sciendo

Financial Internet Quarterly 2021, vol. 17 / no. 1, p. 50-61

DOI: 10.2478/fiqf-2021-0006

Financial Internet Quarterly

APPLICATION OF PREDICTIVE METHODS TO FINANCIAL DATA SETS

Reza Habibi[1]

Abstract

Financial data sets are growing too fast and need to be analyzed. Data science has many different techniques to store and summarize, mining, running simulations and finally analyzing them. Among data science methods, predictive methods play a critical role in analyzing financial data sets. In the current paper, applications of 22 methods classified in four categories namely data mining and machine learning, numerical analysis, operation research techniques and meta-heuristic techniques, in financial data sets are studied. To this end, first, literature reviews on these methods are given. For each method, a data analysis case (as an illustrative example) is presented and the problem is analyzed with the mentioned method. An actual case is given to apply those methods to solve the problem and to choose a better one. Finally, a conclusion section is proposed.

JEL classification: G17

Keywords: data mining, machine learning, meta-heuristic technique, numerical computation, operation research, predictive methods

Received: 05.01.2021 Accepted: 10.03.2021

[1] Iran Banking Institute, Central Bank of Iran, Tehran, Iran, e-mail: habibi1356@gmail.com, ORCID: ORCID 0000-0001-8268-0326.

Introduction

Financial analysis refers to standard practices to give stakeholders an accurate depiction of a company's finances, including their revenues, expenses, profits, capital, and cash flow, as formal records that provide in -depth insights into financial information. Financial data analysis is one of the bedrocks of modern business. While you may already know that financial data analysis is important (mainly because it's a legal requirement in most countries), you may not understand its untapped power and potential. It's clear that financial analyzing tools could serve to benefit your business by giving you a more informed snapshot of your activities. Utilizing financial data with the help of online data analysis tools allows you to not only share vital information both internally and externally but also to leverage metrics or insights to make significant improvements to the very area that allows your business to flow.

Each of these financial KPIs is incredibly important because they demonstrate the overall 'health' of a company – at least when it comes to the small matter of money. These types of KPI reports don't offer much insight in the way of a company's culture or management structure, but they are vital to success, nonetheless. Financial analysis is used by owners, managers, employees, investors, institutions, government, and others to make important decisions about a business. If you're considering investing money in a company, it only makes sense that you'll want to know how well that company is doing – according to a standardized litmus test; not measurements that a company has fabricated to make them look good. This is where the importance of financial data analysis comes into play for investors. This also applies to credit vendors and banks who are considering lending money to a company.

In these situations, you will need to gain an accurate understanding of how likely you are to be paid back so that you can charge interest accordingly.

The velocity, variety and volume of financial data sets have exploded. To analyze this increasing volume of data and observations, techniques such as data science is necessary.

Data science has many applications in various fields of finance, especially in financial engineering. Constructing predictive models, running live simulations, analyzing new data sets and storing diverse data sets are some needs of financial data which is solved by data science techniques like sentiment analysis, real time analysis, customer segmentation, big data analysis and related techniques. Predictive analysis and methods have a main role among data science techniques. They have many applications in risk analysis, forecasting market behavior, customer segmentations, high frequency trading, making calculated predictions, running winner strategies, identifying online precursors for stock market moves and many other fields of finance. For a comprehensive review of applications of data science predictive techniques in financial data sets, see Kovalerchuk and Vityaev (2000).

Several techniques are commonly used as part of the predictive methods. Using several types of predictive methods makes the comparison, trend and structure of the data clear at a glance. In addition there are many types of financial data sets, in practice from different points of view such as risk management, money management, sale management, etc. Therefore, many types of predictive methods are needed to analyse a specified financial data set. In this way, in the current paper, to cover most important techniques used in a financial data analysis for different purposes, almost 22 individual or hybrid predictive methods are considered. They are categorized to four classes namely, data mining and machine learning, numerical analysis, operation research techniques and meta-heuristic techniques. The first category contains CART, cross validation, decision trees, particle filter and EM algorithm, bootstrap, Jacknife, K-means, K-neighborhood, kernel density estimation (KDE), naive Bayes and principal component analysis (PCA). The second category, i.e., numerical analysis involves Kuhn-Tucker, numerical computations, and Spline techniques. The operation research category contains integer and quadratic programming methods, dual and sensitivity analyses, data envelopment analysis (DEA) and TOPSIS as a multicriteria decision-making technique. Finally, simulated annealing is studied as a meta-heuristic technique approach. The current paper studies the application of the above-mentioned methods in various financial example models. Sixteen problems are considered and throughout these examples the computational methods are applied to real data sets. Each method is important for any of the purposes which are described. For some of them data analyses are done and for some of them, the mathematical results are presented. This paper can be viewed as a practical instruction of data analysis useful for almost all participants of financial markets. The following Table gives the structure of the method applied in this paper.

