Heliyon 10 (2024) e23798

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Gaussian Bayesian network model of healthcare, food and energy sectors in the pandemic: Türkiye case

Ersin Sener [a,*], Ibrahim Demir [b]

[a] Department of Mathematics, Faculty of Art & Science, Kirklareli University, Kirklareli, 39000, Turkey
[b] Turkish Statistical Institute, Ankara, Turkey

ARTICLE INFO

Keywords:
Bayesian networks
Energy
Food and beverage
Healthcare
Stocks
Continuous variables

ABSTRACT

Healthcare, food, and energy are the basic needs of life in the globalizing world. Humankind's quest to maintain a healthy life and find ways to meet its needs has continued since its existence. The causal relations between the healthcare, food, and energy sectors are explored using data from the pandemic when COVID-19 was a global risk, and human health sustainability underwent a complicated process. It aims to examine the interaction between the healthcare, food, and energy sectors and model the causal relationship within the framework of probabilistic dependencies. For this purpose, the relationships between these sectors during the pandemic are modeled via Bayesian Networks (BNs). This highly successful inference method makes the complex structure of causal relationships graphically understandable. The data consists of stock returns at the end of the business day between March 11, 2020, when the pandemic was declared, and December 26, 2022. Data on 13 stocks actively traded on the Istanbul Stock Exchange (BIST) during the 700 days were obtained from tr.investing.com. Causal modeling uses Gaussian Bayesian Networks (GBNs) for continuous variables. To make the inferences drawn from the data more successful and minimize the loss of information, the GBN model is built with continuous variables. The posterior Probability Density Functions (PDFs) of the stocks in the network are constructed over the structure of the Directed Acyclic Graphs (DAG) of the BNs, and inferences are made by querying possible cases (tips). Markov Chain Monte Carlo (MCMC) simulations are performed with the posterior PDFs, and measures of the central tendency of the stocks are calculated. GBNs are used to generate daily return estimates for ULKER with the lowest MSE (1.06e-03) and RMSE (3.22e-02) values and ULUUN with the highest MSE (3.43e-03) and RMSE (5.83e-02) values.

1. Introduction

Bayesian Network (BN) is a Probabilistic Graph Model (PGM) that uses information about the probabilistic relationships of a set of random variables to represent relationships between variables with a Directed Acyclic Graph (DAG) [1]. BNs, which use conditional independence and DAG structure to explain the causal relationship between variables, have found wide application in recent years with the popularity of machine learning. BNs are an efficient and versatile tool for making inferences from variables. In addition, the power of BN models to express the relationships between variables and the efficiency of querying the relationships balance each other.

* Corresponding author.
E-mail address: ersinsener@klu.edu.tr (E. Sener).

https://doi.org/10.1016/j.heliyon.2023.e23798
Received 30 January 2023; Received in revised form 6 December 2023; Accepted 13 December 2023
Available online 16 December 2023

In a comprehensive literature review on algorithms for learning BN structure from data, alternative methods for obtaining the DAG model from the data with missing observations and continuous variables are mentioned [2]. BNs have many applications, from bioinformatics to law, from image and natural language processing.

1.1. Literature review

There are also applications of BNs in finance. Some crucial studies encountered in the literature review are discussed in this section. In predicting bankruptcy, unnecessary and less relevant variables are eliminated using correlations and partial correlations between variables, and a Naive BN model is developed with eight continuous variables [3]. Continuous variables are discretized, and the effect of the number of possible states on the model's performance is investigated. It has been observed that BN performs exceptionally well when continuous data is discretized into 2 or 3 classes, and the performance of the model decreases with the effect of overfitting when it is discretized into four or more classes. Instead of discretizing continuous variables, the effect of modeling with continuous distributions on the model's performance is investigated. It has been found that this process does not positively affect the model's performance [3].

