1996. 'Bagging Predictors.' *Machine Learning* 24 (2): 123–40.

———. 2001. 'Random Forests.' *Machine Learning* 45 (1): 5–32.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees.* Belmont, ca: Wadsworth.

Byrt, T., J. Bishop, and J. B. Carlin. 1993. 'Bias, Prevalence and Kappa.' *Journal of Clinical Epidemiology* 46 (5): 423–29.

Choudhary, A. K., J. A. Harding, and M. K. Tiwari. 2008. 'Data Mining in Manufacturing: A Review Based on the Kind of Knowledge.' *Journal of Intelligent Manufacturing* 20 (5): 501–21.

Cohen, W. W. 1995. 'Fast Effective Rule Induction.' In *icml '95: Proceedings of the Twelfth International Conference on Machine Learning,* 115–23. San Mateo, ca: Morgan Kaufmann.

Crook, J. N., D. B. Edelman, and L. C. Thomas. 2007. 'Recent Developments in Consumer Credit Risk Assessment.' *European Journal of Operational Research* 183 (3): 1447–65.

Fawcett, T. 2006. 'An Introduction to ROC Analysis.' *Pattern Recognition Letters* 27 (8): 861–74.

Freund, Y., and R. E. Schapire. 1996. 'Experiments with a New Boosting Algorithm.' In *Machine Learning: Proceedings of the Thirteenth International Conference on Machine Learning,* 148–56. Bari: The International Machine Learning Society.

Hall, M. A. 1998. *Correlation-Based Feature Subset Selection for Machine Learning.* Hamilton: The University of Waikato.

Hssina, B., A. Merbouha, H. Ezzikouri, and M. Erritali. 2014. 'A Comparative Study of Decision Tree ID3 and C4.5.' *International Journal of Advanced Computer Science and Applications* 10 (Special Issue): 13–19.

John, G. H., and P. Langley. 1995. 'Estimating Continuous Distributions in Bayesian Classifiers.' In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence,* 338–45. San Mateo, CA: Morgan Kaufmann.

le Cessie, S., and J. C. van Houwelingen. 1992. 'Ridge Estimators in Logistic Regression.' *Applied Statistics* 41 (1): 191–201.

Lejeune, M. A. P. M. 2001. 'Measuring the Impact of Data Mining on Churn Management.' *Internet Research: Electronic Networking Applications and Policy* 11 (5): 375–87.

Lucas, A. 2001 'Statistical Challenges in Credit Card Issuing.' *Applied Stochastic Models in Business and Industry* 17 (1): 83–92.

Machová, K., F. Barčák, and P. Bednár. 2006. 'A Bagging Method Using Decision Trees in the Role of Base Classifiers.' *Acta Polytechnica Hungarica* 3 (2): 121–32.

Marinakis, Y., M. Marinaki, M. Doumpos, N. Matsatsinis, and C. Zopounidis. 2008. 'Optimization of Nearest Neighbor Classifiers via Metaheuristic Algorithms for Credit Risk Assessment.' *Journal of Global Optimization* 42 (2): 279–93.

Ngai, E. W. T., L. Xiu, and D. C. K. Chau. 2009. 'Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification.' *Expert Systems with Applications* 36 (10): 2592–602.

Ngai, E. W. T., Y. Hu, Y. H. Wong, Y. Chen, and X. Sun. 2011. 'The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of Literature.' *Decision Support Systems* 50 (3): 559–69.

Oreski, S., D. Oreski, and G., Oreski. 2012. 'Hybrid System with Genetic Algorithm and Artificial Neural Networks and Its Application to Re-

tail Credit Risk Assessment.' *Expert Systems with Applications* 39 (16): 12605–17.

Patel, H. G., and K. Sarvakar. 2014. 'Research Challenges and Comparative Study of Various Classification Technique Using Data Mining.' *International Journal of Latest Technology in Engineering, Management & Applied Science* 3 (9): 170–76.

Platt, J. C. 1998. 'Fast Training of Support Vector Machines Using Sequential Minimal Optimization.' In *Advances in Kernel Methods: Support Vector Learning,* edited by A. J. Smola, P. L. Bartlett, B. Schölkopf, and D. Schuurmans, 185–208. Cambridge, MA: MIT Press.

Pejić Bach, M., J. Zoroja, B. Jaković, and N. Šarlija. 2017. 'Selection of Variables for Credit Risk Data Mining Models: Preliminary research.' In *40th Jubilee International Convention on Information and Communication Technology, Electronics and Microelectronics,* edited by P. Biljanović, 1599–604. Rijeka: Croatian Society for Information and Communication Technology, Electronics and Microelectronics.

Pourzandi, M. M., and K. Babaei. 2010. 'Using Genetic Algorithm in Optimizing Decision Trees for Credit Scoring of Banks Customers.' *Journal of Information Technology Management* 2 (4): 23–38.

Powers, D. 2011. 'Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation.' *Journal of Machine Learning Technologies* 2 (1): 37–63.

Priya, P. I., and D. K. Ghosh. 2013. 'A Survey on Different Clustering Algorithms in Data Mining Technique.' *International Journal of Modern Engineering Research* 3 (1): 267–74.

Quinlan, J. 1992. *c4.5: Programs for Machine Learning.* San Francisco, CA: Morgan Kaufmann.

Scholkopf, B., K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik. 1996. *Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers.* Cambridge, MA: Massachusetts Institute of Technology.

Sokolova, M., and G. Lapalme. 2009. 'A Systematic Analysis of Performance Measures for Classification Tasks.' *Information Processing and Management* 45 (4): 427–37.

Strohmeier, S., and F. Piazza. 2013. 'Domain Driven Data Mining in Human Resources Management: A Review of Current Research.' *Expert Systems with Applications* 40 (7): 2410–20.

Thomas, L. C. 2000. 'A Survey of Credit and Behavioral Scoring.' *International Journal of Forecasting* 16 (2): 149–72.

Thomas, L. C., R. W. Oliver, and D. J. Hand. 2005. 'A Survey of the Issues in Consumer Credit Modelling Research.' *Journal of the Operational Research Society* 56 (9): 1006–15.

Wei, J.-T., M.-C. Lee, H.-K. Chen, and H.-H. Wu. 2013. 'Customer Relationship Management in the Hairdressing Industry: An Application of Data Mining Techniques.' *Expert Systems with Applications* 40 (18): 7513–18.

Witten, I. H., and E. Frank. 2005. *Data Mining: Practical Machine Learning Tools and Techniques.* San Francisco, CA: Morgan Kaufmann.

Wu, R.-S., C. S. Ou, H.-J. Lin, S.-I. Chang, and D. C. Yen. 2012. 'Using Data Mining Technique to Enhance Tax Evasion Detection Performance.' *Expert Systems with Applications* 39 (10): 8769–77.

Wu, X., V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. F. M. Ng, B. Liu, P. S. Yu, Z. H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg. 2008. 'Top 10 Algorithms in Data Mining.' *Knowledge and Information Systems* 14 (1): 1–37.

Yap, B. W., S. H. Ong, and N. H. M. Husain. 2011. 'Using Data Mining to Improve Assessment of Credit Worthiness via Credit Scoring Models.' *Expert Systems with Applications* 38 (10): 13274–83.

Zhang, S., C. Zhang, and Q. Yang. 2003. 'Data Preparation for Data Mining.' *Applied Artificial Intelligence* 17 (5–6): 375–81.

Zhang, H. 2007. 'Comments on Data Mining Static Code Attributes to Learn Defect Predictors.' *IEEE Transactions on Software Engineering* 33 (9): 635–37.

Zhang, D. F., S. Leung, and Z. M. Ye. 2008. 'A Decision Tree Scoring Model Based on Genetic Algorithm and K-Means Algorithm.' In *Proceedings of the 3rd International Conference on Convergence and Hybrid Information Technology,* 1043–47. Busan: IEEE.

Main Drivers of the Evolution of Management Accounting Concept in Russia: Global Ambitions vs. Local Way

Pavel Lebedev
IEDC Bled School of Management, Slovenia
p.lebedev@inbox.ru

This paper examines the main drivers that initially shaped and are currently influencing the understanding and development of management accounting in Russia, focusing on the specific roles of various stakeholders and institutions. In this research, the author adopted a grounded theory perspective. Data collection was based on archival methods with interpretive approach to the analysis. At the initial stage a thematic model of narrative analysis was applied, followed by an intensive multi-stages coding process at a subsequent phase of the study. The findings, including an emerged theoretical proposition, provide exploratory and explanatory views on the development and current state of management accounting in Russia. The obtained results suggest, that while management accounting in Russia is still an emerging concept, the influence of external factors represented by various stakeholders and institutions on the development of management accounting in Russia has been spontaneous, characterized by a lack of leadership, and unsustainable, driven by ignorance, opportunism, and self-interest of main actors.

Introduction

Traditionally, most of the research in the area of management accounting and control was conducted in the US, whereas it is only during the last decades it started to spread across continents with more papers published by authors from Europe to Australia (Bredmar 2011; Shields 1998). As English language is a dominating means of professional communications in financial and accountancy profession, most of the research in the area of management accounting and control has been, and still is, conducted in the Anglo-Saxon environment or, at least, is highly influenced by it.

Historically, it is difficult to find information about management accounting practices in the countries other than the main global economic powers. Limited ability of many researchers to read about management accounting in languages other than English or, just on the contrary, one's native language hinders the understanding of the dynamics of changes that are occurring in a nation's (and global) management accounting practices, including the extent of changes in management accounting practice terminology, techniques, purposes of techniques, and styles of how techniques are used (Shields 1998). Thus, the lack of information regarding the development of management accounting in other countries, beyond the mainstream scope, creates a significant knowledge gap and limits our opportunities to learn and understand the dynamics of the profession.

Context-specific studies in management accounting contribute to close the abovementioned gaps. Research on emerging economies and dynamically changing environments fosters the inclusion of these countries into the global agenda, which is equally important for them and for their partners in the developed world. Such studies could effectively complement the existing body of knowledge by providing a perspective on the development of management accounting occurring beyond well and long-established paradigms of Anglo-Saxon management accounting and the 'controlling,' as it is often being referred to in German-speaking countries.

Since the collapse of the communist system, a market economy has evolved in Russia, which provided a competitive setting for Russian companies and a need for innovation and effective strategies (Chenhall, Kallunki, and Silvola 2011). While the country is a global energy and natural resource player as well as a member of the G20 economic powers, it has not been given as much attention as BRIC nations such as India and China in either academic research or the popular business press. The modest stream of existing research has mainly focused on dominating companies in the economy – natural resource-based companies and industries (Puffer and McCarthy 2011). At a global level, there are only a few attempts to look systematically into dynamics and scope of management accounting. Some studies only have a local outlook, meaning that they are never disseminated beyond national boundaries since most of them do not comply with internationally acceptable standards of academic work and Russian is their only working language.

The transition to a model of a developing economy in Russia after the

collapse of the Soviet Union has led to the need to build and implement new systems and institutions that operate on the basis of market mechanisms in all social and economic areas. These processes experienced many hardships and still are far from their completion. The development and implementation of management accounting – a key element of business infrastructure – had an especially thorny path. Although Russia in general is adopting western-styled business practices, commercial, legal and accounting systems, as well as approaches to management and strategy, and formal management accounting controls, however, while western-based control practices are being integrated into routines of Russian firms, the application of contemporary practices continues to be limited (Chenhall, Kallunki, and Silvola 2011).

In a market-driven economy, management accounting lies in the very heart of business. It constitutes a core part of decision-supporting infrastructure and, by extension, is a vital part of the market mechanism. However, in Russia management accounting is still an emerging concept, and only recently has the debate within the profession become more inclusive of global topics (Lebedev 2014). An assessment of the gaps between the actual situation and the idealistic 'best-practice' vision and investigation of the underlying factors of these gaps leads to an informed understanding of the current and prospective dynamics of management accounting in a certain context. This informs the reasons for this study and justifies the effort.

While the modern nature of management accounting and its evolution are discussed in other sources,[1] this paper looks into the factors which drive the dynamics of the phenomenon and determine its development – with a focus on emerging economies as exemplified by Russia. It attempts to investigate how the main drivers which determine the development of management accounting, namely the key stakeholders and institutions, initially shaped and are currently influencing the evolution of the concept of management accounting in Russia, focusing on the post-Soviet period, when the country started its transition towards market economy.

This is a qualitative study. It follows an interpretive approach with an aim to develop grounded theory. The interpretive research genre in management accounting has truly come of age, and the ongoing discussion about it affirms the broad scope of what currently counts as 'good practice' for such methodology (Lukka and Modell 2010). The interpretive approach accounts for the social nature and the contextual peculiarities of management accounting. Interpretive theories aim to understand mean-

ings and actions and how people construct them. Interpretive theories are helpful for several reasons. First, they help conceptualize the studied phenomenon to understand it in abstract terms. Second, they help articulate theoretical claims pertaining to scope, depth, power, and relevance of a given analysis. Third, they help acknowledge subjectivity in theorizing and hence recognize the role of experience, standpoints, and interactions, including one's own standpoint. Finally, they offer an imaginative theoretical interpretation that makes sense of the studied phenomenon (Charmaz 2014).

Qualitative research and interpretive approaches may produce especially fruitful results and contributions if applied in the context of dynamically changing environments. As directly prescribed by Kasanen, Lukka, and Siitonen (1993), researchers may consider the constructive approach when management accounting looks for its identity as a respectable discipline.

The rest of the paper is organized as follows. The next section addresses the research question and briefly illustrates the methodology. The subsequent section presents the findings, which provide an exploratory perspective on the role of various factors and their influence on the development of management accounting in Russia. The section 'Emerging theory' offers a theoretical proposition of explanatory nature based on the exploratory findings from the study, followed by a 'Conclusion' section containing final remarks on the research.

Research Framework

RESEARCH QUESTION

The research focuses on the post-Soviet period of modern Russia, when the country started its transition towards market economy. It constitutes a founding and integral stage of the bigger research project conducted by the author. The current study shapes and informs the subsequent investigation efforts. Figure 1 presents the overall research design framework.

This study mainly addresses the question: 'How the main drivers which determine the development of management accounting initially shaped and are currently influencing the evolution of the concept of management accounting in Russia?,' forming at the same time a basis and contributing to the research on the questions 'How did the management accounting concept in Russia evolve and develop?' and 'What is the current understanding of management accounting in Russia.'
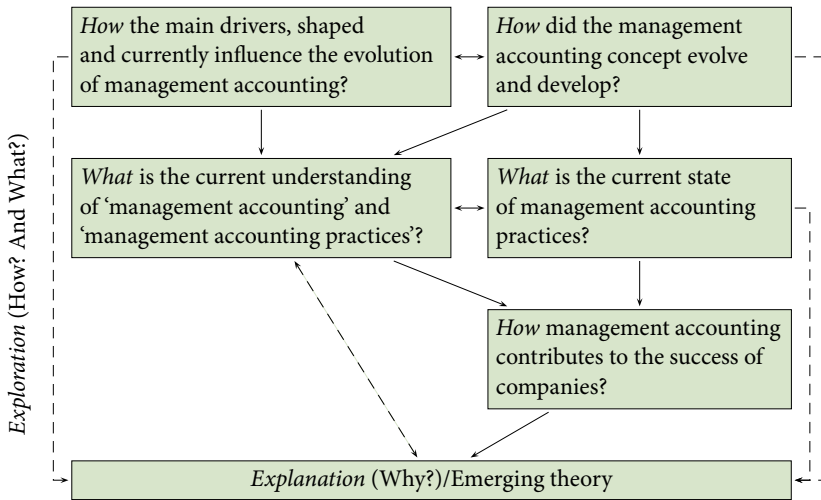
FIGURE 1   Overall Research Design Framework

FIGURE 2   Research Design Summary

METHOD

This is a historical study and is based on archival methods. The author obtained evidence from sources varying from related literature to personal field notes, reflections, observations, and experiences he gained in more than twenty years as a consultant and educator in the field of management accounting. The analysis of this study follows an intuitive style of an interpretive approach. The author applied a thematic model of narrative analysis of evidence (Lebedev 2019e). A summary of the research design is presented in figure 2, which is followed by details covered further in the related subsections.