Table 1: Applied methods

Method	Studied methods			
Data mining	CART	Cross validation	Decision tree	Particle filter
	EM algorithm	Resamplings	K-means	KNN
Numerical analysis	Kuhn-Tucker	Spline		
Operation research	Integer programming	Dual analysis	Sensitivity	DEA
Meta-heuristic	Simulated annealing			

Source: Author's work

The rest of the paper is organized as follows. In the next section, literature review of the above-mentioned methods is presented. In section 3, nine problems are proposed and using the mentioned methods are solved. An actual case is given to apply those methods to solve the problem and to choose a better method. Finally, a conclusion section is presented.

Literature review

Hwang and Yoon (1981) studied the applications of multiple attribute decision making in many fields such as finance. Breiman et al. (1984) used the CART algorithm in many applications such as finance, economics, engineering, statistics and biology. One famous example of a financial application is the credit scoring of bank customers. Olave and Miguel (2000) used the bootstrap method for forecasting exchange rates. They used GARCH modeling for exchange rates to obtain prediction intervals and to present various measures for predictions. Das (2003) applied the K-means clustering method for hedge fund classification. He used asset class, size of hedge fund, incentive fee, risk level and liquidity of hedge fund as cluster variables. Miyazaki (2003) considered the dual analysis on hedging VaR of a bond portfolio using options. Huck and Guegan (2007) used the k neighborhood method for forecasting in predicting the price of commodities, stock indexes and interest rates. Zhou et al. (2008) applied a particle filtering framework for a randomized optimization algorithm. Sensitivity analysis is done to capture the change in a response variable throughout the change in one variable at a specific time, for example, changes in net present value (NPV) because of change in risk free and discount rates. Break even analysis and calculating operating leverage are two examples of sensitivity analysis, see Winston (2010). Lai (2010) used simulated annealing in multifactor equity portfolio management. Robles-Granda and Belik (2010) proposed a comparison study for machine learning classifiers applied to financial datasets.

Mingione (2011) used the principal component analysis for testing financial stability indices for Jamaica with the aim of forecasting two indices of financial vulnerability. Bulgurcu (2012) applied the TOPSIS technique for financial performance evaluation of technology firms in the Istanbul stock exchange market. Hillier and Lieberman (2012) described the financial applications of Kuhn-Tucker equations. Aparicio et al. (2013) used the DEA technique for benchmarking based on genetic algorithms and parallel programming. Cai et al. (2014) used -dual interior-point methods for linear optimization. He used the complexity analysis based on a new parametric kernel function with a trigonometric barrier term. Patil (2014) studied naive Bayes and Jelinek-mercer smoothing for a heart disease prediction system. Ibanez et al. (2016) used the simulated annealing to study the stability of protein interaction networks in cancer and neurological disorders.

Illustrative examples

Here, illustrative examples are given to propose the financial applications of the above-mentioned methods in four categories, namely data mining and machine learning methods, meta-heuristic approaches, operation research based methods and numerical analysis methods.

Data mining and machine learning techniques

CART method

Better stock price predictions yield better trading systems. There are many good econometric methods

such as the generalized method of moments or random walk theory to predict the movement of stock price. However, they are constructed with many hypothetical assumptions. The classification and regression tree method proposes a computational approach for trading strategies (see Breiman et al., 1984). As follows, this problem is studied. The data in this study comes from the stocks of the first seven corporations listed in the S&P500 involving MMM (denote its return by R_t^1), ABT (R_t^2), ABBV(R_t^3), ACN (R_t^4), ATVI (R_t^5), AAP (R_t^6), AES (R_t^7). Four portfolios are constructed based on Markovitz, CAPM, APT, and Fama and French as follows:

Portfolio 1: (0, 0.45, 0.32, 0, 0, 0.12, 0.11),

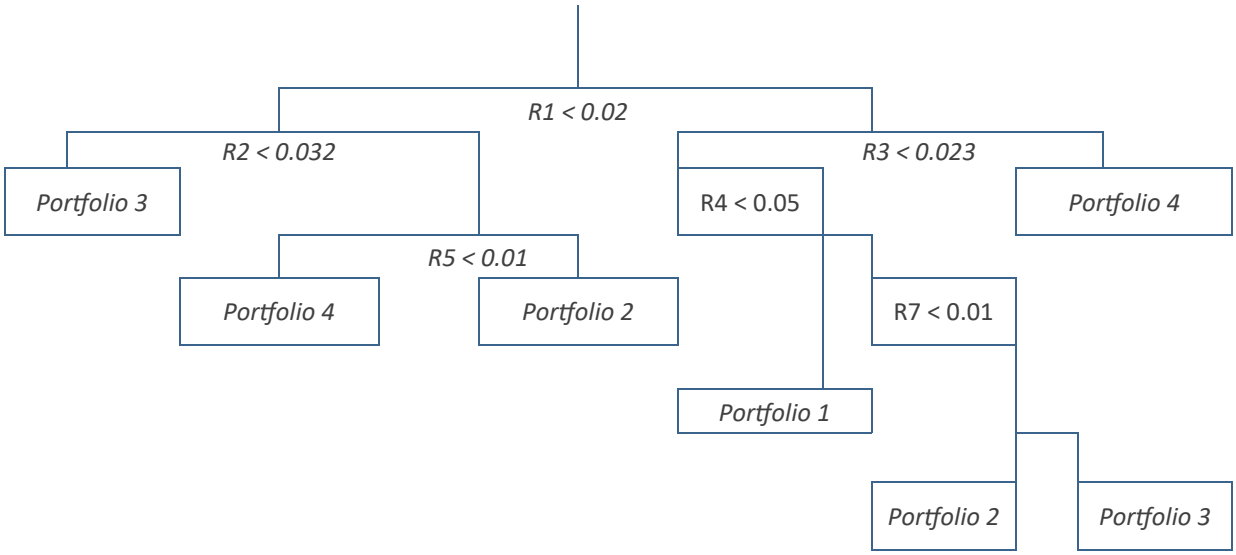
Portfolio 2: (0.35, 0, 0.12, 0.1, 0.1, 0.05, 0.28),

Portfolio 3: (0, 0.3, 0.3, 0.2, 0.2, 0, 0),

Portfolio 4: (0.1, 0.1, 0.2, 0.2, 0.1, 0.08, 0.22).

The classification and regression trees CART are a nonparametric method for construction of decision trees for discrete variables which is called the classification tree and for continuous variables which is referred to as a regression tree. Usually, a large sample divided into two categories exists. The aim is to relate each category to some covariate variables. For example, consider a sample of bank clients. The interest is in predicting whether a specified client may default or not. Covariate variables are age, education, gender and job. The largest tree is designed and then it is revised. Some pollution criteria like a Gini index are calculated. The simplest tree with the minimum pollution criteria is selected. The following diagram presents the result of CART in selection of portfolios. The right direction is "yes" and the left direction is "no". The following Figure gives the related decision tree.

Figure 1: CART algorithm for portfolio selection



Source: Author's work

Naive Bayes and Cross validation methods

The naive Bayes is a classifier method in the machine learning field. In this problem it is used to separate two categories of financial companies with high and low systematic risks. Following Robles-Granda and Belik (2010), the research variables are volatility in the market which is calculated by (high price-low price)/ (high price + low price) and reinvestment rate. The naive Bayes is a non-parametric method for classification using posterior information. Similar to the Bayesian network the posterior of a new generation given the information of the previous generation is calculated and it is used for classification. Naive Bayes is a machine learning technique and it is a learning classifier based on Bayes rule. To derive the posterior f(y|x) use available data to derive likelihood f(y|x) and prior f(y) and then combine this information together with the Bayes rule to obtain the posterior. However, imple-

menting the Bayesian classification to extract likelihood and prior will need estimation of many parameters. The naive Bayes assumes a version of conditional independence. Indeed, the naive Bayes is based on the Bayes rule and conditional independence assumption. Consider attributes $X_1, \dots, X_n$ are given, then

$$P(Y = y_k|X_1, \dots, X_n) = \frac{P(X_1, \dots, X_n|Y = y_k)P(Y = y_k)}{\sum_{k=1}^{\infty} P(X_1, \dots, X_n|Y = y_k)P(Y = y_k)}$$

Therefore, y_k belongs to classifications which maximizes the $P(X_1, \dots, X_n|Y = y_k)P(Y = y_k)$. Indeed,

$$y_k = \text{argmax}_{y_k} P(X_1, \dots, X_n|Y = y_k)P(Y = y_k).$$

Patil (2014) used the naive Bayes to predict heart disease. Here, the data sets are again the first seven companies of the S&P500 like in problem 1. The cross validation is very similar to in-sample-out-sample and rolling analysis of econometric time series. The sample is divided into two folds: test and evaluation sets. The statistical inference is performed on test parts and the performance of the first fold is evaluated by the second fold, i.e., the evaluation fold. Then, the role of test and evaluation folds is replaced and finally an averaging proposes the final decision. The sample may be divided to k-folds. To run the 3 fold cross validation is used. It is seen that the best first category is (ABT, AES, MMM, ACN) and (ABBV, ATVI, AAP).

K-means and K-nearest neighborhood (KNN)

The K-means separates the observation to k homogenous categories by defining k circles. The method first determines the suitable value of k. Then, suitable centers and radiuses are selected to obtain the most homogenous circles. The KNN is also a similar method, however, the best homogenous neighborhood is introduced. It is attempted to propose the smallest k. Indeed, KNN algorithm surveys all cases and using a similarity measure, the new case is classified. It has many applications in statistical pattern recognition. Following Sutton (2012), the algorithm is summarized as follows:

1) a positive integer k is specified, along with a new sample,

2) select the k entries in our database which are closest to the new sample,

3) the most common classification of these entries is found,

4) this is the classification which is assigned to the new sample.