To estimate the price-earnings (P/E) ratio for stocks, the continuous P/E ratio is discretized using the clustering algorithm, and the BN model is obtained from a set of discrete values. The average stock prices of NIKKEI (NIKKEI225) and Toyota Motor Company (TOYOTA) are considered [4]. The correlation coefficient and mean squared error (MSE) are used as comparison metrics, and it is found that BN gives better results than the traditional time series estimation methods (AR, MA, ARMA, and ARCH) [4]. Some studies analyze sectoral stock market indicators using classical methods. For example, the conditional value-at-risk (CoVaR) method [5] is used to determine the influential banks in evaluating the contribution of 52 banks and their subsidiaries operating in the Turkish financial sector to risk diversification and systematic risk between 2005 and 2016 [6].

The effects of the discretization (supervised, unsupervised, and manual) methods of continuous variables on the performance of BNs are investigated [7]. Three methods are used to discretize the soil erosion dataset consisting of 1600 samples, and classification is performed with BNs. It is determined that the average estimation abilities of the discretization methods in BN construction are supervised (73.8 %), manual (69.0 %), and unsupervised (64.8 %), respectively. The advantages and disadvantages of discrete and continuous variables are explained in detail [7].

The daily rates of increase and decrease of the stock market indexes (Financial Times Exchange 100 (FTSE100), Tokyo Stock Exchange (Nikkei225), and Dow Jones Industrial Average (DOW30)) are analyzed using BN via the K2 algorithm [8]. This algorithm is applied to predict the up/down analysis of daily stock market indexes in 2007, and the results are compared with traditional and widely used psychological line and trend forecasting algorithms. As a result of the accuracy comparisons, it was found that the current algorithm (60 %), psychological line (50–59 %), and trend forecasting (50–52 %) [8].

The BN model is proposed to improve the classical trend-following effect in the U.S. stock market index and to encourage investors to turn to markets with positive returns and take short positions in markets with negative returns. Results show that BNs incorporating market variables outperform traditional trend-following and trading strategies [9].

The relationship between Chinese macroeconomic data and the stock market is examined using the dynamic BN method, and it is found that the relationships between the variables depend on the period. Although indexes that can influence the stock market in some industries are found, the relationship between the stock market and macroeconomics needs to be considered in the BNs [10].

To systematically analyze the impact of COVID-19 on financial risk, a Bayesian topology based on gray clusters is constructed, considering the COVID-19 pandemic and economic indicators as nodes [11]. Liquidity, country, and equity market risks are considered in the financial risk assessment. The probability distribution of financial risk levels at different COVID-19 index levels is modeled using BN. The clustering results show that there is a positive relationship between pandemic severity and financial risks. In addition, liquidity and country risk are found to be more sensitive to the COVID-19 index, while equity market risk is not very sensitive to the COVID-19 index [11]. BNs are used for credit score classification on three real-world datasets and compared with Markov Chain Monte Carlo (MCMC) simulations. BNs have shown much better performance with the support of the Markov blanket [12].

In the literature review, it has been observed that BNs are a probabilistic modeling method that offers understandable and quite simple implications for modeling the causal relationship of both continuous and discrete variables. It is observed that expert opinion is used in constructing the network graph structure, especially in the studies carried out in the field of finance using BNs. The researchers construct the graph structure of the BN with expert opinion by directly restricting the implicit relationship between the variables.

However, multivariate statistical methods such as regression and path analysis, based on expert opinion, take place in the literature with vast application areas. These parametric methods have prerequisites to fulfill various assumptions before the calculation steps. Since continuous variables are mostly used in these analyses, there are various assumptions, and sometimes these assumptions cannot be realized. In cases where the assumptions cannot be satisfied, the analysis results are weak, and the interaction of the variables is not clear [13]. Although these methods seem to be on the same ground as BNs since they contain affecting and affected variables, they do not guarantee the independence of the variables conditionally.

Contrary to the shortcomings mentioned above, the BN approach has many advantages, such as the absence of assumptions that affect the result of the analysis, the ability to analyze binary, nominal, and ordinal variables that contain preliminary information in the analysis, and the ability to examine the interaction between the variables.