The section 'Emerging theory' presents an excerpt from the working documentation of the coding process. It contains the summary of the higher level codes and categories which form the grounds for theoreti-

TABLE 1   PEST Analysis for the Development of the Concept of Management
Accounting

PEST groups of factors	Bhimani's factors
Political	Government intervention and regulation.
Economic	Transfer of management accounting ideas and practices across national boundaries by multinational companies.
Social	Academics Education of students and employees Professional associations Consultants
Technological	Technology

NOTES   Base on Bhimani's (1996) model.

cal explanations. The author used Bhimani's (1996) model of the seven
factors that influence management accounting practices in European na-
tions as a framework for structuring the research project. These factors
include the influence of academics, education of students and employees,
government intervention and regulation, the influence of professional as-
sociations, consultants, and technology, and transfer of management ac-
counting ideas and practices across national boundaries by multinational
companies. This presents an appropriate fit with a PEST model of exter-
nal analysis – the most commonly used approach for considering the ex-
ternal business environment across the structure of political, economic,
social, and technological factors (table 1). The internal analysis of the de-
velopment of management accounting in Russia – a perspective from in-
side of companies – is covered in other studies (Lebedev 2018; 2019d).

Bhimani applied the model in his research of management accounting
in 11 European countries and used it 'to provide a historical perspective on
management accounting research and practice in their country, comple-
mented by a discussion of salient raisons d'êtres for the present practice
and thought in the field' (cited by Shields 1998, 503). Table 2 presents a
summary of the factors Bhimani applied and their directions of influence
on management accounting practices.

Findings and Discussion

SUMMARY OF THE FINDINGS

Table 3 summarises the research findings. The Table contains factors (col-
umn 1), the extent to which each factor initially shaped the concept of
management accounting (column 2), and the estimate of current influ-

TABLE 2   Bhimani's Seven Factors that Shape and Influence Management Accounting

Factors	Direction of influence
Academics	Their writings influence debate and practice (e.g., whether full or variable costing should be used, relevant costs for decisions, and, in some countries, 'champion' particular practices).
Education of students and employees	Education develops particular views about management accounting, including the influence some textbooks and teaching cases have on the development and diffusion of management accounting practices.
Government intervention and regulation	For example, to control cost-based pricing and profits to ensure a desired type and level of competition, particularly when there were supply-demand imbalances in a national economy, and, during World War I and World War II, to ration resources and reduce war profiteering.
Professional associations	Advocating particular management accounting practices and best-practices.
Consultants	Advocating particular management accounting practices and/or help firms design and implement practices.
Technology	In particular, the computerisation of accounting systems, which allows more information to be collected, processed, and communicated, in a cheaper and quicker way.
Transfer of management accounting ideas and practices across national boundaries	The transfer occurs by academics, education, consultants, and multinational businesses in a process of their activities.

NOTES   Adapted from Shields 1998.

ence of each factor[2] (column 3). The subsequent sections offer an in-depth discussion of each factor.

THE ROLE OF ACADEMICS: 'THERE IS NO SEX IN USSR'[3]

In this study, the factor 'The role of academics' includes both the domains of research and teaching at academic programs, such as graduate programs. Professional development programs and business education are covered under the factor 'Education.'

The academic community influences the development of a profession through research, publications, debate, and practice. Good practices admit the importance of a right mix of excellence and relevance in research and teaching to ensure the desired outcomes for the stakeholders of the educational process (Abel et al. 2018).

TABLE 3 How Factors Initially Shaped and Are Currently Influencing Management
Accounting in Russia

Factors	Initially shaped	Currently influence
Academics	Insignificant	Moderate
Education	Moderate	Moderate
Government intervention and regulation	Insignificant	Insignificant
Professional associations	Insignificant	Insignificant
Consultants	Significant	Significant
Technology	Moderate	Moderate
Transfer of practices by academics, education, consultants, and multinational businesses across national boundaries.	Insignificant	Moderate

NOTES Adapted from Lebedev 2014.

Traditionally strong in natural sciences, Russian academia dramatically falls behind in social and economic sciences, having been influenced by the ideology of Marxism-Leninism for almost a century. The inherited Soviet practices of teaching and research were far both from relevance to the stakeholders' needs and aspirations (including students, employers, and society) and from excellence. The context of the transition time was characterised by lack of professionalism, lack of experience and understanding of a market economy, lack of financing, low status of education as opposed to the newly emerging rich class, whose representatives promptly created wealth without having any education at all. Hence, no surprise that, challenged with a task to start overnight teaching the concepts of the market economy, the old generation of academia failed – not only individual instructors, but all the system performed weakly.

Thus, during the first decade of transition (the period 1990–2000), the academic community did not play any significant role in the development of management accounting in Russia. Most instructors in educational institutions remained the bearers of the old Soviet approaches and traditions, no educational and methodological infrastructure were available in the form of modern technologies of teaching, relevant textbooks, research findings and developments; no connection to practice existed. The following statement summarises a common attitude towards the challenges the new realities brought, and a common lack of understanding of what

management accounting is and why it is needed (Dean of a leading Russian university, interview by the author, April 24, 2010):

> Why do we need this 'management accounting' at all? Who is going to teach it and how? You know, that there is only one professor at our department who has ever been there [abroad]. And even he doesn't understand what it is about, at least it seems so. And look at all these old ladies in our department. They will never cope! I wish they [the Government] would never push the accounting reform any further.

In spite of the fact that educational institutions had started to gradually include the course 'Management accounting' into the curriculum (in some institutions and universities it was named 'Controlling'), its content often did not have a direct connection with similar subjects in Western educational institutions. Most of the contents of 'Management accounting' courses were devoted solely to production costing; moreover, they did not cover more advanced topics, even in the domain of costing.

The situation slightly changed during the subsequent decades, which has led to some moderate influence of academia. Increasing exposure to the good international practices and efforts of the state to increase the competitiveness of educational system had some positive effects on the content of academic conversations.

In May 2013, in accordance with Presidential Decree of the Russian Federation N. 599 'On measures to realize state policy in the sphere of education and science,' a Russian Academic Excellence Project 5-100 was launched. It was intended to last for eight years and proclaimed such goals as maximizing the competitive position of a group of leading Russian universities in the global research and education market, keeping up with the current pace of international development, internationalization in all spheres, development of infrastructure to recruit the best scientists, faculty, managers, and students, development of an outstanding academic reputation by conducting breakthrough research, and recruiting the world's leading scientists (Prezident Rossiyskoy Federatsii 2012). Although the achievements of the initial goals are controversial and modest, the project has boosted research and publications.

To date, most of the difficulties of the initial period of the transition have been successfully overcome. The substantial content of curricula in management accounting and controlling is in many ways consistent with the programs of Western universities. This allows to characterise the degree of influence of academic's current role as moderate, but a lack of

connection between teaching and practice still often occurs. Therefore, many programs are perceived as purely academic and theoretical, and far from the real world experience.

THE ROLE OF EDUCATION: 'MONEY MAKES THE WORLD GO ROUND' (OR 'WE'LL TEACH YOU WHATEVER YOU PAY US FOR')

Over the entire transition period, the role of educational programs in the development of management accounting can be defined as moderate. Importantly, in this study the factor 'Education' covers professional development programs and business education (i.e., seminars, trainings, and educational courses conducted by new-generation business schools, corporate universities, and private educational providers), whereas the specifics of graduate programs are included in the factor 'The role of academics.'

The moderate influence of such programs is conditioned by the following groups of factors: teachers and instructors' relative professionalism, participants' interest and motivation, and 'advanced' profile of companies sending the participants for training, compared to the overall situation. Instructors teaching in the abovementioned educational institutions were, as a rule, representatives of consulting community, not of an academic environment. This allowed to adapt the content of the programs and made them more applied, and also to use up-to-date information and 'best-practice' experience.

The main customers of such educational programs were usually represented by the companies of the 'new era' (i.e., the companies that were created after the collapse of the Soviet Union and that share the values of corporate training). Accordingly, the participants of such programs were, as a rule, representatives of the most 'market-oriented' part of the economy. Indeed, they were the employees of these advanced companies, which, at an early stage of their functioning, understood the value personnel's training and development can convey for the business. Accordingly, these people were carriers of a certain cultural code which not only admitted, but also supported, in every possible way, various training and development practices.

The main outcome of the influence of education on the development of understanding of management accounting was an attempt to introduce the world-class thinking on management accounting, its approaches, and its tools, at least to a part of the local professional community. This ex-

posure was not an easy exercise. Most of the content was perceived as something that would never work, given the 'specifics' of the national context. The 'specifics of our business' or the 'specifics of our context' are commonly used phrases in the Russian corporate environment, and often hide reluctance to make a try of anything new.

As business education was strongly commercialised, customer satisfaction became one of the concerns of educational providers. However, it was understood rather simplistically, without any reflections on the complex nature of the customer in business education. The most important was the immediate satisfaction of the participants of educational programs, as measured by questionnaires after the end of the course. Numerous attempts were made to provide some adaptation of the taught content to consider any possible 'specifics' and, thus, ensure high scores; however, it looked more like flirting with customers. It led to the increasing offer of various attractively named programs with doubtful quality and inferior content.

Briefly, the most important influence of the factor 'Education' was in that it activated some dynamics to thinking and rethinking of the essence of management accounting. The effect was similar to planting many seeds, some of which could later give birth to a plant of knowledge and understanding, while many others would not, depending on their own quality and quality of the soil.

THE ROLE OF GOVERNMENT INTERVENTION AND REGULATION: 'WHAT MANAGEMENT ACCOUNTING? PLEASE, DO NOT DISTURB.'

The influence of the government and its regulation on the development of management accounting, both in the early years of its formation and later on, were insignificant. The state was not able to see the benefits of management accounting, did not acknowledge its existence, and was not interested in it, having another agenda of the development of accounting in the country.

Traditionally, the role of the state is paramount in Russia. This is true, among other things, in the area of accounting regulation. As one of the most interested financial information users and of the most powerful stakeholders in the relations of financial information flows across the economy, the state fully used its position to push the accounting nationwide towards the direction of tax-accounting and taxation control. This was especially evident during the first decade of the transition period,

when no distinctions between financial and tax accounting existed and the statutory accounting was harmonized with tax accounting and control. Formally, the situation changed with the adoption of a tax code in the period 1998–2000, which meant that what was not included in its scope was intended to be financial accounting (Gosudarstvennaia Duma Rossiiskoy Federatsii 1998). However, while tax accounting was perceived as a burden with which a company was obliged to comply (sometimes just to a certain extent), financial accounting was perceived as an unnecessary burden. At that time, most people in the companies did not have a culture of working with numbers, neither the skills of financial evaluation and analysis. There was a lack of demand for financial information from managers, who, instead, could have used it to make better decisions. They preferred to make them opportunistically and intuitively. Further, there was no demand for management accounting, because 'accounting' was the key burdensome word.

During the subsequent years, one single attempt to incorporate the concept of management accounting into the legislation was made through the inclusion of the subsection of accounts which deemed to serve the needs of management accounting into the official chart of accounts (Ministerstvo Finansov Rossiyskoy Federatsii 2000). This was the only artefact, which, in the opinion of the developers of the document, should have supported the process of costs analysis. It was provided that, 'for management accounting purposes, a company may open accounts in the section 30 for each cost element,' and a handful of accounts (namely 10 accounts only from 30 to 39, with some of them blank) was offered. This confirms a very narrow view of management accounting, and its identification solely with cost and production accounting, even in cases when some understanding of it existed and the overall rejection of the phenomenon. All was on behalf of the state. Last, but not least, it went practically unnoticed by the accounting community.

THE ROLE OF PROFESSIONAL ASSOCIATIONS: SO MUCH TO DO, SO LITTLE TIME

Professional associations are organisations serving the interests of a profession, the individuals engaged in the profession, and the public interest (see https://en.wikipedia.org/wiki/Professional_association). They shape the professions by promoting and advocating best practices and approaches to professional activities, and by providing professional education, development, and certifications. These state-of-the-art solutions are

distilled through research, collaboration, and extended dialogue among stakeholders. An exemplary case is an alliance between the two most prestigious global associations: the American Institute of Certified Professional Accountants (AICPA) and the Chartered Institute of Management Accountants (CIMA), which in 2017 founded the Association of International Certified Professional Accountants, representing more than 650,000 members and students in 179 countries. Global management accounting principles these associations have developed and advocated guide the best practice and inform management accounting profession all over the world (The Chartered Institute of Management Accountants 2017).

Russian professional organisations in the field of accounting are limited in their number and in the number of their members. They are created on a different basis; they were not a good will of their members, but the result of personal goals and a political sophistication of their founders. They have neither sufficient resources, nor will to carry out research to generalise and promote best practices. Relationships within organisations are excessively politicized and built around the establishment of personal influence of organisational leaders. Membership in these organisations is perceived not as an act of goodwill, but as a transaction. Accordingly, members do not have any expectations regarding the support of the association for their professional activities. The transactional nature of the participation in such organisations assumes the deal: the members pay modest annual fees and formally participate in obligatory continuing professional development programs the associations provide (another stream of revenue), and receive in exchange the certifications they might need for compliance regulations. Thus, unfortunately, in Russia professional associations are viewed more as an element of compliance, rather than as a mutually enriching mechanism for communication between professionals who are united by common values. Only in rare instances, community members see themselves as a means of interaction, communication, and exchange of experience.

The case of the Institute of Professional Accountants of Russia (IPAR) is exemplary. The IPAR claims to be the largest professional accounting association in Russia (see https://www.ipbr.org). However, finding an exact number of active members is tricky. On its website, the IPAR states that annually it provides educational programs to 70000 members. The operational and financial models of the organisation lack transparency. Dozens of organisations are in some way affiliated to the IPAR, as the

Trade Register of Russia evidences, and most of them have a registered seat in various regions of Russia. As to the IPAR itself, as of June 16, 2019, it had six founders: three educational institutions, a state-owned publishing company, and a research institute under the Ministry of Finance. This research institute ceased to exist as a separate entity in 2006, and was merged with another educational institution and another professional association which had been inactive for years and was finally liquidated on June 7, 2019[4] (Federal'naia nalogovaia sluzhba Rossii 2019). Obviously, these entities are not about the professional community; rather, they represent a strong and specific business case of legitimising the interests of a narrow circle of players who are able to exploit political and social capital to take an opportunity to talk on behalf of the profession.

In summary, during all the three decades this study considers, professional associations in Russia did not have any significant impact on the development of management accounting.

Importantly, the situation the author described is exclusively related to national professional associations; the role and influence of international professional associations acting in Russia, such as the Association of Chartered Certified Accountants (ACCA) or the Chartered Institute of Management Accountants (CIMA), as well as a number of others, are different and worth considering in a separate study. In general, given their prestigious reputation and high standards, they were serving a very narrow segment of companies and members, who conform or aim to conform to world-class practices.

THE ROLE OF CONSULTANTS: GO WEST (BUT NOT TOO FAR)

Consultants influence the development of management accounting by advising, implementing, and advocating certain management practices, promoting tools and concepts, and expressing respected expert opinion, which altogether are often regarded as a frame of reference. Their authority and credibility to do this are determined by accumulated skills and competences, and long-term toiling for the sake of clients and developing relationships, which assumes a huge investment of money, time, and effort.

In Russia, the development of consultancy as a profession, including consultancy in management accounting, followed a specific path, which resulted in a rather perverted form of outcomes, especially in the early years of the transitional economy, in the period 1990–2000. Indeed, at that time, due to the lack of balance of opinions, consultants were of-

ten perceived as the only possible truth (not to mention the often doubtful quality of the advice given), but also as an evil, sometimes necessary, but mostly not. The underlying values of consultants' professional activity were far from the foundations of sustainability and serving the public well. They had a sharp commercial pragmatic focus and opportunist approach to conducting projects, with no fear of unrealistic promises to their clients.