Applying the k-means algorithm with k = 3 to the problem of 3.1, the following category is derived. (ABT, AES, AAP), (ACN, ABBV) and (ATVI, MMM). The KNN method with k = 3 results the following categories (ABT, AAP), (MMM, ATVI), (ABBV, ACN, AES).

Jacknife vs Bootstrap

Jacknife belongs to the class of re-sampling methods like the bootstrap method. Suppose that $\hat{\theta}_n = f(x_1, \dots, x_n)$ is the estimate of parameter θ using the sample $x_1, \dots, x_n$. The Jacknife estimate of is given by

$$\frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{-i},$$

where $\hat{\theta}_{-i} = f(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$. Jacknife removes the bias of $\hat{\theta}_n$.Next, suppose that $x^*_{1,i}, \dots, x^*_{n,i}$, i = 1, 2, …, B is the B resamples of $x_1, \dots, x_n$ with replacement. Let $\hat{\theta}^*_i$ be the estimate of θ using $x^*_{1,i}, \dots, x^*_{n,i}$. Then, the bootstrap estimate of θ is given by

$$\hat{\theta}_n{}^* = \sum_{i=1}^{B} \hat{\theta}^*_i/B.$$

Habibi (2011) studied the VaR estimates under the GARCH modeling using the bootstrap method. Let r_t be the return of a specified portfolio. Then, for the normally distributed portfolio, the value at risk is given by

$$\text{VaR}_\alpha = -\mu + z_\alpha \sigma$$

where μ and σ are the mean and variance of portfolio return. However, the quantile z_α may be obtained by the bootstrap method. Another method is Jacknife. The following Table gives the maximum and median of errors of VaR between bootstrap and Jacknife methods.

Table 1: The max and med of errors

α	0.1	0.05	0.025	0.008	0.0025	0.0005	0.00025
max	1.4	1.80	1.250	1.190	1.1200	1.0200	1.02000
med	1.1	1.50	1.000	1.020	1.0300	0.9900	0.98000

Source: Researcher results

Particle filter and EM algorithm

The Plain Particle Filter Framework (PPF) algorithm of Zhou et al. (2008) provides a good description for particle filter. The PPF algorithm in this problem is given as follows:

PPF algorithm

1) Initialization. Sample $\{x_0^i\}_{i=1}^N$ i.i.d from initial distribution P_0. Set t = 1.

2) Importance sampling. Sample x_t^i from $p(x_t|x_{t-1}^i), i = 1,2,\dots,N, k = 0,1.$

3) Bayes updating. Let $\hat{p}_t(x_t) = \sum_{i=1}^N w_t^i \delta(x_t - x_t^i)$, where δ the Dirac delta function is. Weights are calculated as $w_t^i \propto p(y_t|x_t^i), i = 1,2,\dots,N,$ and normalized.

4) Re-sampling. Sample $\{x_t^i\}_{i=1}^N$ i.i.d from $\hat{p}_t(x_t)$ and go to step 2.

Idvall and Jonsson (2008) applied the EM algorithm to algorithmic trading problem. Under the CAPM model $E(r_t) = r_f + \beta(E(r_m - r_f))$. Then,

$$\begin{cases} r_t = x_t + \varepsilon_t \\ x_t = r_f + \beta(E(r_m - r_f)) \end{cases}$$

This defines a state space model. Applying the particle filter the following Table for the Value at Risk of a stock is obtained.

Table 2: VaR under different scenarios

Scenarios	S_1	S_2	S_3
Expected loss	14503	14668	17655
VaR(0.95)	43736	43775	52554
VaR(0.975)	55775	55988	96888
VaR(0.99)	51154	52257	122442
VaR(0.995)	74776	75136	168332

Source: Researcher results

KDE method

The kernel density estimation (KDE) proposes a density based on a function called kernel and some parameters called bandwidth. There are many kernels like cosine, Gaussian, Epanechnikov and Triangular kernels. The general formulae for kernel density estimation are $\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^n K(\frac{x-x_i}{h})$. Grith et al. (2010) derived the risk neutral density using the kernel density estimation. neutral measure the discounted expectation of a financial derivative is its price. Here, using the historical data of the monthly call option of Intel Corporation during 2000-2015 and using the following kernels this density is estimated as follows. The following Table gives the maximum (max) and median (med) errors and asymptotic mean integrated squared error of bandwidth h. Also, related weights are given in the next Table.