In our literature review of BNs with stock market indexes and stocks, the discrete values of the variables are discussed, and it is observed that modeling for continuous variables is not preferred. Although BNs produce highly successful models for discrete and continuous variables, information loss is inevitable due to the discretization of the data. However, using the natural structure of the data is very important when making predictions with Bayesian methods.

Thus, these advantages of BNs, a very user-friendly method that guarantees the conditional independence of the variables, contains the conditional (non-dependency) relations between the variables and provides an understandable graph structure, have been utilized. This opportunity offered by BNs is used in this study to investigate the relational structure of stocks, one of the financial indicators.

The COVID-19 pandemic has once again highlighted the vital role of the healthcare, food, and energy industries in meeting people's basic needs. The difficulties in maintaining the balance between supply and demand in this process and the war between Ukraine and Russia, which triggered the global food and energy crisis, are dragging today's economy into depression. During the pandemic, Türkiye played a significant role in the global healthcare, food, and energy sectors amidst the food and energy crises that followed the health crisis. This interaction in global financial markets will inevitably trigger the relationship between the indexes in Türkiye. The main thing is to observe how this relationship is reflected between stocks in this term. Between March 11, 2020, and December 26, 2022, the healthcare, food, and energy sectors have a logical and implicit relationship. In this study, 13 companies' stocks traded on the Istanbul Stock Exchange (BIST) are discussed with available data. Returns of stocks according to their closing prices at the end of 700 business days constitute the data set. BNs are used to prove that the logical implicit relationship of stocks is also provided as probabilistic.

BNs have many applications for discrete and continuous variables [14]. It was found that discrete variables were used in 52.6 % of the studies on BNs conducted between 1990 and 2010 [15]. The estimation performance of the BN depends on the method used to discretize the data [7]. Although the first studies with BNs are discussed and draw some inferences from continuous data [3,4], and obtained network models, discrete variables are alternatively modeled in the field of economics [16,17], and machine learning [10,11, 18] are frequently used. However, it is expected to reduce the loss of information from the data as much as possible, eliminate cumulative errors in the modeling phase, and obtain results as close to reality as possible. The continuous values of stocks are modeled with BN to observe the change more clearly and draw more realistic results.

BNs, one of the PGMs, are used to identify the relational structure of the stocks, infer probabilistic information, and model them mathematically. Expert opinion is used to logically infer the causal relationship from the network structure obtained in the context of probabilistic dependencies between BNs and stocks. In this study, using BNs for learning and inference, the DAG structure is learned from the dataset to make predictions and obtain a successful model. To the best of our knowledge, the stocks traded in the healthcare, food, and energy sectors are not modeled with BNs for continuous variables. In this respect, this study will be the first application of Gaussian Bayesian Networks (GBNs) to fill the gap in this area.

2. Materials and methods

In this section, the dataset used in the study and the operations performed on the dataset are mentioned. In addition, for GBNs, which allow graphical inference using probabilistic relations for continuous variables, the PC (Peter and Clark) algorithm is used to obtain a DAG structure from continuous variables, and the steps of the PC algorithm are given.

2.1. Stocks dataset

There are four companies in the healthcare sector, 31 companies in the food sector, and 23 companies in the energy sector, which are a total of 58 companies traded on BIST in January 2023. The stocks that continued their activities in these sectors and ranked in the top 10 on a sector basis in order of transaction volume are evaluated from March 11, 2020, when the pandemic period began, to December 26, 2022. Under these conditions, a total of 13 stocks are analyzed: two from the healthcare sector, seven from the food sector, and four from the energy sector. The return rates of the stocks according to the closing prices at the end of 700 working days are compiled as a dataset from the website tr.investing.com. The daily stock returns are percentages and continuous variables.

The codes and explanations for the stocks are given in Appendix 1. The selected stocks have been traded on BIST since the beginning of the pandemic and have the most securities and the largest transaction volume. The descriptive statistics of the stocks are presented in Table 1.