In the first decade of the economic transition in Russia (1990–2000), there were three main streams of origination of the consultancy profession in the area of management accounting:

- Russian repatriates who had gained some exposure to Western practices (mainly in the United States) as students, postgraduate students, and trainees, for example. Such possibilities allowed to understand the borders were open and Russian citizens were able to travel abroad.
- Local consultants who emerged as a result of a booming demand for accountants and basic accountancy services, which were determined by the liberalisation of economy and the creation of a huge number of private enterprises. These consultants were mainly auditors acting as sole proprietors or they were founders of audit companies which offered consulting services in addition to core activities.
- Western consultants who were available through the numerous technical assistance programs reputable international organisations and institutions (e.g., World Bank, International Finance Corporation, European Bank for Reconstruction and Development, and European Commission's 'Technical Assistance to the Commonwealth of Independent States' programme) ran at that time.

Global international consultancies (e.g., The Boston Consulting Group, McKinsey & Company, PricewaterhouseCoopers, and Deloitte) exclusively advised the government and big international companies. These segments represented a very narrow and closed part of the economic landscape, without any significant influence beyond the 'elite' boundaries.

Table 4 summarises the specifics of the main streams of the origination of consultancy in management accounting and their influence on management accounting in Russia.

Russian repatriates represented the most influential group of consultants, as they offered the most practical solutions compared to the 'nice-

TABLE 4   Main Streams of Origination of Consulting Practices in Management
Accounting in Russia

Category	Russian repatriates	Local consultants	Global consultants
Origin	Local, repatriated after some exposure to global settings.	Local	Global
Source of experience and expertise	Mainly US	Russia	Mainly US/Germany/Other European countries.
Type of services rendered in Russia	Balanced mix of trainings and consulting projects.	Mainly consulting projects.	Mainly trainings and seminars.
Content/Focus	Reporting	Basic tools and concepts, attempts to address specific clients' needs.	Basic tools and concepts.

to-have' storytelling of the global consultants or a handful of tailored so-
lutions of local consultants. The focus on reporting the repatriate consul-
tants' group chose was also successful from the point of view of expansion
of their influence and their personal professional promotion. To a very
great extent, it had determined the thorny 'path' of management account-
ing in Russia, as opposed to selecting a 'highway' of commonly accepted
practices which the rest of the world wisely preferred. This specific un-
derstanding of management accounting consultants advocated assumed
that management accounting was first and foremost responsible for the
preparation of truthful financial statements in the interests of the owners
of a company. Thus, consultants mainly contributed to the formation of
a very specific understanding of management accounting as a superior
accounting system in terms of quality of financial reports, compared to,
for example, financial accounting. This manifested their main influence
on the development of management accounting.

The abovementioned focus on reporting was meant to close the in-
formation gap which had been arising due to the imbalanced system of
reporting that had been historically developed in the country during the
first decade of the transition of economy. As the author noted above, the
official accounting served mainly the interests of the state, primarily the
interests of tax authorities. The lack of a culture of paying taxes, with a
general reluctance to do it, and the absence of a culture and skills of us-
ing of any kind of financial reporting for decision-making contributed to

make companies' owners and managers perceive and approach statutory and tax accounting as an evil or a burden.

According to consultants, managerial accounting was intended to be the solution that would allow preparing 'real' reports, presenting truly what was happening in the company, as opposed to the other perspectives financial and tax accounting provided. Hence, such terms as 'true balance,' 'real balance,' and 'true profit and loss statement' were often used in consulting companies' promotional materials and consultants' publications. Also, even more extreme cases occurred, as it was not unusual to see such advertisements as 'we will provide you with management accounting based on international financial reporting standards' (see https://www.specialist.ru/course/upmsfo; Zolotukhina 2015).

The second most popular theme consultants promoted was automation of management accounting. This is discussed in detail in the following section.

Briefly, the involvement of consultants in the development of management accounting nationwide was characterised by a pragmatic egoistic approach. It assumed a push selling strategy of whatever could be sold to the unsophisticated clients who were constantly looking for a magic pill and did not wish to make any additional effort or suffer any extra pain of organisational transformation or take their own responsibility for organisational change. It was a favourable context which accommodated any type of ploy, fake solutions, and promises never to be fulfilled. The case of a Management Accounting Council under the Ministry of Economic Development of Russia provides a relevant and typical picture.

The Management Accounting Council under the Ministry of Economic Development of Russia was created in 1999 as a successful lobbying initiative of several consulting companies, in fact qualifying for a definition of micro-business. The quasi status 'under the Ministry' does not assume any legitimate status, nor constituted the arrangement any formal part of the Ministry, nor assumed it any power to issue any legislative acts. The ploys of such kind are often-used vehicles in Russia to foster and legitimise initiatives of selected interested companies that try to gain some advantage from being close to the power.

It is and it was then very difficult to find any official information regarding the council. On the web page of the project (which has not been updated nor supported for more than a decade) which was affiliated to one of the founding companies of the council, a then-manager of this company, and a member of the council stated that the body 'consists of

25 people, including representatives of enterprises, scientific and educational institutions, and consulting companies' (Gershun n.d.). The council proclaimed that it would: expand the range of enterprises that used the best management accounting mechanisms in their activities for making management decisions; stimulate the use of the best experience of foreign countries in the development of financial management systems and management accounting in the Russian context; develop basic methodological recommendations on the organisation and maintenance of management accounting in enterprises, taking into account Russian and foreign experience; explain to enterprises' management and owners the importance of management accounting; accelerate the process of forming a professional community; support the process of uniting professionals in the field of management accounting (Ibid.).

The most ambitious task the council tried to accomplish was an attempt to develop methodological recommendations on management accounting, which covered a range of topics such as goals and objectives of management accounting, management accounting tasks, main components of the management accounting system in an organisation, basics of management reporting, and requirements for management reporting organisations. In essence, the document presented a compilation of the basic concepts from textbooks, which were further simplified. It goes without saying that the document went unnoticed. However, what is worth pointing out is the shameless attempt to present it as a piece of legislation. Indeed, the document was prefaced by the following statement:

> Approved by the Management Accounting Council under the Ministry of Economic Development of Russia. Recommended for use by Russian enterprises by the Management Accounting Council under the Ministry of Economic Development of Russia. Minutes of the meeting of April 22, 2002, n. 4. (Ekspertno-konsul'tativnyi sovet po voprosam upravlencheskogo ucheta pri Minekonomrazvitiia Rossii 2002).

Obviously, minutes of a meeting of an odd structure, approving summaries from international textbooks on management accounting, and recommending it for usage by Russian enterprises were not a normative act, and by no means it could be. Rather, it demonstrates the mindset, ambitions, and cynicism which informed many similar arrangements.

TABLE 5 Characteristics of Various Types of IT Solutions for Accounting in Russia Since the 1990s

Category	Mainstream local	Other local	Compact global	Big global
Examples of IT solutions[6]	1C	Galaktika, Best, Info-Bukhgalter	Navision, Axapta, Hansa	Oracle Financials, SAP
Designed for a company	Any size	Small and mid-sized	Mid-sized	Big
Availability of consultants	Plenty	Limited number	Few	Few
Cost of implementation	Low	Moderate	High	High
Chances of successful implementation	High	High/Medium	Low	Low
Selling strategy	Pull	Pull/Push	Push	Push

THE ROLE OF TECHNOLOGY: 'WE ARE THE CHAMPIONS (OR AT LEAST WE TRY OUR BEST)'

Technology and the computerisation of accounting systems allow more information to be collected, processed, and communicated, in a cheaper and quicker way. Advanced solutions as manufacturing resource planning (MRP), enterprise resource planning (ERP), and business intelligence (BI) systems foster the process of analysis and decision-making, thus potentially advancing management accounting systems and creating competitive advantages to a company. Besides, the implementation of technology usually requires rethinking and remodelling of business processes, which, in turn, direct the effort towards the revision of the company's strategic agenda. All these elements require external (i.e., the consultants) and internal (i.e., updated skills and competences of the staff) competences. Thus, it could be reasonably assumed that a successful implementation of technology, or even an attempt to do so, could move the company to the right side along the maturity level continuum of management accounting.[5]

The booming market for computerisation and automation of accounting in Russia had a moderate influence on the development of accounting during all the three decades of the transition of economy, since the 1990s. This influence was determined by the types of solutions which were available on the market (table 5).

1C is the most popular accounting software in Russia. Founded in 1991, the 1C company has specialised in development, distribution, publishing, and support of mass-market software, as of 2019 having 8000 dealers from 600 cities and about 7500 teams constituting the 1C: Franchising partner network (see http://1c.ru/eng/title.htm). The scale of the company, its network of partners, reasonable prices, country-wide availability of support, and easiness of implementation ensured high penetration of this software. The overall influence of the implementation projects on the development of management accounting was an introduction to some very basic concepts of management accounting 1C consultants provided to the companies in order to facilitate such projects. However, in many cases, the consultants themselves had very modest understanding of the subject.

The introduction to the Russian market of IT products such as SAP, Axapta, and other similar world-class systems also had a certain impact on the practice of management accounting. As a rule, the company's suppliers offered comprehensive solutions and had appropriate consulting units. Employees of such units were consultants whose profile is described above and is mainly corresponding to those of Russian repatriates. The implementation solutions such companies offered were expensive and had a very broad scope, including business diagnostics, business processes analysis and reengineering, and HR consulting, for example. Being unable to offer solutions in the context of 'as is,' they advocated a large-scale change, which constituted a huge portion of the overall budget, as well. Needless to say, the results of such projects were questionable, if any, and took years of interaction between companies and consultants. However, they have resulted in some knowledge and understanding among the involved employees. This understanding was limited and, in most cases, far from being practical, but its positive influence was serving mainly as a multiplier of awareness.

THE ROLE OF TRANSFER OF PRACTICES BY ACADEMICS, EDUCATION, CONSULTANTS, AND MULTINATIONAL BUSINESSES ACROSS NATIONAL BOUNDARIES: LISTEN TO THE GURU'S VOICE

International companies have had a certain impact on the development of managerial accounting and its understanding in Russia. When these companies accessed the Russian market, they brought with them management technologies. This fully affected the practices of management

accounting the local companies under influence employed. The transfer of technologies by such international companies was usually carried out as follows. First, foreign employees arrived and implemented their parent company's internal technologies and practices, with minimal differences, trying to take into account the Russian context (including Russian legislation and regulations). These employees occupied top managerial posts, and they also managed key functional areas. Russian specialists were invited to take on roles in the middle management, if they had foreign practical experience or completed foreign educational programs. The role of ordinary staff was reserved to local employees. This created the conditions for the realisation of the multiplier effect, when the concepts which were introduced from abroad were taken as a sample. The unconditional authority of international companies, which was especially observed in the first decade of the development of management accounting in Russia, led to the fact that all the management concepts that went beyond the boundaries of these companies and became known to others were at once legitimised and considered to be the only true ones.

Emerging Theory

This study constitutes to a part of bigger investigation of the development of management accounting in Russia in the post-Soviet era and it focuses on the external drivers which formed and informed the understanding of the concept of management accounting (Lebedev 2019b). Having a qualitative interpretive nature, this study forms a basis for grounded theoretical conceptualisation regarding the development of management accounting in Russia.

Corbin and Strauss (1990) provide a systematic overview of canons and procedures of the grounded theory approach. They include the following:

- data collection and analysis are interrelated processes;
- coding;
- concepts are the basic units of analysis;
- categories must be developed and related;
- sampling in grounded theory proceeds on theoretical grounds;
- analysis makes use of constant comparisons;
- patterns and variations must be accounted for;
- process must be built into the theory;
- writing theoretical memos is an integral part of doing grounded theory;

- hypotheses about relationships among categories are developed and verified as much as possible during the research process;
- a grounded theorist needs not to work alone;
- broader structural conditions must be brought into the analysis, however microscopic in focus is research.

A classic set of analytic moves in qualitative analysis follows the next sequence:

- assigning codes or themes to a set of field notes, interview transcripts, or documents;
- sorting and sifting through these coded materials to identify similar phrases, relationships between variables, patterns, themes, categories, distinct differences between subgroups, and common sequences;
- isolating these patterns and processes, and commonalities and differences, and taking them out to the field in the next wave of data collection;
- noting reflections or other remarks in jottings, journals, and analytic memos;
- gradually elaborating a small set of assertions, propositions, and generalizations that cover the consistencies discerned in the database;
- comparing those generalizations with a formalized body of knowledge in the form of constructs or theories (Miles, Huberman, and Saldaña 2014).

Evidently, in the heart of qualitative analysis lies coding. Codes are labels that assign symbolic meaning to the descriptive inferential information compiled during a study; they are attached to data 'chunks' of varying size and can take the form of a straightforward, descriptive label or a more evocative and complex one (e.g., a metaphor) (Miles, Huberman, and Saldaña 2014, 71).

Over the course of the project, the author applied the following coding techniques. At an initial phase of coding a deductive approach was taken, based on a preliminary developed provisional start list of codes. The list was based on the prior work, knowledge, and experience, on the conceptual framework for the research, list of research questions and the preunderstanding brought into the study from the previous phases of the project. The start list consisted of fifty preliminary codes. Two exploratory methods, holistic and provisional coding, supported this pro-

TABLE 6 Theory Building: An Excerpt from the Working Documentation of the Coding Process

Codes	Categories	Concept
Soviet traditions in academia Poor infrastructure for education Nihilism Scepticism Firm belief in own obsolete knowledge	Ignorance	The influence of external factors on the development of management accounting in Russia
Newly emerging informational needs Gap between needs and outcomes of management accounting Unwillingness of companies to make efforts Conservatism of decision makers Fashion for foreign solutions Expectations of a 'magic pill'	Opportunism	has been spontaneous (lack of leadership) and unsustainable (driven by ignorance, opportunism, and self-interest of main actors).
Lack of professional community Short-term orientation Unsustainable consulting practices Promises never to be fulfilled Commercial focus Cynicism	Self-interest	
Unwillingness to take responsibility Unwillingness of companies to make efforts Non-involvement of non-financial managers Expectations of a 'magic pill' Transactional approach to change management	Lack of leadership	

cess by serving as its foundation, and it was enriched by In Vivo and evaluation coding techniques where appropriate. As noted by Miles, Huberman, and Saldaña (2014), for all approaches to coding, several codes will change and develop as field experience continues. At the same time, other codes flourish and too many segments get the same code, thus creating the problem of bulk. These issues lead to inductive coding when other codes emerge progressively and some of the existing codes are being changed, amended, updated, and rearranged, which was exactly how the project developed.

Table 6 presents an excerpt from the working documentation of a coding process. It demonstrates the logic of arriving to theoretical conclusions. The constructivist grounded theory arose as an alternative to objectivism. Objectivist grounded theorists aim to conceptualize the data without taking an interpretive stance. For example, Glaser (2002) treats data

as something separate from the researcher, which implies that data are untouched by the competent researcher's interpretations. However, constructivists study how and why participants construct meanings and actions in specific situations (Charmaz 2014). Theories try to answer questions by offering accounts for what and how a phenomenon happens and may aim to account for why it happened, and theorizing consists of the actions involved in constructing these accounts (Charmaz 2014). Saldaña (2013) with reference to Tavory and Timmermans (2014) and Gibson and Brown (2009) suggests that a social science theory has four main characteristics, as it is traditionally conceived:

- it predicts and controls action (through an if-then/when-then/since that's why logic);
- it accounts for variation in the empirical observations;
- it explains how and/or why something happens by stating its causes and outcomes;
- it provides insights and guidance for improving social life.