Table 3: Max errors, med errors, bandwidth h

kernel	formula	max	med	h
Gaussian	$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	0.020	0.010	0.01
Cosine	$\frac{\pi}{4}\cos(\frac{\pi}{2}u)I(\lvert u\rvert \le 1)$	0.004	0.002	0.02
Epanechnikov	$\frac{3}{4}(1-u^2)I(\lvert u\rvert \le 1)$	0.030	0.010	0.04
Triangular	$(1-\lvert u\rvert)\, I(\lvert u\rvert \le 1)$	0.010	0.005	0.02

Source: Researcher results

Table 4: Weights of variables

Ratios	Contains	Input/Output	Weight
Liquidity Ratio	Current Ratio	in	0.45
	Quick Ratio	in	0.30
	Working Capital Ratio	in	0.25
Activity Ratio	Accounts Receivable Turnover Ratio	in	0.30
	Inventory Turnover Ratio	in	0.30
	Asset Turnover Ratio	in	0.20
	Collection Period Ratio	in	0.20
Leverage Ratio	Debt Ratio	in	0.30
	Interest coverage Ratio	in	0.35
	Equity Ratio	in	0.35
Profitability Ratio	ROA	out	0.26
	ROE	out	0.27
	Return On Current Assets	out	0.25
	Return On Equity	out	0.10
	Operating Profit to Sales	out	0.12

Source: Researcher results

Meta-heuristic methods

Simulated annealing

Consider the case of part 3.1. For a given function defined in a large discrete support domain, there is a meta-heuristicoptimization approach called simulated annealing to approximate the global optimum. There are some other search methods like brute-force search or gradient descent, however, simulated annealing is preferred when local optimum is more important than the global optimum. It is based on accepting worse solutions like other meta-heuristics methods. The logic behind it is similar to the process of heating and cooling a material. This process is done to alter the physical properties of material because of changes in internal structure. The cooled metal gets its new structure and receives its new properties. Lai (2010) applied the simulated annealing method in multifactor equity portfolio management. Ibanez et al. (2016) applied the simulated annealing method to study the stability of protein interaction networks in cancer and neurological disorders. The technique of Lai (2010), here, is used in the portfolio of case 3.1. It is seen that the weights are (0.1, 0.2, 0.15, 0, 0.3, 0.1, 0.05).

Operation research technique, empirical results

Topsis

Consider the case 3.1. The TOPSIS is a multi-criteria decision-making like AHP, proposed by Hwang and Yoon (1981). It contains seven steps including an evaluation matrix consisting of alternatives and b criteria, normalizing matrix, calculating the weighted normalized decision matrix, determining the worst alternative and the best alternative, calculating the related distances, calculate the similarity to the worst condition, and finally ranking the alternatives. In this problem, the category of TOPSIS is similar to KNN method, i.e., (ABT, AAP), (MMM, ATVI), (ABBV, ACN, AES).

Dea method

Data Envelopment Analysis (DEA) is a widely used management tool. This method first was introduced in the seminal work by Charnes, Cooper, and Rhodes. It is usually used for evaluating efficiency in business, bank-

ing and management. The DEA compares each decision maker unit (DMU) with the best DMU. For each DMU a set of varying level inputs and outputs exist. For example, each bank has a certain number of managers, size, and stocks (the inputs). Also, banks have a number of outputs like checks, loans, and so on (the outputs). Based on inputs and outputs, DEA tries to determine which of the banks are most efficient. Feroz et al. (2003) proposed a DEA solution for financial statement analysis. Aparicio et al. (2013) did benchmarking in data envelopment analysis based on genetic algorithms. Here, this method is applied to find the weight for financial statements of corporations applied in part 3.1, where related weights are (0, 0.35, 0.22, 0.05, 0.05, 0.22, 0.2)

Dual analysis

Dual analysis is a common approach to the optimization problem. In the linear programming, the solution of dual and primal problems are equal, however, it is not necessary that this event happens in the general optimization setting. In economic perspective, primal and dual problems may be viewed as resource allocation and resource valuation. A novelty application of equivalence of primal and dual solutions is the equivalence of a two-person zero sum game and linear programming. Usually, a difficult primal problem changes to a simple dual problem and vice versa. The weak duality theorem implies that if (x_1, …, x_n) is a feasible solution for the primal minimization linear program and (y_1, …, y_m) is a feasible solution for a dual maximization linear program, then $\sum_{i=1}^{m} b_i y_i \leq \sum_{j=1}^{n} c_j x_j$ where b_i and c_j are the coefficients of the respective objective functions (see Cai et al., 2014). Let p_t^i be the price of i-th i = 1,2, …, n financial asset at time t. The portfolio w_1, …, w_n is an arbitrage portfolio at maturity T if $\sum_{i=1}^{n} w_i p_i^0 = 0$ and with probability one. This problem defines a linear programming as follows:

$$\min Z = \sum_{i=1}^{n} w_i p_i^0,$$

such that $\sum_{i=1}^{n} w_i p_i^T \geq 0$ and w_i's are real numbers and $\sum_{i=1}^{n} w_i = 1$. For n = 2 the dual program is given by

$$\min Z^* = y_1$$
$$y_1 + p_i^T y_2 = p_i^0.$$

Solving this linear programming it is seen that

$$w_1 = \frac{p_2^T}{p_2^T - p_1^T}.$$

Applying this method to the case of 3.1, weights (0, 0.35, 0.2, 0.15, 0.05, 0.22, 0.2) are derived.