The daily percentage returns of the stocks have an average that is very close to zero and standard deviations that vary between 2

Table 1
Descriptive statistics of stocks.

Sector	Stocks	Mean	St. Dev.	Skew.	Kurtosis	Min.	Max.
FOOD & BEVERAGE	ULKER	0.0015	0.0224	0.0128	2.9806	−0.0789	0.0842
	ULUUN	0.0050	0.0407	0.0154	2.9745	−0.1678	0.1554
	TUKAS	0.0025	0.0313	−0.0258	2.9810	−0.1117	0.1242
	AEFES	0.0025	0.0242	−0.0034	3.0400	−0.0945	0.1077
	CCOLA	0.0028	0.0232	0.0143	3.0167	−0.0761	0.0969
	KERVT	0.0030	0.0316	−0.0279	3.0607	−0.1445	0.1192
	PENGD	0.0032	0.0388	−0.0259	2.9733	−0.1509	0.1825
HEALTHCARE	LKMNH	0.0038	0.0349	0.0188	3.0713	−0.1370	0.1374
	MPARK	0.0031	0.0286	0.0089	3.0231	−0.1081	0.1193
ENERGY	ENJSA	0.0026	0.0229	−0.0093	2.9245	−0.0737	0.0931
	ODAS	0.0039	0.0352	−0.0151	2.9279	−0.1074	0.1240
	AKSEN	0.0055	0.0299	0.0145	3.0088	−0.1060	0.1353
	ZOREN	0.0033	0.0313	0.0012	2.9632	−0.1154	0.1226

and 4%. The highest daily loss rate among the stocks is ULUUN, with 16.78 %, while the highest gain rate is PENGD, with 18.25 %.

In graphing and modeling with BNs, when the variables have priors, these probability values are used. In the absence of a priori information (distribution parameters of the data, e.g., mean, standard deviation, etc.), the distribution of the data is expressed as a function of local distributions according to Bayes' rule, and this function has a priori structure [16,17]. It has been observed in research with continuous variables that the data generally fits normal or Gaussian distributions [10,18]. The data set is expected to satisfy the assumption of a normal distribution [5,8]. Therefore, the kurtosis and skewness values of the data are examined to observe the normal distribution of the dataset and the values are presented in Table 1.

Skewness, also defined as a lack of symmetry, is a measure of symmetry. The data distribution is symmetrical if it is equal on both sides of the mean. Kurtosis is the degree of flattening near the center of the density function curve [19]. In other words, datasets with high kurtosis tend to have strong outliers. Datasets with low kurtosis tend to have slight outliers or no outliers. The value for skewness is expected to be close to zero, while the value for kurtosis should be close to 3 [20]. According to Table 1, the skewness of the stocks is calculated around 0, and the kurtosis is close to 3. The histograms in Fig. 1 show that the data are approximately normally distributed.

Both Table 1 and Fig. 1 show that the daily stock price changes have a normal distribution. BNs provide sufficient information about the structure of the variables before modeling. It is necessary to ensure the assumption of low correlations between the variables in GBN modeling of continuous variables [21]. The correlations between stocks are calculated to observe this assumption and presented as a heat map in Fig. 2.

The correlation between the stocks shows a relationship between KERVT and PENGD with a correlation value of 0.55, and this relationship is at a reasonable level. The relationships between the other stocks are medium and low. Thus, the assumption of a low correlation of GBN is fulfilled, and the data is normally distributed.

2.2. Bayesian Networks

BN is a network model that combines the structure of a DAG and information about conditional probabilities to explain the causal relationship between PGMs. BNs have gained acceptance with a mixed representation of probability theory by incorporating graph theory and probability information with a graphical representation of their structure [22].