At the finalizing stage of the study the author developed a following theoretical proposition concerning the development of management accounting in Russia:

> The influence of external factors on the development of management accounting in Russia has been spontaneous (lack of leadership) and unsustainable (driven by ignorance, opportunism, and self-interest of main actors).

This theoretical proposition explains the findings regarding the development of management accounting in Russia that the author explored throughout the project, and serves as the basis for further research, including the possibility of reframing and operationalizing it for future theory testing studies.

Conclusion

With a lack of similar contextual studies, going beyond traditional Anglo-Saxon scope of research in management accounting, this work makes a theoretical contribution by complementing existing related conversations in science and professional practice with an attempt to improve an understanding of the development of management accounting in Russia.

The study revealed that management accounting in Russia is still an emerging concept. It has gone a long route from its widely held understanding, which is somewhat equivalent to shadow accountancy in the

90s to management reporting systems or a vague concept of the system of information collection, processing, and distribution for the support of management decisions. The role of main actors/stakeholders in the development of management accounting was in general insignificant. At an early stage of development of management accounting in Russia, it was primarily consultants who influenced and informed the concept, with an increasing role of academics and education, as well as cross-border information exchange during the last decade. The specific understanding of management accounting the first generation of consultants had brought in had significantly determined its 'own' way of development, which only currently is becoming more converged to the global understanding of the concept. The influence of external factors on the development of management accounting in Russia has been spontaneous, characterized by a lack of leadership and unsustainable, driven by ignorance, opportunism, and self-interest of main actors.

Practical implications also include the possibility to employ the results of the research to support decisions in the transformation of the financial function along its maturity continuum (both conducted internally by managers and/or with external support of consultants and advisors). This study could also be used as an educational content for programs in professional and business education.

Future research on the topic would be highly appreciated. It should focus on the immediate voice of business. It would be interesting to see primary evidence of management accounting development and its dynamics to understand the bigger picture and case studies to scrutinize the experiences of companies. Another important track could be advancing and reframing of the formulated theoretical proposition for the theory-testing studies and conducting such studies.

Research results could have been influenced by contextual nature and subjectivity of interpretative analysis, possible incompleteness and subjective selection of archival and field data, and limited resources for analysis and interpretation, given the huge scale of the project. The author considered these potential limitations as much as possible at the planning and design stages of the project, and continuously assessed them along the study.

Notes

1 Two suggested studies (Lebedev 2019a; 2019e) provide an outlook on the evolution of the role of management accounting, from transactions pro-

cessing to financial leadership (2019a), and the evolution of a context-specific understanding of management accounting concept in Russia (2019e).

2 The author selected and justified the scale in a progress of extensive coding procedures he applied at the stage of the data analysis.

3 'There is no sex in USSR' is a popular catchphrase and refers to the comments of Liudmila Ivanova. Ms. Ivanova's comments were aired on a perestroika-era TV Show, US-Soviet Space Bridge. An American asked about sex in the Soviet Union. Ms. Ivanova intended to say "There is no sex in the USSR ... there is love," but the last words were cut off of the broadcast' (An American in Moscow 2010).

4 According to the Trade Register of Russia, there is another association with a similar name, registered by 8 physical persons on June 25, 2009. It has a different registration number; however, it is registered on the same business address.

5 Lebedev's (2019c) study of management accounting maturity level continuum provides a detailed discussion of the concept.

6 Examples of software do not reflect changes in brand names or brand ownerships which have occurred as a consequence of M&A deals in IT industry.

References

An American in Moscow. 2010. 'There Was No Sex in USSR.' https:/usa-moscow.blogspot.com/2010/04/there-was-no-sex-in-ussr.html

Abel, D., D. Purg, A. Braček Lali, and N. Kleynand. 2018. 'Manifesto: Changing the Course of Management Development; Combining Excellence with Relevance.' Report, Bled.

Bhimani, A., ed. 1996. *Management Accounting: European Perspectives.* Oxford: Oxford University Press.

Bredmar, K. 2011. 'Theoretical Foundations of the Concept of Management Control: A References Analysis.' *International Journal of Management* 28 (2): 412–27.

Charmaz, K. 2014. *Constructing Grounded Theory.* 2nd edition. London: Sage.

Chenhall, R. H., J.-P. Kallunki, and H. Silvola. 2011. 'Exploring the Relationships between Strategy, Innovation, and Management Control.' *Journal of Management Accounting Research* 23 (1): 88–128.

Corbin, J. M., and A. Strauss. 1990. 'Grounded Theory Research: Procedures, Canons, and Evaluative Criteria.' *Qualitative Sociology* 13 (1): 3–21.

Ekspertno-konsul'tativnyi sovet po voprosam upravlencheskogo ucheta pri Minekonomrazvitiia Rossii. 2002. 'Metodicheskie rekomendatsii

po organizatsii i vedeniiu upravlencheskogo ucheta.' https://www
.twirpx.com/file/152176/

Federal'naia nalogovaia sluzhba Rossii. 2019. 'Predostavlenie svedenii iz
EGRIUL/EGRIP.' https://egrul.nalog.ru/index.html

Gershun, A. n.d. 'Razvitiie upravlencheskogo ucheta v rossii i rol' Ekspertno-
konsul'tativnogo soveta po upravlencheskomu uchetu pri Minekonom-
razvitiya RF.' http://consulting.ru/209ias2

Gibson, W., and A. Brown. 2009. *Working with Qualitative Data.* London:
Sage.

Glaser, B. G. 2002. 'Constructivist Grounded Theory?' *Forum: Qualitative
Social Research* 3 (3). http://dx.doi.org/10.17169/fqs-3.3.825

Gosudarstvennaia Duma Rossiiskoy Federatsii. 1998. *Nalogovyi kodeks
Rossiyskoi Federatsii: Chast' pervaia.* Moscow: Gosudarstvennaya Duma
Rossiyskoy Federatsii.

Kasanen, E., K. Lukka, and A. Siitonen. 1993. 'The Constructive Approach
in Management Accounting Research.' *Journal of Management Ac-
counting Research* 5:243–64.

Lebedev, P. 2014. 'Evolution of Management Accounting Concept in Rus-
sia: In a Search of Identity.' *Procedia: Social and Behavioral Sciences*
156:580–84.

———. 2018. 'Management Accounting in Russian Mid-Sized Compa-
nies: Results of an Extended Survey-Based Study.' In *Proceedings of
the 17th International Scientific Conference Globalization and Its Socio-
Economic Consequences,* 1196–203. Rajecke Toplice: University of Zilina.

———. 2019a. 'Defining Financial Leadership: Reflections on Current
Conversations about a Changing Role of Management Accounting.'
Paper presented at the 6th International Multidisciplinary Scientific
Conference on Social Sciences & Arts SGEM 2019, Vienna.

———. 2019b. 'Management Accounting in Russian Mid-sized Private
Companies: A Report on a 2014–2019 Research Journey.' Paper pre-
sented at the 6th SWS International Scientific Conference on Social
Sciences 2019, Albena.

———. 2019c. 'Management Accounting Maturity Levels Continuum Mod-
el: A Conceptual Framework.' *European Journal of Economics and Busi-
ness Studies* 5 (1): 24–36.

———. 2019d. 'Management Accounting Practices in Mid-Sized Compa-
nies in Emerging Economies: An Evidence from Russia.' In *Knowledge
– Economy – Society: Challenges for Contemporary Economies – Global,
Regional, Network and Organizational Perspective,* edited by J. Kopeć
and B. Mikuła, 93–103. Cracow: Cracow University of Economics.

———. 2019e. 'Three Decades of Management Accounting in Russia: The
Evolution of Understanding of Management Accounting Concept.' In

Proceedings of the 6th International Scientific Conference Contemporary Issues in Business, Management, and Economics Engineering 2019, 353–364. Vilnius: Vilnius Gediminas Technical University.

Lukka, K., and S. Modell. 2010. 'Validation in Interpretive Management Accounting Research.' *Accounting, Organizations and Society* 35 (4): 462–77.

Miles, M. B., A. M. Huberman, and J. Saldaña. 2014. *Qualitative Data Analysis: A Methods Sourcebook.* 3rd edition. Thousand Oaks: Sage.

Ministerstvo Finansov Rossiyskoy Federatsii. 2000. 'Ob utverzhdenii plana schetov bukhgalterskogo ucheta finansovo-khozyaystvennoi deyatel'nosti organizatsii i instruktsii po yego primeneneniiu.' Ministerstvo Finansov Rossiyskoy Federatsii, Moscow.

Prezident Rossiyskoy Federatsii. 2012. 'Ukaz "O merakh po realizatsii gosudarstvennoy politiki v oblasti obrazovaniya i nauki."' Prezident Rossiyskoy Federatsii, Moscow.

Puffer, S. M., and D. J. McCarthy. 2011. 'Two Decades of Russian Business and Management Research: An Institutional Theory Perspective.' *Academy of Management Perspectives* 25 (2): 21–36.

Shields, M. D. 1998. 'Management Accounting Practices in Europe: A Perspective from the States.' *Management Accounting Research* 9 (4): 501–13.

Tavory, I., and S. Timmermans. 2014. *Abductive Analysis: Theorizing Qualitative Research.* Chicago: University of Chicago Press.

Zolotukhina, T. 2015. 'Postroyenie upravlencheskogo ucheta na osnove otchetnosti po msfo.' https://finotchet.ru/articles/175/

Exploration and Exploitation of Good Practices by Technology Parks: Empirical Findings from Poland

Anna Wójcik-Karpacz
Jan Kochanowski University in Kielce, Poland
anna.wojcik-karpacz@ujk.edu.pl

Jarosław Karpacz
Jan Kochanowski University in Kielce, Poland
jaroslaw.karpacz@ujk.edu.pl

Joanna Rudawska
Jan Kochanowski University in Kielce, Poland
joanna.rudawska@ujk.edu.pl

In the context of current economic climate in which enterprises revise business models as well as seek new opportunities and their significance for economic growth policies in a situation where the government recognises the need to support micro-, small and medium-sized enterprises, analysing paths for learning good practices by the technology parks (TPS) is a valuable scientific venture. The aim of this article is to identify ways of learning good practices by technology parks operating in Poland. Data collection techniques engaged standardised and unstructured interviews conducted with 14 managers of the technology parks operating in Poland. In the research, the ambidexterity approach was adopted as one that fits into the current trend of strategic management based on paradoxes, assuming the simultaneity of exploratory and exploitative learning. Results showed that in those TPS which were analysed, maintaining attractive offers of support for current and potential beneficiaries is based on the systematic adaptation of already-recognised good practices and experimentation with new good practices. Research findings indicated ambidexterity as a condition which is not alternative, but necessary for the development of TPS.

Introduction

Dynamic capabilities and their consequences for organisations have been the subject of scientific inquiry for a long time. Despite the fact that

such research works were generally carried out among enterprises (Bruni and Verona 2009, 101–17; Gebauer, Paiola, and Edvardsson 2012, 321–39; Wójcik-Karpacz 2019, 111–2), this problem equally affects other types of organisations, including technology parks (TPS) in which such research was not conducted at all. Changes in the sources of financing TPS' activities in Poland that have been recorded in recent years, which are under the influence of 'withdrawal' of European Union structural funds from existing financing sources and the emergence of a different type of financing, among others, such as local entities, big and small and medium-sized enterprises (SMES) and micro-sized enterprises, encourages them (TPS) to meet the expectations of high quality services reported by entrepreneurs as well as institutions financing services for SMES.

Based on the findings of previous works which indicate that the growing importance of empowerment of clients and other stakeholders than ever before (Urbanowska-Sojkin 2013, 224–35) makes reconfiguration of operational practices more possible, and that requires looking for new ways of doing things. This also applies to the TPS, because they belong to business environment institutions which through their activities are to increase the portfolio of resources available to enterprises and the possibilities of their instrumentation in a commercially viable way (Wójcik-Karpacz and Mazurkiewicz 2015, 575–90; Wójcik-Karpacz and Rudawska 2016, 248–64). A challenge for the TPS, regardless of their business models, which should follow the development of good practices, is that continuing their operations in each period of time requires them to use dynamic capabilities whose development is highly based on good practices (GPS) (Daniel and Wilson 2003, 282–96). In the modern world of paradoxes, the TPS have to cope with tensions which are evoked by these paradoxes. One of them is the paradox of exploration and exploitation (March 1991, 71–87), if they strive to reconcile adaptive capability with innovation and introducing anticipatory changes. What is more, the eldest Polish technology parks have only 24 years of experience. In comparison to the first USA research parks (1948), they are all still in a development phase, trying to reach competitive advantage also by learning from different sources. Therefore, the aim of this article is to identify ways of learning good practices by the technology parks operating in Poland.

This article extends the previous research and focuses on dynamic units in different sizes. The authors extended results of previous research and found that the technology parks operating in Poland are ambidextrous organisations as well because they use both exploratory and ex-

ploitative learning of good practices to become and maintain the status of modern and dynamic TPS. Research findings allow, therefore, to believe that ambidexterity is the right path for developing good practices by the TPS operating in Poland.

Good Practices: Theoretical Approach

In the literature, the concept of good practices is widely defined. A good practice may be a postulate, guideline or rule which indicates what actions should be taken to perform a specific job more efficiently and effectively. They are defined as: 'methods and tools faster and better than others' (Hoag and Cooper 2006, 151), 'activities aimed at optimal performance of work' (Dani et al. 2006, 1717–28) or 'ways of performing activities in order to improve functioning of organisations' (Codling 1995, 34). There are also examples of referring good practices to the potential of organisation, which results from the accumulated experience of employees. Other authors, in turn, identify good practices with know-how and tacit knowledge (O'Dell and Grayson 2006, 601–22).

Good practices are not only breakthrough and exemplary initiatives, but above all, ways of acting resulting from experience. A basis for considering activity as a good practice is to confirm its effectiveness in solving a specific problem (Kerzner 2009, 373). What is important, operational capabilities are considered strong when a given organisation has achieved the best practices (Pfeffermann 2017, 191–212). This means that the development of TPS' operational capabilities is highly based on the best practices (Daniel and Wilson 2003, 282–96). Technology parks' operational capabilities mean achieving technical efficiency and 'doing things right' in basic business operational, administrative and management functions (Pfeffermann 2017, 191–212). An observable expression regarding the advantages of homogeneous good practices is that one may observe 'similarities' in the activities of technology parks which are successful.