Operation research technique, theoretical results

Sensitivity analysis

Besides the optimized solution in a linear programming, the sensitivity analysis is also important. That is what event happens for an optimal solution when data values are changed. Standard approaches in linear programming presents many good answers to this question. For example, we refer to what occurs to the optimal tableau (see Winston, 2010). A natural question arises that under what changes in p_2^T, p_1^T the arbitrage opportunity is removed. This needs the sensitivity analysis with real data sets.

Kuhn-Tucker equation

For a general problem $\min_{x \in R^n} f(x)$ subject to $h_i(x) \leq 0, i = 1,2, \ldots, m$ and $l_j(x) = 0, j = 1,2, \ldots, r$, the Kuhn-Tucker conditions (see Hillier and Lieberman, 2012) are:

(i) $0 \in \partial f(x) + \sum_{i=1}^{m} u_i \partial h_i(x) + \sum_{j=1}^{r} v_j l_j(x)$

(ii) $u_i h_i(x) = 0$ for all i

(iii) $h_i(x) \leq 0, i = 1,2, \ldots, m, l_j(x) = 0, j = 1,2, \ldots, r$

(iv) $u_i \geq 0$ for all i.

The Kuhn-Tucker solution for this problem is:

$$u_1 \left(\sum_{i=1}^{n} w_i - 1 \right) = 0,$$

and

$$u_2 \sum_{i=1}^{n} w_i p_i^T = 0.$$

For n = 2 then $w_1 + w_2 = 1$ and $u_1 - p_i^T u_2 = p_i^0$, this yield the $w_1 = \frac{p_2^T}{p_2^T - p_1^T}$. An important result is the equilibrium condition for removing the arbitrage opportunity is $\frac{p_1^0}{p_i^T}$ be independent of i.

Quadratic programming

When the objective function is quadratic and constraints are linear, the program is quadratic program-

ming. The general form of quadratic programming is given by:

$$\text{Min} f(x) = \frac{1}{2} x'Bx - x'b,$$

subject to Ax ≤ b where B is a squared matrix and A is a general matrix, see Winston (2010). Next, suppose that three assets exist, then

$$\text{Min} Z = p_1^0 w_1 + p_2^0 w_2 + p_3^0 g(w_1, w_2)$$

$$p_1^T w_1 + p_2^T w_2 + p_3^T g(w_1, w_2) \geq 0$$

$$w_1 + w_2 + g(w_1, w_2) = 1.$$

This is a non-linear programming. For example, for $g(w_1, w_2) = w_1^2 + w_2^2 + a w_1 w_2$ this is a quadratic programming. The Lagrange multiplier equation is:

$$\Lambda = p_1^0 w_1 + p_2^0 w_2 + p_3^0 g(w_1, w_2) - \lambda_1 (p_1^T w_1 + p_2^T w_2 + p_3^T g(w_1, w_2)) - \lambda_2 (w_1 + w_2 + g(w_1, w_2) - 1).$$

The Lagrange multiplier equations are:

$$p_1^0 - \lambda_1 \left(p_1^T + p_3^T \frac{\partial g}{\partial w_1} \right) - \lambda_2 \left(1 + \frac{\partial g}{\partial w_1} \right) = 0$$

$$p_2^0 - \lambda_1 \left(p_2^T + p_3^T \frac{\partial g}{\partial w_2} \right) - \lambda_2 \left(1 + \frac{\partial g}{\partial w_2} \right) = 0$$

Integer programming

Decision variables of linear programming are continuous. For example, a producer can easily produce 100.34 gallons. However, sometimes this is not a realistic assumption. In this case, the decision variables are integer and the problem is referred as the integer programming. The Bayesian perspective to this problem is to choose a binary random variable J = 0,1 such that

$$\text{Min} Z = p_1^0 J + p_2^0 (1 - J)$$
$$p_1^T J + p_2^T (1 - J) \geq 0.$$

This is an integer programming.

Numerical analysis methods

Numerical computation

The Black method is an appropriate method for approximation of the price of financial derivatives. The following Table gives these prices. Here, k = 500, T = 1, r = 0.07, σ = 0.2. The option has dividend yield. Here, it is assumed that the dividend yield is one USD for each month. To approximate the Black method sometimes techniques like Spline are applied. The following Table gives the results. The Black method is not applicable for American type options.