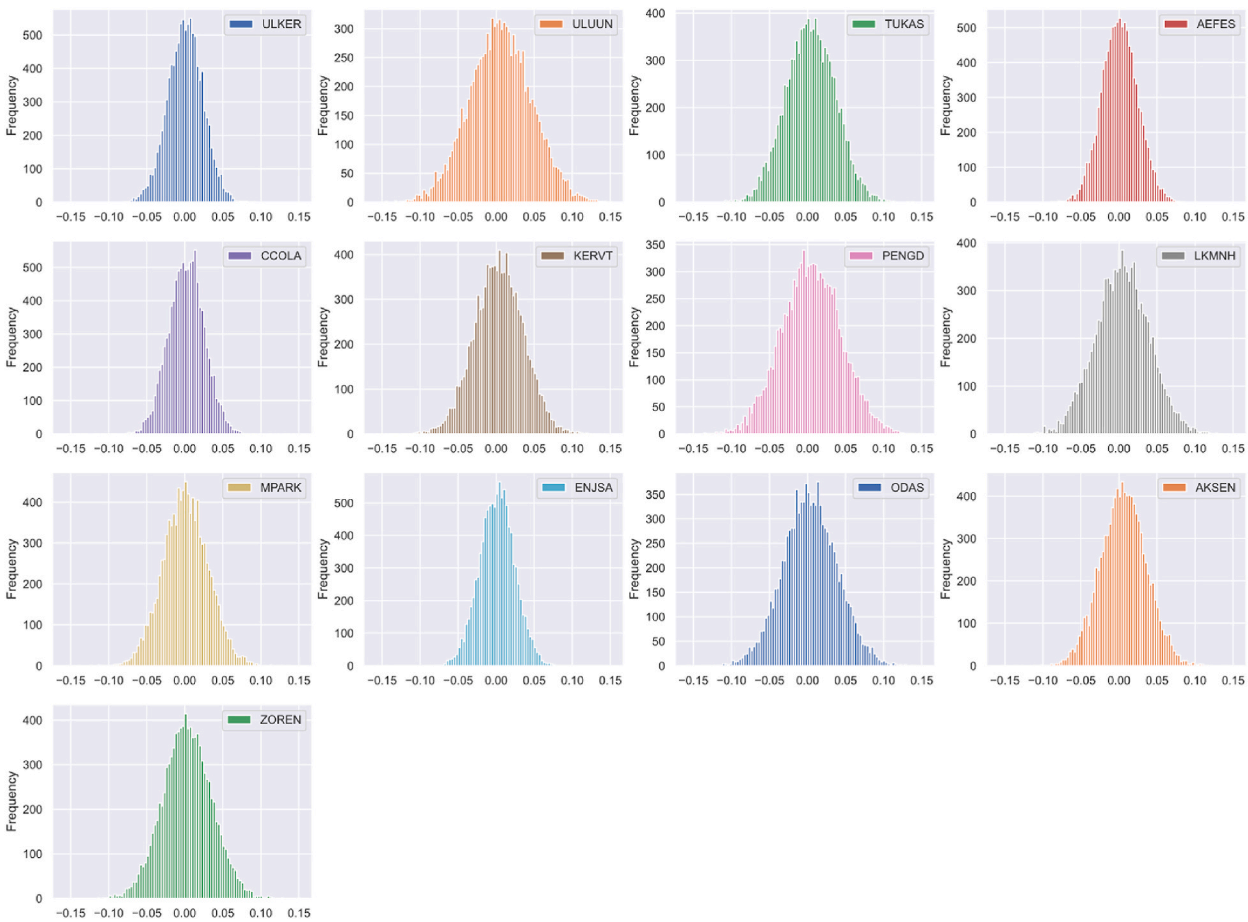

Fig. 1. Histogram of stocks.