What are *Ambidextrous Organisations:* Literature Findings

In strategic management, the uncertain context of environment requires the development of dynamic capabilities. Having such capabilities enables such organisations as technology parks to better perform their functions and, consequently, to contribute to more effective support for innovative enterprises operating in their areas at present and the future. In order for this to happen, technological parks also need to reconfigure their operational capabilities to strike a balance between the explo-

ration of new good practices and the exploitation of current ones (Kriz, Voola, and Yuksel 2014, 287–99) because both activities are needed for them to survive and develop (March 1991, 71–87). Exploratory and exploitative activities, despite the fact that they seem to be contradictory, do not need to be ruled out, and ambidextrous organisations, thanks to their intelligence, are able to reconcile them (He and Wong 2004, 481–94). In the literature, the organisation's capability to balance exploratory and exploitative activities is called ambidexterity (O'Reilly and Tushman 2004, 185–206). Ambidexterity means the organisation's capability to operate in different (oppositional) directions at the same time (Bratnicka 2015, 67–82; Zakrzewska-Bielawska 2017, 174–90). Researchers considering the problem of ambidexterity were focusing on various approaches, placing it, among others, in the theory of organisational learning which is the key trend for intelligent organisations (Nosella, Cantarello, and Filippini 2013, 450–65). Organisational learning refers to the problem of seeking a balance between two competing aims. On the one hand, it is a search for and acquisition of new good practices, i.e. exploration; on the other hand, exploitation of existing good practices, which is based on processes used to improve incrementally existing operational capabilities and knowledge. Exploration leads to adaptation through developing new good practices, while exploitation aims at using existing good practices. In order to achieve high effectiveness in the long term, the TPS must explore good practices, learn new ways of doing things and, at the same time, use good practices which have already been implemented, i.e. select ways to achieve results considered satisfactory over a longer period of time. So if exploration and exploitation are put in opposition to each other, a conclusion may be drawn that both sides of this contradiction are necessary to achieve effectiveness by the TPS in the long run. Ambidextrous organisations must therefore balance exploratory and exploitative learning. It is proved that organisational learning capability positively influences knowledge management and organisational business performance (Turulja and Bajgorić 2018, 1–18; Okreša Đurić and Maleković 2018, 239–46). But, any kind of organisational learning is self-propelling and may lead to falling into a 'success trap' or 'failure trap' (Im and Rai 2008, 1281–96). The theory of organisational learning offers support for understanding the links between the 'success traps' and the development of dynamic capabilities. Literature indicates that when the exploration of good practices displaces exploitation, there is the 'failure trap' (Levinthal and March 1993, 95–112). However, 'success trap' reflects the organiza-

tion's excessive emphasis on learning through exploitation, or a passive learning process based on experience which focuses on operational routines and replication of operational capabilities, and thus replication of previously used good practices. This makes it difficult to enter a higher level of intentional learning in order to change the status quo and develop new knowledge and operational capabilities. Therefore, if the TPS successfully improve their capabilities and increase their effects resulting from exploitation of good practices, their willingness to change decreases. As a consequence, the possibilities to influence changes in the environment may be suppressed, and the most important good practices related to recent successes become, over time, the most inflexible element of such technology parks (Karpacz 2017, 375). But any success achieved in the past is a predictor of existing capabilities and does not always lead to the 'success trap.' Only those technological parks which experience organisational inertia preventing them from breaking the bonds of existing good practices fall into it; while the environment changes, they remain in a trap of the recent success (Karpacz 2017, 373). Therefore, the technology parks' managers should avoid sinking into the trap of existing good practices resulting from past successes and cautiously develop new ones allowing them to respond to changes in the environment. What is important, 'success traps' may have a greater impact on the results of technology parks than 'failure traps.' The immediate value of information about success is higher than that about failure. Information about success tells employees what is going on; while information about failure excludes one of many possible directions of action (Wang, Senaratne, and Rafiq 2015, 26–44). To explore good practices, the technology parks simultaneously use resources and their ability to create and capture values from new good practices which their top management has identified as the most promising ones. At the same time, adopting new good practices may require the selective withdrawal of old practices, and will even require changing business models of some parks (Dyduch 2017, 253–66).

In reference to the above literature findings, the question arises whether ambidexterity is necessary for the survival and development of all TPS, or only those whose industry profile belongs to the high-tech industry due to relatively much higher expectations of enterprises – TPS' tenants resulting from the latest scientific, technical and technological achievements? Literature findings indicate that while organisations operate in hyper-competitive environments, ambidexterity is perceived as a key success factor; while in a relatively stable environment (of less complexity

and dynamics of change as well as greater predictability), it is not necessarily like that (Kriz, Voola, and Yuksel 2014, 287–99). The current state of research does not give a clear answer to the above question. On the one hand, it indicates that in conditions of a complex and constantly changing environment, and such one occurs in the high-tech industry, the technology parks should shape and develop dynamic capabilities and, at the same time, their ability to balance exploratory and operational activities (Kriz, Voola, and Yuksel 2014, 287–99) which are guided by other strategic aims (He and Wong 2004, 481–94). These findings indicate that in such conditions ambidexterity allows the technology parks to cope better. On the other hand, it suggests that the technology parks which address their offer to enterprises operating under more stable conditions may be less inclined to shape and develop ambidexterity because the environment protects them to some extent. Therefore, for this type of technology parks, ad hoc problem solving may be more cost-effective than creating dynamic capabilities in case they have to react to a non-typical and/or temporary change in the needs of enterprises – TPS' tenants. Continuing this discussion, it is reasonable, however, to pay attention to the fact that the current negligence of investments in developing new good practices, although having high efficiency in the present, has poor prospects for the future. Having all this into consideration, although much has already been explained about ambidexterity in the literature, this does not concern such organisations as technology parks, because so far such research has not been conducted. Filling this cognitive gap out found its expression in the subsequent parts of this article.

Description of Research Sample and Data Collection

RESEARCH TECHNIQUES

Data collection techniques engaged standardised and unstructured interviews conducted with 14 managers of the technology parks operating in Poland. Over 1/3 of the general population was analysed.

Questions were being asked in the same order and were open-ended. During the research, the interview scenario was being used, but the questions were being asked in a free manner and on the basis of the so-called thematic threads. These questions allowed for obtaining a lot of valuable information and gave the opportunity to structure certain pieces of information. During the interviews, the information about how the technology parks explore and exploit good practices was obtained, among oth-

ers. After the transcription of interviews, the collected data was subjected to further analysis. The results of grouping and comparing respondents' profiles are presented in the next part of this article.

DATA

Pilot qualitative empirical research on how to learn good practices by the technology parks were being conducted from September 2018 to March 2019.

General population consisted of 37 technology parks operating in Poland (as of 1 September 2018). Due to the fact that a 'technology park' construct is commonly used for each type of park in the literature, this construct was also used to define the different names of parks functioning in practice (Wójcik-Karpacz 2019, 111–12).

Exploratory and exploitative learning of good practices was considered as an organisational phenomenon. Organisational level of analysis of this phenomenon was an argument for engaging managers of technology parks as respondents, as they have adequate knowledge about the activities of the organisations by which they are employed.

Ambidextrous Organisations in Practice: Research Findings

Ways of technological parks' behaviour, which may be more oriented to exploitation or more to the exploration of good practices, were the reasons for asking a research question: how do the technology parks operating in Poland learn good practices? Recognising these ways of technology parks' behaviour gives insight into their learning methods which, at the same time, are the basis for recognising them as ambidextrous organisations, or not.

Analysis of the content of responses provided by the TPS' managers made it possible to identify the paths of exploratory learning of good practices by the TPS operating in Poland.

WAYS OF EXPLORATORY LEARNING OF GOOD PRACTICES BY TPS

TP1 *Entrepreneurs, technical university.*
   While talking with an entrepreneur, by a university, by a technical university (department of machine construction). Not from other parks or from abroad.

TP2 *Employees who take part in international projects, clients.*
   If we see something good, we try to transfer it, we read a lot, we are

taught a lot by life, employees, clients who force us to do it; we are a commercial park, if something is wrong, we have nothing to live on.

TP3 *Domestic trade fairs, foreign parks, and institutions.*

In general, trade fairs when it comes to business, and traveling around the world and seeing different places when it comes to a public sphere.

TP4 *Foreign parks.*

We draw much from abroad, examples of internationalisation of enterprises.

TP5 *TPS' employees participating in international projects, workshops, study visits, other workshops, foreign parks, foreign incubators, foreign accelerators.*

Abroad, to a large extent, and basing on the workshops we have been to or in projects, in the Interreg projects like FastForward where we have partners from 11 European countries and we visit and meet each other in incubators, accelerators; so there, for sure. Moreover, a lot of our staff graduated the Top 500 programme (4 persons), we are steeped in that. Also, from the staff in our park.

TP6 *National media, network of employees' social relations, letters of intent, mutual cooperation agreements, foreign media, other domestic and foreign parks, business environment organisations, domestic and foreign conferences.*

We have a two-dimensional model of communication in terms of our good practices. The first one comprises signed letters of intent and agreements on mutual cooperation; the second one involves foreign and domestic media, groups of parks cooperating with us, and we participate in all possible conferences in order to exchange experiences and search for new solutions. We are currently implementing a few projects in a public sphere, namely for the Ministry of Justice and the Ministry of Science and Higher Education, and we establish cooperation through consulting companies, but also private knowledge of park's employees in the use of EU funds, acquisition of money for expansion and introduction of new products or services, mainly due to the fact that the park is scientific, it has that part in the name and commercialisation is very difficult and we have to look for sources of financing, which is not easy, but we more and more often succeed.

TP7 *Employees, study visits, membership in international associations.*

We travel and it is always very inspiring, we read, we are present in international associations and it also allows us to keep our finger on the pulse.

TP8 *Domestic and foreign study visits in other TPS, other organisations.*
Mainly during domestic and foreign study visits where we look for inspiration, we look at what proves itself in other technology parks, and we wonder if it can be applied in ours, but also in other organisations which are not parks, but some of their activities can be translated into our case.

TP9 *Employees, other domestic TPS, study visits.*
There are two channels, it can be said that on the one hand there are experiences of other parks; here, we try to cooperate with parks of our region, there are 3 technological parks in the region, which are partners with each other in terms of services they provide. We have very partner relationships, we do not become competitors, even when we apply for funds, even from the same source; all of this does not make us not try to be partners and exchange our experiences. The second channel concerns meetings abroad, that is, some trips, but also an analysis of everything that is happening in the world, through a review of the literature.

TP10 *Employees.*
Based on our own experience, we come up with what could be catchy and then we try to organise it.

TP11 *Other domestic and foreign TPS.*
We have been traveling for a long time, we observe various solutions in Europe and the United States and in Poland, and since the very beginning we have seen this model of the park and known how this model should function.

TP12 *Employees, other domestic and foreign TPS, other organisations, including those belonging to business environment institutions.*
Observing and following other technology parks are obviously one of the basic methods of drawing inspiration to use good practices. First of all, these are practices implemented in organisations with a similar business profile; and secondly, we can get first-hand feedback about the benefits of implementing a given solution. We often observe other organisations which partially implement similar or the same tasks as our TP, but they are not technology parks. For example, due to a similar field of operation, we are eager to use the experi-

ence of the Polish Chamber of Chemical Industry (PCCI) which implements a number of information, education and lobbying tasks for the broadly-understood chemical industry. On the other hand, there are many institutions dealing with investor services, such as Special Investment Areas or the Polish Investment and Trade Agency, which are also observed by us in terms of their standards and effects of their work.

TP13 *Improvement of procedures used in own TPS so far, mainly other domestic TPS of a similar industry profile, other business environment institutions.*

Good practices in the Science and Technology Park (STP) in Opole originate mainly from inspiring by practices in other parks or by improving already implemented procedures. Carrying out evaluation surveys among clients, thanks to which we know what to improve in our practices.

STP in Opole as one of the youngest parks in Poland draws huge inspiration from the activities of other parks and systematically monitors good practices implemented by them. We mainly take examples of parks of a similar business profile (conducting research and development activities and in the field of IT).

We often observe and exchange good practices with other institutions of business environment (IBE) in the Opolskie region, among others, Opole Chamber of Commerce, Opolskie Centre for Economy Development, Opole Center for Local Democracy. We observe them in terms of educational or project tasks.

TP14 *Employees, other domestic and foreign TPS, study visits, project meetings, investment missions, thematic workshops, conferences, thematic publications.*

Mainly from other parks, we go there and observe. We participate in project meetings, study visits, investment missions, we ask in detail what and how works well for them, what has not been met with interest, we take pictures. Identification of good practices and their transfer among different centres usually take places through thematic workshops, conferences, study visits or publications in this field.

Research results indicated the existence of various paths of exploratory learning of good practices by the technology parks. Study visits at other domestic and foreign TPS are a very common way of acquiring knowl-

edge about good practices and then copying them in own organisation. Technology parks also share information about good practices at regularly organised thematic conferences, meetings as part of ongoing projects and, sporadically, during thematic workshops. A frequent method of developing good practices is the experience of technology park's employees who, during their day-to-day operations for enterprises functioning both within the TPS and external ones using the TP's services or infrastructure, discover further opportunities to enrich the offer.

On this basis, it may be assumed that only the best practices are important from the perspective of employees working in various positions in the organisational hierarchy. Employees share the best practices and want to replicate successes of other TPS, which supports the argument that successes are more important than failures.

However, it should be borne in mind that although good practices copied have some similarities, they differ among the technology parks which apply them due to specific factors (such as personnel and managerial staff; modern infrastructure and technology; size of office and laboratory space; equipment in laboratories, etc.), as well as internal processes necessary for their implementation, which lead to various results. Technology parks which regularly improve and reproduce their good practices shall probably have an advantage over those which are slower in this respect. In addition, they can make better use of the situation of relative stability in their environment to prepare themselves for more drastic changes at a later date, and even to create ground-breaking changes in their offer. In practice, this means that the technology parks are not focused on solving problems mainly through the further development of existing good practices. However, the implementation of new good practices should not lead to negligence or unreasonable rejection of existing 'proven' good practices, but to complement them. To some extent, the recognition of this issue is brought by subsequent fragments of managers' responses about the TPS' ways of using good practices.

WAYS OF EXPLOITATIVE LEARNING OF GOOD PRACTICES BY TPS

TP1 *Employees, enterprises – TPS' tenants.*
   Nobody will understand services in the park without going to a laboratory and not seeing how employees cooperate with an enterprise while making measurements there.

TP2 *Employees, TPS' managers, enterprises – TPS' tenants.*

There are some standards which must be fulfilled, but all other things must dynamically change. If it is good, it is ok. Everything is forced by the market, we are trying to keep high standards. Usually, we try bottom-up implementation, because it is the best then. Not in a way that I will come up with something, but it originates from employees. Those who have the first contact with clients. They know their needs and are able to communicate them to the management which then is able to take further actions.

TP3 *Employees, benchmarking of good practices in other TPS as a reference point for modification of good practices by a park, clients.*
Implementation by mapping. There is no such thing as stability, in the long run, every day brings new challenges, needs and this good practice is modified; it must be located in a place where it is implemented and adapted to the culture, needs and resources, and so on.

TP4 *Employees, benchmarking of good practices in other TPS.*
We are implementing everything step by step, checking whether, for example, the 'SME Internationalisation Exchange' (SIE) project works, where we absorb practices and promote our solutions by ourselves.

TP5 *Employees, benchmarking of good practices in other TPS.*
We are trying to carry out implementation activities very hard, because some things which we have borrowed from somewhere, for example, from Spain, a new thing (it does not exist in Poland) we want, let say, to implement in our park. So firstly, we will try and practice it on a smaller group and then implement it somehow among our tenants, and on outside, as well.

TP6 *Employees, benchmarking of good practices in other TPS, including foreign ones, benchmarking of good practices among enterprises operating in TPS in various countries.*
It is not always a constance. It evolves over a long period of time. We expect that there will be an even more competitive solution, that is matched. Sometimes it evolves in a direction we have assumed, and sometimes it evolves in a completely different direction. The example of an incubator, which we saw during the visit in China, showed us that small and medium-sized enterprises are developing, and we have enterprises which work with Apple, Microsoft. So even our small businesses can exchange experiences with Chinese ones.

TP7 *Employees, enterprises – TPS' tenants.*

We implement them and adapt them to our needs and the needs of our clients.

TP8 *Employees while performing operational activities for enterprises/ communicating needs by enterprises operating on the TPS' areas.*
While continuing good practice, we expect that it will continue to prove itself in practice. There are services which do not evolve and it is enough that this standard is good, and we maintain it. It has to be adapted all the time, only the scheme/draft can be the same. One has to evolve all the time to develop. When it comes to the scope of consultancy services for enterprises from other parks, they function to this day, although they are slightly changed.

TP9 *TPS' business model, employees, enterprises – TPS' tenants.*
I will refer to what I have said regarding all services, but also a certain concept of park management. In the long run, that is at the moment of introducing a good practice, we try to define how we understand it, how we will take these actions to implement it. But it is also acting somehow on a living organism on whose other side is a subject which is a recipient, it has expectations, and its expectations do not have to be consistent with what we assumed at the beginning, so as a result of cooperation, this final shape is given.

TP10 *Employees.*
We assume that the enterprises will cooperate with each other, not only door to door, but also building to building, as we have 4 buildings and it must be animated so that everyone knows each other.