Table 5: Option pricing: Black approximation

Option type	10	40	80	120	250	Black value
European Call	45.34	45.21	45.07	44.97	45.98	44.98
American Call	46.31	46.07	45.93	45.83	45.83	*
European put	31.54	31.41	31.27	31.18	31.18	31.18
American put	31.99	31.89	31.76	31.88	31.88	*

Sources: Researcher results

Actual case

Here, four methods including CART, K-means, TOPSIS, and simulated annealing are compared in the case of a customer churn example with real data set. However, because of security arguments, its name should be hidden. Consider a simulated bank where about five percent of its customers are churn. Here, churn customers are persons that have not contributed with the bank during the current six months. It is interesting to estimate this proportion using sample surveys. A real

data set is used including variables age, region, education, gender, cheque situation, occupation, account-balance, and turnover. About discrete variables, Region has four levels indexed by 1,…,4, education has four levels Diploma or lower is 1, Bachelor is indexed by 2, Master by 3 and PhD or upper is 4. Cheque situation is zero is one if person has a dishonored cheque and zero otherwise. Occupation four levels and Churn has two levels: zero for loyal and 1 otherwise. Other variables are continuous variables. Using the Cochrane formula 384 is anticipated. They are categorized to two classes: churn and loyal. About nine percent of the sample is churn that is 35 persons. As follows, results of each method are given:

1) the K-means clusters are given,

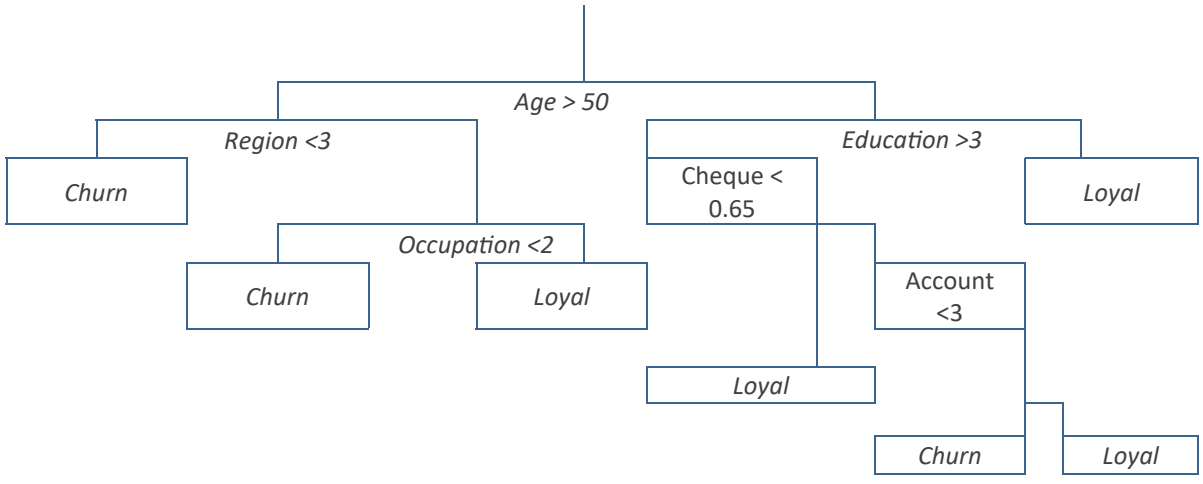
Table 6: K-means cluster of variables

Attribute	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Age	51.54	46.08	26.59	27.16
Region	1.27	1.96	3.19	4.19
Education	2.71	2.94	3.71	3.98
Cheque Situ.	0.12	0.36	0.82	0.94
Occupation	3.25	2.94	0.62	1.18
Account-Bal.	0.98	1.25	4.57	3.24
Turnover	0.34	0.42	1.25	1.49
Churn	1.00	1.00	0.00	0.00

Source: Researcher results

2) the tree of CART method is given as follows: It is seen that variables turnover and occupation have no effect on the tree structure,

Figure 2: CART algorithm for customer churn



Source: Researcher results

3) the simulated annealing (SA) gives weights (0.22, 0.05, 0.27, 0.11, 0.14, 0.18, 0.03) for each variable to construct discriminate function,

4) the TOPSIS provides gives weights (0.25, 0.12, 0.1, 0.07, 0.12, 0.29, 0.05) for each variable to construct discriminate function.

The following Table gives the probabilities of types I and II, for each method.

Table 7: Error I and II probabilities of each method

Method	K-means	CART	TOPSIS	SA
Error I prob.	0.04	0.07	0.01	0.03
Error II prob.	0.02	0.03	0.06	0.04

Source: Researcher results

According to the probabilities of error types I and II, it can be seen that, in this case, the best procedure is TOPSIS, the second best is simulated annealing, then K-means and finally CART.

Conclusions

Probability of error type I are sorted from smallest to largest for TOPSIS, SA, K-means and CART. Also, the smallest to largest probability of error type II belongs to the K-means, CART, SA and TOPSIS. Although almost all four methods seem reasonable applications in practice. These methods are chosen as selected methods of the mentioned categories. The criteria for choosing methods are types I and II errors probabilities. Although, these probabilities are sorted and methods are chosen, similar to the hypothesis testing approach, by keeping the type I error probability at the level of 0.05, then the best methods are K-means, SA and TOPSIS.