TP11 *Employees.*
From the very beginning we have been implementing and applying such a practice, we provide enterprises with a certain infrastructure for the beginning of their operations, then some of the enterprises have their infrastructure, a rented part, and this barrier to entry into the high technology industry is overcome in that way. These are, of course, risky projects, because the investment risk falls on the park, failure of the project made on the park's infrastructure imposes on the park the necessity of looking for a new client and incurring losses in the period of when this infrastructure was out of use. But thanks to this, several and even several dozen very good high technology entities were established and it is worth doing it.

TP12 *Employees while performing operational activities for enterprises and meetings with entrepreneurs.*

Firstly, the Płock Industrial and Technological Park (PITP) tries to use good practices in its own organisation.

However, there are cases where some good practices are not available for use by the PITP, if only due to a different size or organisational structure, or because of different scope of tasks we implement. However, we try to pass on information about good practices further. As a majority shareholder of Centrum Edukacji Sp. z o.o. we pass this knowledge and influence its implementation to this company. We also share some good practices during our meetings with entrepreneurs gathered in the PITP, Chemical Cluster or in general with enterprises which participate in our events.

TP13   *Employees, TPS' managers.*

Before implementation, good practices are discussed at employee meetings, and then presented to the Board of Managers for verification and acceptance. We implement such a practice 'experimentally' to check its effectiveness. If it is accepted, we implement it permanently and if it is necessary, we try to improve it even more so that it can be adapted to our needs.

TP14   *Employees, TPS' managers.*

As we have a good practice, we test it, because it is a very important stage; we modify, for example, a recruitment process, we separate stages and questions being asked. Then, we test again and implement it after the corrections with appropriate provisions, e.g., an ordinance.

Processes change over time because they evolve, are improved and become 'our action' based on a previously discovered and 'followed' good practice. We often pass them on meetings, presentation displays and conferences where we share our knowledge. But, the quality of servicing clients is always the result, nothing changes here.

Answers of technology parks' managers indicate that exploitative learning of good practices permeates all hierarchical levels of these organisations, usually from staff to top management. At the same time, it is exposed in different ways and to different degrees in individual TPS.

Top management strengthens the entrepreneurial behaviours of managers of lower levels and personnel members. Managers of different levels of technology parks' management recognise and cultivate as well as manage and encourage bottom-up initiatives of employees and representatives of other scientific and business institutions, thanks to which trans-

forming results of research and development projects into technological innovations may become much faster and more effective, as well as improve the TPS' effectiveness and functioning, at the same time.

In this bottom-up approach, the entrepreneurial behaviour of personnel members is therefore a key mechanism enabling not only the implementation of aims set for individual technology parks, but above all, to guarantee access to modern infrastructure and technology. In this case, the behaviour of TPS' employees is also strongly associated with exploitative learning of good practices and is of an entrepreneurial, but not inhibitory nature.

These findings indicate that technology parks striving to provide convenient development conditions create effective solutions, especially in the area of systematic matching of their offer to the needs of enterprises which are or will be the TPS' tenants.

Conclusions and Summary

This article presents empirically verifiable knowledge about exploratory and exploitative learning as paths for the development of good practices by the analysed technology parks in Poland were presented. Research findings indicate that neither exploitation nor exploration may be dispensed with as both of these activities are needed to maintain modern and dynamic technology parks which facilitate the development of enterprises operating in the areas administered by them. Image of the analysed technology parks operating in Poland, which emerged after conducting a kind of 'inventory,' shows exploratory and exploitative activities within the same park and a balance between them. This means that ambidexterity is the right path for developing good practices. Practices recognised as the best ones become a desirable standard. Hence there are searching for them and then they are followed. TPS may be treated as ambidexterious organisations because they simultaneously carry out exploratory and exploitative activities, reconciling the tensions which arise between them. Implementation of such a reciprocal approach by the TPS is difficult because exploration and exploitation require different models of organisational learning, as well as various components of resources and capabilities whose implementation provides the technology parks with specific benefits, primarily in the form of increased efficiency and effectiveness of operations as well as growth in innovation. Therefore, the technology parks which may now be falling into 'success trap' or 'failure trap,' will probably transform into ambidextrous organisations over time in or-

der to benefit not only from new knowledge and searching for new good practices, but also from operational efficiency and good practices already used.

PRACTICAL IMPLICATIONS

A very often implemented method of learning good practices lies in the potential of technology parks themselves, the source of which is the accumulated experience of managers and non-managers, both through exploratory and exploitative learning of good practices. This proves that the technology parks' managers do not lose sight of long-term efficiency, which may be ensured by successive good practices. To conclude, it may be stated that the TPS make strategic integration between these activities in an effort to maintain balance. In practice, it means that these technology parks which improve and reproduce good practices will have an advantage over those which are slower in this respect.

LIMITATIONS AND SUGGESTIONS FOR FUTURE IMPLICATIONS

With regard to qualitative research, there is a doubt as to its universality (to what extent will particular cases allow for understanding other cases) and objectivity (or if another respondent perceived a given phenomenon in the same way).

Conclusions and generalisations refer only to part of the parks' population units because the research sample is not representative. However, the research results may form the basis for comparative research in the field of learning of good practices by the technology parks.

In the future, it is recommended to analyse the antecedences of selection of exploratory and exploitative learning by technology parks and results obtained in connection with their implementation. Also, an interesting direction of further research seems to be a deeper recognition of the context not only from the perspective of strategies, structures or processes, but also the competence and commitment of managerial staff.

References

Bratnicka, K. 2015. 'Reconsideration of Organizational Ambidexterity: A Dialectical Multilevel Approach.' *International Journal of Contemporary Management* 13 (4): 67–82.

Bruni, D. S., and G. Verona. 2009. 'Dynamic Marketing Capabilities in Science-Based Firms: An Exploratory Investigation of the Pharmaceutical Industry.' *British Journal of Management* 20 (1): 101–17.

Codling, S. 1995. *Best Practice Benchmarking: A Management Guide.* Aldershot: Gower.

Dani, S., J. A. Harding, K. Case, R. I. M. Young, S. Cochrane, J. Gao, and D. Baxter. 2006. 'A Methodology for Best Practice Knowledge Management.' *Journal Engineering Manufacture* 220 (10): 1717–28.

Daniel, E. M., and H. N. Wilson. 2003. 'The Role of Dynamic Capabilities in E-Business Transformation.' *European Journal of Information Systems* 12 (4): 282–96.

Dyduch, W. 2017. 'Strategiczny projekt organizacji wspierający rozwój zdolności dynamicznych.' In *Strategiczny wymiar dynamicznych zdolności polskich przedsiębiorstw,* edited by K. Krzakiewicz and Sz. Cyfert, 253–66. Poznań: Wydawnictwo UEP.

Gebauer, H., M. Paiola, and B. Edvardsson. 2012. 'A Capability Perspective on Service Business Development in Small and Medium-Sized Suppliers.' *Scandinavian Journal of Management* 28 (4): 321–39.

He, Z. L., and P. K. Wong. 2004. 'Exploration vs. Exploitation: Empirical Test of the Ambidexterity Hypothesis.' *Organization Science* 15 (4): 481–94.

Hoag, B., and C. L. Cooper. 2006. *Managing Value-Based Organizations: It's Not What You Think.* Cheltenham: Edward Elgar.

Im, G., and A. Rai. 2008. 'Knowledge Sharing Ambidexterity in Long-Term Interorganizational Relationships.' *Management Science* 54 (7): 1281–96.

Kerzner, H. 2009. *Project Management: A Systems Approach to Planning, Scheduling and Controlling.* Hoboken, NJ: Wiley.

Kriz, A., R. Voola, and U. Yuksel. 2014. 'The Dynamic Capability of Ambidexterity in Hypercompetition: Qualitative Insights.' *Journal of Strategic Marketing* 22 (4): 287–99.

Levinthal, D. A., and J. G. March. 1993. 'The Myopia of Learning.' *Strategic Management Journal* 14 (S2): 95–112.

March, J. G. 1991. 'Exploration and Exploitation in Organizational Learning.' *Organization Science* 2 (1): 71–87.

Nosella, A., S. Cantarello, and R. Filippini. 2013. 'The Intellectual Structure of Organizational Ambidexterity: A Bibliometric Investigation into the State of the Art.' *Strategic Organization* 10 (4): 450–65.

O'Dell, C., and C. J. Grayson. 2004. 'Identifying and Transferring Internal Best Practices.' *Handbook on Knowledge Management* 1:601–22.

Okreša Đurić, B., and M. Maleković. 2018. 'Knowledge Management and Conceptual Modelling Towards Better Business Results.' In *Proceedings of the ENTRENOVA – ENTERPRISE RESEARCH INNOVATION Conference,* 239–46. Zagreb: IRENET – Society for Advancing Innovation and Research in Economy.

O'Reilly, C. A., and M. L. Tushman. 2008. 'Ambidexterity as a Dynamic Capability: Resolving the Innovator's Dilemma.' *Research in Organizational Behavior* 28:185–206.

Pfeffermann, N. 2017. 'Role of Communication as a Dynamic Capability in Business Model Innovation.' In *Revolution of Innovation Management,* edited by A. Brem and E. Viardot, 191–212. Hamburg: Hochschule Hochschulzentrum Hamburg.

Turulja, L., and N. Bajgorić. 2018. 'Knowing Means Existing: Organizational Learning Dimensions and Knowledge Management Capability.' *Business Systems Research* 9 (1): 1–18.

Urbanowska-Sojkin, E. 2013. *Ryzyko w wyborach strategicznych w przedsiębiorstwach.* Warszawa: Polskie Wydawnictwo Ekonomiczne.

Wang, C. L., C. Senaratne, and M. Rafiq. 2015. 'Success Traps, Dynamic Capabilities and Firm Performance.' *British Journal of Management* 26 (1): 26–44.

Wójcik-Karpacz, A. 2019. *Zdolności organizacyjne w kontekście wyników przedsiębiorstwa: aspekty teoretyczne i wyniki badań empirycznych.* Warszaw: Polskie Wydawnictwo Ekonomiczne.

Wójcik-Karpacz, A., and Mazurkiewicz, Sz. 2015. 'Rola parków technologicznych w kształtowaniu przedsiębiorczego rozwoju przedsiębiorstw.' *Marketing i Rynek* 5: 575–90.

Wójcik-Karpacz, A., and J. Rudawska. 2016. 'Instytucjonalne formy wspierania przedsiębiorczości w Polsce.' *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu* 419:248–64.

Zakrzewska-Bielawska, A. 2017. 'Ambidexterity jako zdolność dynamiczna w odpowiedzi na niepewność otoczenia.' *Studia Oeconomica Posnaniensia* 5 (9): 174–90.

Applied Text-Mining Algorithms for Stock Price Prediction Based on Financial News Articles

Adrian Besimi
South East European University, North Macedonia
a.besimi@seeu.edu.mk

Zamir Dika
South East European University, North Macedonia
z.dika@seeu.edu.mk

Visar Shehu
South East European University, North Macedonia
v.shehu@seeu.edu.mk

Mubarek Selimi
South East European University, North Macedonia
ms21693@seeu.edu.mk

This article includes a developed model and well-defined process that one should undertake in order to contribute in the prediction of the potential stock price fluctuation solely based on financial news from relevant sources. We are providing background information on this topic adding the role of text mining in general, furthermore supporting the idea with the study of relevant research articles to narrow the focus on the problem we are researching. Our proposed model relies on existing text-mining techniques used for sentiment analysis, combined with historical data from relevant news sources as well as stock data. In confirming the model, after the experiment we have provided the results of the simulation, which are opening the ground for further explorations in this sensitive area of prediction.

Introduction

The data produced and the speed at which data is provided on the Internet nowadays has increased to a degree and at a rate that is impossible to process. This trend, on the other hand, has challenged the research in many areas, such as data mining and text mining, which are the focus of

our study. These two areas have emerged in the last decade mainly due to research in artificial intelligence, machine learning, and inferential statistics (Vale 2018).

Stock market data and relevant news associated with fin-tech industry are increasing rapidly as well. Many investors that are handling stock market transactions have a major interest in understanding more about the future of stock markets for the purpose of being able to do an educated guess and/or predict any future investment. Ensuring some level of prediction in market fluctuation can assist investors in the form of decision support systems and integrated with existing automatic trader agents that would ensure better prediction on future trades. Fully predicting the market fluctuation means in practice becoming a billionaire over the night and all the time minimizing financial losses, which is not possible for many reasons. Recent scholars argue that news articles are among influential sources that may affect stock market prices and they should be carefully considered by investors when planning future investments. By definition, any stock price is simply defined by supply and demand of the market, but it is argued by scholars Nikfarjam, Emadzadeh, and Muthaiyah (2010) and Kaya and Karsligil (2010) that another important variable when decision is made to invest or not is also related to verifiable news from financial news sources. This is hard and time-consuming task because it requires to read and analyze a lot of news published on several occasions by various news sources/providers (Nikfarjam, Emadzadeh, and Muthaiyah, 2010; Kaya and Karsligil 2010).

Information published in news articles influence, to a varying degree is influencing the decision of the stock traders, especially if the given information is unexpected. It is important to analyze this information as fast as possible, so it can be used as an advantage to help traders to make trading decisions before the market has had time to adjust itself to the new information (Aase 2011).

One important application of using text mining is text sentiment analysis, also known as opinion mining, a technique that digs deep into the content of the text file and extracts the sentiment of it. Sentiment analysis classifies textual data into positive texts, negative text and neutral text sentiments which is later used for the purpose of categorizing any text documents into the given sentiment (Aase 2011; Khedr and Yaseen 2017).

The model proposed in this paper is going to leverage the Naïve Bayesian classifier for document classification to make a prediction for whether the stocks will go up or down, based on a dataset that is generated from the process proposed later in this paper.

Literature Review

Several scholars, specifically Kim, Jeong, and Ghani (2014) prove to some extent in their work that the relevant news are closely related to stock price movements in the market. With the current trends in big data and content creation on the Internet and the enormous amount of unstructured text data available, the mobile channels, and Social Network services, scholars have attempted to predict stock movements using such text data as in the case of Kim, Jeong, and Ghani in 2014.

Many scholars tried different approaches in research to prove that there is a potentially strong correlation among financial news articles and stock price fluctuations, as is the case of Khedr and Yaseen (2017) that we mentioned earlier. In their paper they propose an approach to use sentiment analysis in financial news, along with features extracted from historical stock prices to apply prediction for the future behaviour of stocks. According to their findings, the proposed model has achieved high accuracy using sentiment analysis in categorizing news polarities by applying Naïve Bayes algorithm. In their case the accuracy of the model is up to 86.21%. By moving on with their experiment in prediction, during their next attempt in analyzing these news articles, they have included numerical attributes which in their case increased the accuracy to 89.80%.

The paper, published by Hagenau, Liebmann and Neumann (2013), examines the hypothesis if any stock price prediction based on textual content from the financial news can be further improved. In this paper, the authors have upgraded the text mining methods by adding expressive feature to represent the text and by adding more variables, such as employing market feedback in the feature selection process. According to the authors, this selection of the features does significantly improve the accuracy due to the fact that this approach removes the unnecessary so called 'less-explanatory features,' i.e., noise, which itself helps the classifier to overcome the over-fitting during classification of the text. In the case when the feedback-based feature selection is combined with 2-word combinations, the authors results show an accuracy of 76%. These results are different from common sentiment analysis approaches since the 2-words combination gives more information and potentially more meaning to the sentiment classification.

A lot of research has been carried by scholars in the area of prediction of stocks as well. A project by Joshi, Rao, and Bharathi (2016) is taking financial news articles about a given company, and they use these data to try to predict the future movements of the stock again by applying sen-

timent analysis. The approach is like in the other cases with an idea to identify how stocks have reacted if news has polarity. Authors in this case have taken the past three years of data from Apple Inc. stock prices as well as news articles. Similarly, to previous scholars, the polarity of the news is labelling these articles and based on these data they are building the training set. The approach in this paper is dictionary-based that contains for positive and negative words that is build based on financially specific words. Further, they have pre-processed the data which resulted in having their own finance specific stop words and dictionary. Using their own dictionary, they have implemented three models for classification and tested them. After comparing the results, they have concluded that Random Forest algorithm resulted in better accuracy for the test cases ranging from 88% to 92%. This algorithm was followed by Support Vector Machines with again very good accuracy of 86%. In their case the Naïve Bayes algorithm performance was the lowest with 83%.

There is some promising research published that applies deep learning techniques and has resulted with higher accuracy ratings. Of interest is the published paper by Tabari et al. (2019) that shows a comparison of diverse algorithms specifically applied in stock market tweets. This research shows quite promising results, with accuracy ranging up to 92.7 % (using Convolutional Neural Networks). However, even though deep learning approaches can be scaled for using news articles, other authors report much lower accuracy rates Kim and Jeong (2019).

One major advantage of using Naive Bayes algorithms is its well-known ability to improve by introducing new data. In our case, previously analyzed news articles can be fed to the algorithm and treated as prior probability. With this, new, previously unknown words will gain weight and affect prediction when encountered in the future. This is one major drawback of the proposed model from Joshi, Rao, and Bharathi, mentioned above, since it only works for a pre-defined dictionary of words.

Another similar approach of finding the correlation amongst the content of news articles and stock prices for the purpose of predicting the stock markets was implemented by Kaya and Karsligil (2010). They collected news articles published in the last year period and combined with the stock prices for the same period. These articles were then labelled as positive or negative sentiments categorization based on their effects on stock price. Their approach is a little bit different in the sense that for them it was important to use the price changes for categorization of the news. While analyzing the textual data, authors use and approach of word

doubles of a noun and a verb as features and not only single words. The support vector machines (svm) method was used in this case which resulted with 61% accuracy.

These scholars and articles mentioned in the section above are the core of our model and study that we conducted.

Methodology

PROBLEM DEFINITION

Financial analysts that are handling investments and transactions in stock markets around the world have a huge headache on making decisions that will be effective and bring more money to the investor or maximizing profits by trading. They are aware that any news, either good or bad can directly affect the stock market. The job of these experts relies on analyzing everything from the media outlet. This is time consuming and the amount of data is getting larger all the time. The methodology that we are arguing as many other scholars mentioned above, including also Falinouss (2017), is that an advanced text mining algorithm can assist these experts and provide them with knowledge just by processing resources related to text and news.

The price movements from the past are not always a good indicator on the future movements and are not a guarantee of smart investment, which makes news articles analysis a better predictor on stock market movements. Falinouss in 2007 proposed to research about the impact of textual data in predicting the financial markets movements. In his study he also developed a system which uses similar approaches as previous case of text mining techniques and their influence on the stock market. This according to Falinouss (2007) can help financial analysts to act immediately upon new news articles as they get published.

We propose a model of predicting stock price fluctuations or movements by analyzing financial news articles on one hand and historical stock prices on the other hand. To accomplish this objective, a complete process of data mining and text mining was developed to predict the price movements for the 3 companies listed public, which are explained in the subsection below.

PROPOSED MODEL FOR STOCK PREDICTIONS
BASED ON FINANCIAL NEWS

In our study we worked towards analyzing data, concretely news articles and historical stock prices to make future predictions about stock direc-

tion. To achieve this, qualitative and quantitative data are crucial. Many steps are conducted to achieve the aim of this research, starting from data gathering.

The data is collected for a period of one year, starting from the 1st of March 2018 until 1st of March 2019. In order to make the prediction we used different variables, such as the polarity of the news (either positive or negative), the rate of change in stocks quotes (an average of 5 days), a source of the news article as well as the company name.

The following is the process consisted of eight steps needed to be performed in order to predict stock price from the financial news:

1. Identifying the news sources and targeted companies
2. Data collection and data cleaning of news articles
3. Sentiment Analysis of news articles
4. Data collection of stock prices
5. Calculating Rate of Change (ROC)
6. Categorizing the data
7. Applying Naive Bayesian classifier
8. Training

Identifying the news sources and targeted companies is crucial to understand the data. The information collected must be relevant and trustworthy. As such, the relevant data from financial news articles from top reliable sources have been identified as: The Washington Post, CNN, MarketWatch, BGR, Fox Business, The Street, The Verge and Breitbart. The targeted companies for our study are: Tesla, Facebook and Apple. News sources are proven to be reliable in the market as the most unbiased, whether the targeted companies have been chosen randomly from technology, software and automotive industry. Tesla has been added because it is a typical example of a lot of news noise and several fluctuations of stock prices.

Data collection and data cleaning of news articles. Links of the news from the sources mentioned in step 1 are collected using Web Scraper extension of Google Chrome browser. After having all the links, we built a python script based on Scrapy (an open source and collaborative framework for extracting the data you need from websites, see http://www.scrapy.org). framework that is extracting data from the links and organizing them in the following structure: article's Title, date, author and the text content. Appropriate data cleansing has been applied

to remove unnecessary HTML tags as well as to format the data from different sources to one standard (see table 1 and table 2).

Sentiment analysis of news articles was applied to every news record based on the news content by using Vader Sentiment Analysis. VADER (Valence Aware Dictionary for sEntiment Reasoning) is a pre-built sentiment analysis model included in the NLTK (Natural Language Toolkit) package. It can give both positive/negative (polarity) as well as the strength of the emotion (intensity) of a text. VADER however is focused on social media and short texts, unlike Financial News which are almost the opposite. We updated the VADER lexicon with words plus sentiments from other sources/lexicons such as the Loughran-McDonald Financial Sentiment Word Lists, to be appropriate for our collected financial news (Yip 2018). At the end of this step we had the polarity of the news content recorded in our dataset.

Data collection of stock prices for each of the targeted companies was done from Yahoo Finance portal, where the following information was collected: *date, open price, high, low, close price, volume* and *Adj close*. These data are important for correlation with the appropriate news from our first data set.

Calculating Rate of Change (ROC). The ROC and *Future ROC* are the two variables that are calculated from the data set from Step 4. The rate of change (ROC) in stocks in an average of 5 days is an existing formula that refers to the last 5 days of stock fluctuation. In our case we also added a column with the Future ROC (the ROC after 5 days), having in mind that the effect of this positive or negative news will be reflected in the future and not the past. Since we are dealing with historical data, the Future ROC is easy to calculate.

Categorizing the data must be done in order to apply Naive Bayesian classifier. In the data set that we have all news collected with their features, we added two new columns: *Sentimentof_text* that could be 'positive' if the sentiment score is greater than zero and 'negative' if the sentiment score is less than zero. We don't take in consideration the neutral score of the text content because that could result in majority of neutral results. The second column is the *ROC_Sentiment* that can be 'positive' if the Future ROC is greater than zero and 'negative' if the future ROC is less than zero.

Naive Bayesian classifier was used to make the prediction of the future stock movements. The naive Bayes applies the well-known Bayes' theorem, where by using a 'naive' assumption that any set of features are

independent for a given class (Tang, Bo; Kay, Steven; He, Haibo; 2016). To prepare the data set to make predictions with the NB, we added a new column with the name class that is 'UP' if the *Sentimentof_text* is 'positive' and the *ROC_Sentiment* is 'positive,' and if the *Sentimentof_text* is 'negative' and the *ROC_Sentiment* is 'negative' then the class is *'down,'* otherwise is *'neutral'* classification. The training dataset results are summarized in table 3.

Training. The data that is collected (see table 1 and table 2) contains records for 12 months from which 10 months will be used to train the model and the last 2 months will be used for the test set, to evaluate how it performs. In total 18236 records will be used as training dataset and the remaining 1990 records (roughly 10%) out of 20226 will be used as a test set.

We created 2 models to see how they perform. The following variables are used to train and test the first model: *Source, Company, Sentimentof_text* and the *5-day ROC,* while in the second model only variables of: *Source, Company* and *Sentimentof_text.*

Results

As explained in the steps undertaken to perform our prediction, the data collection results are shown below. We succeeded to collect the news articles from 8 different news sources, totalling 20226 news articles, split into table 1 for Training Set (18236 records/articles) and table 2 for Test Set (1990 records/articles).

The training dataset results are summarized in table 3, where for each company in our target the list of classification results is shown. As a general finding is that the algorithm applied classifies 15.71% of the articles in the training set as *'down'* (meaning the stock will go down in the following days), 50.71% is classified as *'neutral'* (there is no clear picture on what the prediction will be) and 33.59% of the data as *'up'* (meaning stock will go up). The *'up'* classification is relevant to our study and can be used for simulating investments on our test data from the test set.

In the first prediction that uses the following variables: *Source, Company, Sentimentof_text* and the *5-day ROC* model, the test set classification from 1900 records being tested, resulted in 564 *'down'* and 1426 *'up'* classes for stock price direction were predicted. The achieved accuracy of 94.29% in this prediction model shows that there is a very high chance to predict the stock price movements. By this, our arguing that based on several attributes from new articles we can reach a certain level of predic-

TABLE 1 Total News Articles Obtained for Apple, Tesla and Facebook Organized by Source For Training Set (March 2018–December 2018)

Variable	Categories	Frequencies	Percentage
Source	BGR	1073	5.884
	Breitbart	435	2.385
	CNN	687	3.767
	Fox Business	813	4.458
	The Street	3810	20.893
	The Verge	2847	15.612
	The Washington Post	6051	33.182
	Market-Watch	2520	13.819
Company	Apple	7591	41.626
	Facebook	7513	41.199
	Tesla	3132	17.175

TABLE 2 Total News Articles Obtained for Apple, Tesla and Facebook Organized by Source for Test Set (January 2019–March 2019)

Variable	Categories	Frequencies	Percentage
Source	BGR	185	9.30
	Breitbart	167	8.39
	CNN	211	10.60
	Fox Business	147	7.39
	The Street	590	29.65
	The Verge	603	30.30
	The Washington Post	87	4.37
	Market-Watch	0	0
Company	Apple	1144	57.49
	Facebook	416	20.90
	Tesla	430	21.61

tion, and give directions to financial experts, is valid as in our case where we reached a certain level of prediction based on several attributes. Still, though Efficient Market Hypothesis (EMH) clearly states that financial stock prices cannot be predicted, because there is no 100% prediction. The accuracy rate of the first model is high and there is a strong relationship between financial news and stock price movements.

TABLE 3   Training Set Classification Data Organized by Company and Frequency

Company		Down	Neutral	Up	Total
Apple	N	1,006	3,930	2,655	7,591
	%	5.52	21.55	14.56	41.63
Facebook	N	1,390	3,683	2,440	7,513
	%	7.62	20.20	13.38	41.20
Tesla	N	468	1,634	1,030	3,132
	%	2.57	8.96	5.65	17.17
Total	N	2,864	9,247	6,125	18,236
	%	15.71	50.71	33.59	100.00

To test the second model 3 variables as input are given: *Source, Company* and *Sentimentof_text* to predict the class up, down or neutral. In comparison with the first model that has an accuracy of 94.29%, the second model has 49.49% which is significantly with lower accuracy than the first model, that has just one more variable – the *5-day* ROC. This prediction rate is less than the guessing probability (50%), and as such this model is irrelevant. It can be stated that aside from sentiment of text, stock data variables as in this case of '*5-day* ROC' plays a vital role in prediction of the future stock price movements. Only a single relevant variable as the rate of change of a given stock for the last five days does impact the prediction in the model where we deal with stock rates. It is important to include the most relevant variables when applying prediction in the FinTech industry.

It can be stated that the weak form of EMH is true. Weak form efficiency states that all future price movements follow a random walk, unless there is some change in some fundamental information. It does not state that prices adjust immediately in the advent of new fundamental information, which means that some forms of fundamental analysis and news article analysis might provide excess returns. This is because they trade on new information and does not use any historical information to look for patterns.

Simulation

In order to apply and validate the text-mining algorithms and classification techniques mentioned in this paper to predict the financial news, we have conducted a simulation based on exact data from the market and

the cross reference of the dates with the predictive algorithm. Simulation data is conducted with 10.000$ investments per company per day.

Table 4 contains sample data for our simulation. Bellow is the description of the fields in order to better understand when the 'investment' is to be made:

- *DoW* represents the Day of the week that we have news articles processed and categorized the sentiment. It is an important variable to understand if we are dealing with weekends.
- *Date* is the actual date when the news have appeared and the stock Close price date.
- *Company* is company name that we have collected data.
- *Close* is the closing price of the stocks for the given date and company.
- *5-day* ROC is the historical rate of change for the last 5 days.
- *Invest?* is a probability calculation that is derived as an average of all sentiments for all news in our dataset. For instance: *if there is one news article and it is positive towards Facebook, then this percentage is 100%. In case if there is one negative and 3 positive news for a given company, then the probability is 75% to invest.*
- *Investment* is the same amount used for the purpose of simplification and simulation: $10,000.00.
- *Profit/Loss* is calculated based on the prediction if we go for investment. For instance, *the second row results in investment for Facebook. In this case 'investment' probability is 100% and in the simulation we 'invest' total $10,000. The difference in Close price for Facebook for 1/1/2019 and Facebook for 1/2/2019 in stock prices is ($10.000/135.68 = 76.283 stocks, or equal on 2nd of January as 76.283x$135.68 = $10,350.14) and profit/loss is $350.14.*

Rules and assumptions for 'investment' are the following:

- The 'investment' is done a day after the news have been published (the effect of the published news will be seen the next day). In case of weekends, the investment is done next Monday.
- The purchased stocks will be sold the next day, for simplification the cost of closed stock price for that day has been taken into the simulation.
- The difference represents either Profit/Loss.

TABLE 4 Profit and Loss Sample Table for the Test Set simulation

(1) (2)		(3)	(4)	(5)	(6)	(7)	(8)
3	1/1/2019	Apple	157.740	18.423429	80.00	10,000.00	11.41
3	1/1/2019	Facebook	131.089	−60.73621	100.00	10,000.00	350.14
3	1/1/2019	Tesla	332.799	110.98007	100.00	10,000.00	(681.49)
4	1/2/2019	Apple	157.919	20.466856	67.74	10,000.00	(996.07)
4	1/2/2019	Facebook	135.679	−59.230769	66.67	10,000.00	(290.39)
4	1/2/2019	Tesla	310.119	96.601993	57.14	10,000.00	(314.72)
5	1/3/2019	Apple	142.190	8.4674699	54.93	10,000.00	426.89
5	1/3/2019	Facebook	131.740	−60.414660	62.50	10,000.00	471.38
5	1/3/2019	Tesla	300.359	90.197561	54.55	10,000.00	576.97

NOTES Column headings are as follows: (1) DOW, (2) date, (3) company, (4) close, (5) 5-day ROC, (6) invest (%), (7) investment ($), profit/loss ($).
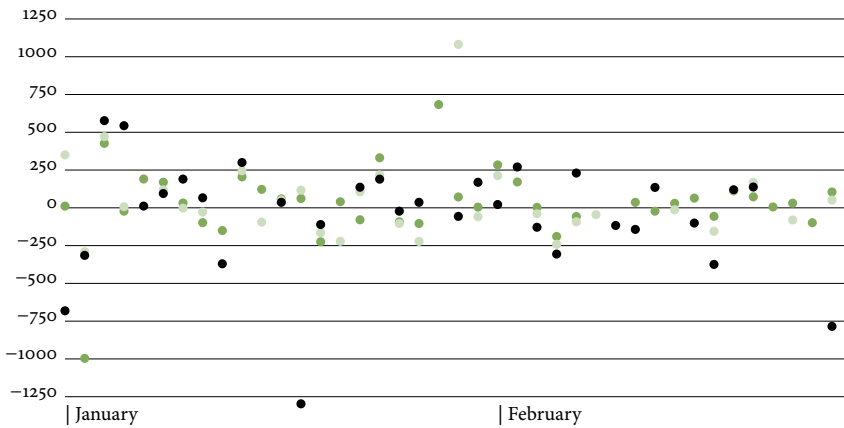

FIGURE 1   Chart of Profit/Loss simulation on Test Set Data Based on Classification Model (Naive Bayes) Used in This Study (dark green – Apple, light green – Facebook, black – Tesla)

Figure 1 clearly shows the fluctuation on the investment simulation based on real data from stock closing prices, for the three companies in combination with our model as explained in previous table. *Tesla* data shows more fluctuation and as such we excluded in our next chart (figure 2) to see if the model prediction can be used for investment.