Acknowledgments

The author is grateful to the referee for several suggestions for improving the paper.

References

Aparicio, J., Lopez-Espin, J.J., Martinez-Moreno, R., Pastor, J.T. (2013). Benchmarking in Data Envelopment Analysis: an Approach Based on Genetic Algorithms and Parallel Programming. *Advances in Operations Research, 29,* 1-9.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone. C. J. (1984). Classification and Regression Trees. *The Wadsworth Statistics/Probability Series*, 358-370. Wiley.

Bulgurcu, B. (2012). Application of TOPSIS Technique for Financial Performance Evaluation of Technology Firms in Istanbul Stock Exchange Market. *Procedia-Social and Behavioral Sciences, 62,* 1033-1040.

Cai, X.Z., Wang, G.Q., El Ghami, M., Yue, Y.J. (2014). Complexity Analysis Of Primal-dual Interior-point Methods for Linear Optimization Based on a New Parametric Kernel Function with a Trigonometric Barrier Term. *Abstract and Applied Analysis, 11*, 23-38.

Das, N. (2003). Hedge Fund Classification Using K-Means Clustering Method. RTH International Conference on Computing in Economics and Finance. Seattle. USA: University of Washington.

Feroz, E.H., Kim, S., Raab, R.L. (2003). Financial Statement Analysis: a Data Envelopment Analysis Approach. *Journal of the Operational Research Society, 54,* 48-58.

Grith, M., Hardle, W.K., Schienle, M. (2010). Nonparametric Estimation of Risk-neutral Densities. *SFB 649 Discussion Paper 2010-021*. Berlin: Humboldt-Universität.

Habibi, R. (2011). A Simple Estimate of VaR under GARCH Modeling. *Ekonomia, 14,* 127-137.

Hillier, F.S., Lieberman, G.J. (2012). *Introduction to Operation Research*. USA: McGraw Hill.

Huck, N., Guegan, D. (2007). On the Use of Nearest Neighbors in Finance. *Technical report*. Department of Economics. Festion et Equipe University. France.

Hwang, C.L., Yoon, K. (1981). *Multiple Attribute Decision Making: Methods and Applications*. New York: Springer.

Ibanez, K., Guijarro, M., Pajares, G., Valencia, A. (2016). A Computational Approach Inspired by Simulated Annealing to Study the Stability of Protein Interaction Networks in Cancer and Neurological Disorders. *Data Mining and Knowledge Discovery*, *30,* 226-242.

Idvall, P., Jonsson, C. (2008). Algorithmic Trading: Hidden Markov Models on Foreign Exchange Data. Masterthesis. Department of Mathematics, Linkopings Universitet.

Kovalerchuk, B., Vityaev, E. (2000). Data Mining in Finance: Advances in Relational and Hybrid Methods. USA: Kluwer Academic Publishers.

Lai, C.C. (2010). Simulated annealing in multifactor equity portfolio management. *Proceedings of the International Multi-Conference of Engineers and Computer Scientists*, IMECS 2010, Hong Kong.

Olave, P., Miguel, J.A. (2000). Bootstrap Method in Exchange Rate Forecasting. *Technical Report*. Spain: University of Zaragoza.

Mingione, F. (2011). Forecasting with Principal Components Analysis: An Application to Financial Stability Indices for Jamaica. *Technical Report*. Bank of Jamaica. Jamaica.

Miyazaki, K. (2003). Dual Analysis on Hedging VaR of Bond Portfolio using Options. *Journal of Operation Research, 4,* 448-466.

Patil, R. (2014). Heart Disease Prediction System Using Naive Bayes and Jelinek-mercer Smoothing. *International Journal of Advanced Research in Computer and Communication Engineering 3,* 6786-6789.

Robles-Granda, P.D., Belik, I.V. (2010). A Comparison of Machine Learning Classifiers Applied to Financial Datasets. *Proceedings of the World Congress on Engineering and Computer Science*. WCECS October 20-22, San Francisco, USA.

Sutton, O. (2012). Introduction to k Nearest Neighbor Classification and Condensed Nearest Neighbor Data Reduction. *Technical report*. UK: University of Leicester.

Tracey, M. (2009). Principal Component Value at Risk: an Application to the Measurement of the Interest Rate Risk Exposure of Jamaican Banks to Government of Jamaica (GOJ) bonds. *Technical reports*. Financial Stability Department, Research & Economic Programming Division. Bank of Jamaica.

Winston, W.L. (2010). *Operation Research: Applications and algorithms*. USA: Cengage Learning India Pvt Ltd.

Zhou, E., Fu, M.C., Marcus, S.I. (2008). A Particle Filtering Framework for Randomized Optimization Algorithm. *Proceedings of the 2008 Winter Simulation Conference*. USA.