Figure 2 shows improvements and majority of profit cases on the investment simulation when *Tesla* is excluded, and the only trade is done with *Apple* and *Facebook*. In order to support this figure, table 5 represents
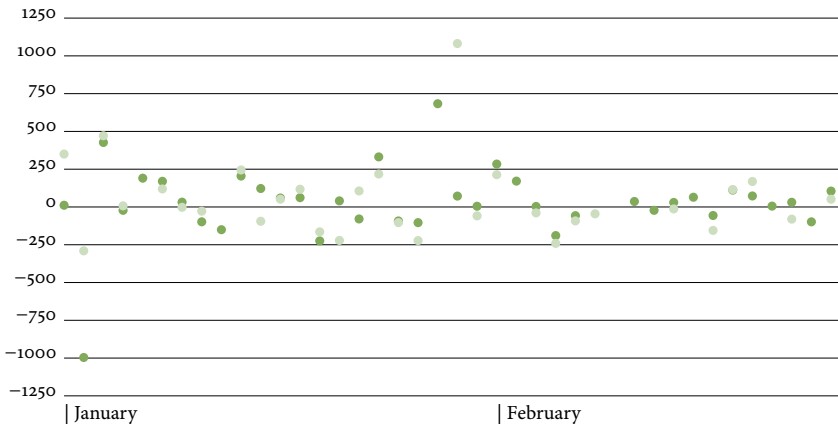
FIGURE 2   Chart of Profit/Loss Simulation on Test Set Data Based on Classification
Model (Naive Bayes) Used in This Study, Excluding Tesla Company (dark
green – Apple, light green – Facebook

real data for our simulation for the 2 months test data used in this exper-
iment. The values show daily profit/loss for all three companies based on
the news sentiments that we processed.

In table 5, Profit and Loss table for Apple, Facebook and Tesla based
on the Test Set simulation with \$10.000 investments was conducted. The
simulation conducted does not show 100%-win case for the classification
of stock prediction and as such it does not apply to all companies. The
difference where there are better results relies on the targeted compa-
nies, such as Apple and Facebook, which are more stable ones rather than
Tesla, which as a case had different fluctuations that in long term did not
bring good results in our simulation.

In this simulation, Tesla's predictions based on our model result in Loss
where the other two companies Apple and Facebook in the long run result
in profit of around \$2.600 if the algorithm is run every day where the data
is available with daily amounts of \$10.000 per company investments.

Conclusion

The trading of stock in public companies is an important part of the econ-
omy, so in this study stocks have been analyzed through using data min-
ing and text mining techniques to make a prediction for stock price di-
rections of the stocks for 3 companies listed public.

To achieve a prediction we gathered data, collected relevant financial
news articles from reliable sources with both qualitative and quantitative

TABLE 5   Profit and Loss Table for Apple, Facebook and Tesla Based on the Test Set Simulation

Date	Apple	Facebook	Tesla	Grand total ($)
1 January	11.411	350.141	−681.490	−319.939
2 January	−996.074	−290.388	−314.717	−1601.179
3 January	426.893	471.382	576.975	1475.250
6 January	−22.258	7.249	543.611	528.602
7 January	190.631		11.644	202.275
8 January	169.817	119.273	94.826	383.916
9 January	31.962	−2.080	190.234	220.116
10 January	−98.180	−27.739	66.383	−59.536
13 January	−150.371		−370.328	−520.699
14 January	204.667	244.859	299.940	749.466
15 January	122.166	−94.663		27.503
16 January	59.378	51.512	36.411	147.301
17 January	61.594	117.329	−1297.112	−1118.189
21 January	−224.461	−164.622	−110.501	−499.584
22 January	40.443	−221.590		−181.147
23 January	−79.262	106.029	136.306	163.073
24 January	331.369	218.062	189.702	739.132
27 January	−92.545	−103.348	−22.219	−218.113
28 January	−103.647	−222.418	36.439	−289.626
29 January	683.347			683.347
30 January	72.012	1081.638	−56.676	1096.974
31 January	4.807	−58.791	169.044	115.060

Continued on the next page

data. This combined with the second type of data of stock prices were used in our study. For every article, a sentiment score (positive and negative) of the text content is calculated.

We have found out that a model that does not include price fluctuations and wholly relies on text content to predict the stock price fluctuation is not accurate at all. Including additional variables improves significantly the prediction. In our case the variable '5-day ROC' plays an important role in predicting the future stock prices.

This article, except for proposing the model used and the process undertaken to arrive at the desired data set, contains results from the sim-

TABLE 5   *Continued from the previous page*

Date	Apple	Facebook	Tesla	Grand total ($)
3 February	284.050	213.626	21.781	519.456
4 February	171.094		270.382	441.477
5 February	3.445	−39.145	−128.520	−164.220
6 February	−189.394	−241.070	−306.096	−736.560
10 February	−57.509	−92.034	230.216	80.673
11 February		−45.238		−45.238
12 February			−116.737	−116.737
13 February	36.433		−142.779	−106.347
14 February	−22.249		135.300	113.052
18 February	29.926	−12.924		17.002
19 February	64.354		−100.773	−36.419
20 February	−56.386	−155.020	−374.471	−585.876
21 February	111.657	115.596	119.492	346.746
24 February	72.845	168.633	137.762	379.240
25 February	5.740			5.740
26 February	30.975	−80.424		−49.449
27 February	−98.359			−98.359
28 February	105.112	51.409	−784.356	−627.836
Grand Total	1135.432	1465.244	−1540.327	1060.350

NOTES   Empty fields in the rows above are possible due to the fact that not all companies chosen for investment every day, based on the news sentiment probability of investment.

ulation of the model. Previous models for sentiment analysis of financial news articles are limited in news articles from relevant sources and as such, based only on sentiment of the news do not provide enough information for future movements. Our model in this paper adds more variables to the dataset in order to give more accuracy to the prediction.

As the results are probabilistic weights (predictions), the simulation we conducted does not show 100%-win case for the classification of stock prediction and as such it does not apply to all companies. The difference where we have better results relies on the targeted companies, such as Apple and Facebook, which are more stable ones rather than Tesla, which as a case had different fluctuations that in long term did not bring good results in our simulation.

Future work would follow with the research on the characteristics of

the companies that would fit to the model, with the tendency to prove that the proposed model is universal for the specific companies within specific variables, adding more tests and simulations as well.

It could also prove valuable to evaluate deep learning algorithms for the purpose of sentiment analysis. These algorithms are yet to show good results when larger text bodies are used, however, for short tweets they are very accurate. Furthermore, in this paper we evaluated only a three-class problem in the context of sentiment analysis. It would be of interest to approach using a multiclass prediction model and see how diverse sentiments would affect the stock market. This is an area where nn algorithms would prove to be much more beneficiary. Finally, even though we consider only highly respected news sources for our analysis, we could further drill down to add the author of the source as an attribute. A renowned author would probably have more weight in the context of affecting the stock market with his articles.

References

Aase, K. G. 2011, 'Text Mining of News Articles for stock Price Prediction.' Master's thesis, Institutt for datateknikk og informasjonsvitenskap.

Falinouss, P. 2007. 'Stock Trend Preidction Using News Articles: A Text Mining Approach.' Master's thesis, Luleå University of Techonology.

Hagenau, M., M. Liebmann, and D. Neumann. 2013. 'Automated News Reading: Stock Prices Prediction Based on Financial News Using Context-Capturing Features.' *Decision Support Systems* 55 (3): 685–97.

Joshi, K., J. Rao, and H. N. Bharathi. 2016. 'Stock Trend Prediction Using News Sentiment Analysis.' *International Journal of Computer Science & Information Technology* 8 (3): 67–76.

Kaya, M. Y., and M. E. Karsligil. 2010. 'Stock price Prediction Using Financial News Article.' In *Proceedings of the 2nd International Conference on Information and Financial Engineering,* 478–82. Chongqing: ieee.

Khedr, A. E., and N. Yaseen. 2017. 'Predicting Stock Market Behavior Using Data Mining Technique and News Sentiment Analysis.' *International Journal of Intelligent Systems and Applications* 9 (7): 22–30.

Kim, H., and Y.-S. Jeong. 2019. 'Sentiment Classification Using Convolutional Neural Networks.' *Applied Sciences* 9 (11): 2347. https://doi.org/10.3390/app9112347

Kim, Y., S. R. Jeong, and I. Ghani. 2014. 'Text Opinion Mining to Analyze News for Stock Market Prediction.' *International Journal of Advances in Soft Computing and its Application* 6 (1): 1–13.

Nikfarjam, A., E. Emadzadeh, and S. Muthaiyah. 2010. 'Text Mining Approaches for Stock Market Prediction.' In *Proceedings of the 2nd Inter-*

national Conference on Computer and Automation Engineering, 256–60. Singapore: IEEE.

Tabari, N., A. Seyeditabari, T. Peddi, M. Hadzikadic, and W. Zadrozny. 2019. 'A Comparison of Neural Network Methods for Accurate Sentiment Analysis of Stock Market Tweets.' In *ECML PKDD 2018 Workshops,* edited by C. Alzate, A. Monreale, L. Bioglio, V. Bitetta, I. Bordino, G. Caldarelli, A. Ferretti, R. Guidotti, F. Gullo, S. Pascolutti, R. G. Pensa, C. Robardet, and T. Squartin. Cham: Springer.

Vale, M. N. d. 2018. 'Dow Jones Index Change Prediction Using Text Mining.' Instituto Alberto Luiz Coimbra De Pós-Graduação E Pesquisa De Engenharia, Rio de Janeiro.

Yip, J. (2018), 'Algorithmic Trading Using Sentiment Analysis on News Articles.' https://towardsdatascience.com/https-towardsdatascience -com-algorithmic-trading-using-sentiment-analysis-on-news-articles -83db77966704.

Ocenjevanje kreditnega tveganja v podjetništvu: izbira lastnosti

Mirjana Pejić Bach, Nataša Šarlija, Jovana Zoroja, Božidar Jaković in Dijana Ćosić

Cilj raziskave je raziskati vpliv različnih algoritmov za izbiro značilnosti v procesu ocenjevanja kreditnega tveganja pri financiranju podjetništva s strani hrvaških finančnih institucij. Uporabljeni so demografski in vedenjski podatke, pa tudi različni algoritmi za razvoj klasifikacijskega modela. Poleg tega je ocenjenih več algoritmov za izbiro spremenljivk, ki med drugim temeljijo tudi na natančnosti klasifikacije. Algoritem zaporedne minimalne optimizacije v kombinaciji z algoritmom razreda CfcSubsetEval in ConsistencySubsetEval za izbiro spremenljivk se je izkazal za najbolj natančnega pri napovedovanju kreditne sposobnosti in posledično tudi najbolj uporabnega za ocenjevanje kreditnega tveganja.

Ključne besede: rudarjenje podatkov, ocenjevanje kreditov, izbira spremenljivk, drevesa odločitev, klasifikacija

Klasifikacija jel: c61, e51

Managing Global Transitions 17 (4): 265–287

Glavni dejavniki razvoja konceptov vodenja računovodstva v Rusiji: globalne ambicije proti lokalnemu načinu dela

Pavel Lebedev

Članek obravnava glavne dejavnike, ki so bili predhodno oblikovani in trenutno vplivajo na razumevanje in razvoj vodenja računovodstva v Rusiji, s poudarkom na specifičnih vlogah različnih deležnikov in institucij. Avtor je v tej raziskavi deloval v skladu z utemeljeno teoretično perspektivo. Zbiranje podatkov je temeljilo na arhivskih metodah, z interpretativnim pristopom k analizi. Na začetni stopnji je bil uporabljen tematski model pripovedne analize, ki mu je sledil intenziven večstopenjski postopek kodiranja. Ugotovitve, vključno z novonastalimi teoretičnimi predlogi, podajajo raziskovalne in pojasnjevalne poglede na razvoj in trenutno stanje vodenja računovodstva v Rusiji. Pridobljeni rezultati kažejo, da je - kljub dejstvu, da je vodenje računovodstva v Rusiji še vedno razvijajoč se koncept - vpliv zunanjih dejavnikov, ki jih zastopajo različni deležniki in institucije, na razvoj vodenja računovodstva v Rusiji spontan, zanj pa je značilno pomanjkanje organiziranega vodstva in nevzdržen oportunizem, ki ga poganja nevednost in samozaverovanost glavnih akterjev.

Raziskovanje in izkoriščanje dobrih praks tehnoloških parkov na Poljskem: empirične ugotovitve
Anna Wójcik-Karpacz, Jarosław Karpacz in Joanna Rudawska

V trenutnih gospodarskih ramerah, v katerih podjetja spreminjajo poslovne modele in iščejo nove priložnosti in pomene za politiko gospodarske rasti v razmerah, v katerih celo vlada prepoznava potrebo po podpori mikro, malim in srednje velikim podjetjem, je analiza učenja dobrih praks s pomočjo tehnoloških parkov (TP) z znanstvenega vidika zelo pomembna. Cilj tega članka je opredeliti načine učenja dobrih praks s pomočjo tehnoloških parkov, ki delujejo na Poljskem. Tehnike zbiranja podatkov vključujejo standardizirane in nestrukturirane intervjuje, opravljene s 14 direktorji tehnoloških parkov. V raziskavi je bil pristop obojestranskosti sprejet kot tisti, ki se najbolj prilega trenutnemu trendu strateškega upravljanja, temelječega na paradoksih, ob predpostavki, da hkrati potekata tako proces raziskovalnega, kot tudi uporabnega učenja. Rezultati raziskav so pokazali, da v analiziranih TP-jih ohranjanje privlačne ponudbe podpore sedanjim in potencialnim bodočim uporabnikom temelji na sistematičnem prilagajanju že priznanih dobrih praks in hkrati eksperimentiranju z novimi dobrimi praksami. Ugotovitve raziskav so prikazale obojestranski pristop kot pogoj, ki ni le mogoč, temveč nujno potreben za nadaljnji razvoj TP.

Algoritmi rudarjenja besedil, uporabljeni za napovedovanje cen delnic na podlagi člankov v finančnih novicah
Adrian Besimi, Zamir Dika, Visar Shehu in Mubarek Selimi

Članek opisuje razvit model in natančno opredeljen postopek, ki bi ju bilo mogoče izvajati z namenom dodatne pomoči pri napovedovanju potencialnega nihanja cen delnic izključno na podlagi finančnih novic iz ustreznih virov. Zagotavlja osnovne informacije o tej temi, z dodatkom opisa vloge rudarjenja besedil, ideja pa je podprta tudi s preučevanjem ustreznih raziskovalnih člankov, z namenom usmerjanja pozornosti na problem, ki je predmet raziskave. Naš predlagani model se opira

na obstoječe tehnike rudarjenja besedil, ki se uporabljajo za analizo ob-
čutkov, v kombinaciji s predhodnimi podatki iz ustreznih virov novic in
podatki o stanju delnic. Pri potrditvi modela smo po poskusu podali še
rezultate simulacije, ki odpirajo možnosti za nadaljnje raziskave na tem
občutljivem področju predvidevanja.

Ključne besede: rudarjenje besedil, finance, novice, crawling, zaloge,
cene, napovedi

Klasifikacija JEL: C89, G17

Managing Global Transitions 17 (4): 335–351