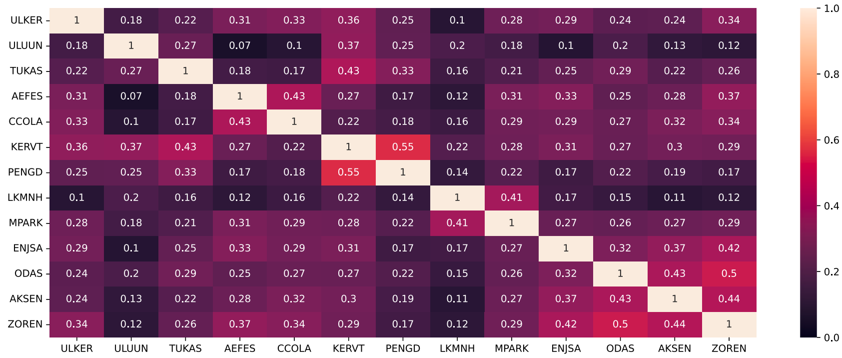

Fig. 2. Correlations of stocks and heat map.

BN is a beneficial graphical network model that contains information about conditional probabilities with parameter information and the DAG structure for causal inference. A BN represents the joint distribution of the set $X_1, X_2, \ldots, X_n$ with n (discrete) variables, with DAG and conditional probability tables (CPT) [21]. Each node in the network is associated with the CPT of each state of its parents. The nodes that are affected are defined as child nodes, and those that are affecting are defined as parent nodes. Fig. 3 shows a simple BN example. In Fig. 3, leaf nodes D, E, and C are green, parent nodes B and F are yellow, and ancestor node A is red.

BNs contain two main pieces of information: first, DAG to determine the relationship between variables, and second, mathematical equations containing conditional probability values to draw inferences about the model's parameters. Considering the graph G shown in Fig. 3, by the definition of D-separation [21], each X node is conditionally independent of all nodes in G that have no descendants, is given $Pa(X)$. This is known as the Markov assumption. The graph of a BN can be determined by the parents of each variable.

For the example in Fig. 3, the structure of the BN can be given as: $Pa(A) = \varnothing$, $Pa(B) = A$, $Pa(F) = \varnothing$, $Pa(C) = B$, $Pa(E) = B, F$ ve $Pa(D) = B, F$. The joint probability distribution of a BN can be represented as in Eq. (1) as the product of the conditional probabilities of each variable whose parents are given [22].

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|Pa(X_i)) \tag{1}$$

The joint probability for the graph G shown in Fig. 3 is calculated as $P(A,B,C,D,E,F) = P(A)P(F)P(B|A)P(C|B)P(E|B,F)P(D|B,F)$.

2.2.1. Gaussian Bayesian network

As for the general use of BNs, the variables take on discrete values, or they are discretized, with some methods used to discretize continuous variables [7,23,24]. However, discretization of data leads to increased computational steps and loss of information. Information loss can be prevented by working with continuous variables. Although there are studies on families of distributions to draw probabilistic inferences from continuous variables, the Gaussian distribution is often used [22]. These are PGMs for continuous variables, also known as GBNs.

GBN is BN whose variables are continuous, and all Probability Density Functions (PDFs) are Gaussian distributed with constant variance. That is, let $X_1, \ldots, X_n$ be continuous parents of the continuous variables X_{n+1} the coefficients $\beta_0, \beta_1, \ldots, \beta_n$ and the Conditional Probability Distribution (CPD) for the parameters σ^2 as in Eq. (2) [25].

$$f_{X_{n+1}|X_1,\ldots,X_n}(X_{n+1}|X_1, \ldots, X_n) = N(\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n; \sigma^2) \tag{2}$$

Basically, GBNs [21];

1. Are networks with a multiple tree structure.
2. The variables have low correlations and Gaussian distribution (normal).
3. The variables and their parents have a linear relationship as in Eq. (3):

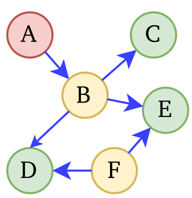

Fig. 3. A simple Bayesian Network.

$$X = \beta_1 U_1 + \beta_2 U_2 + \ldots + \beta_n U_n + W_X \tag{3}$$

where, U_i's are the parent of the variable X, β_i's are constant coefficients, and W_X is zero-mean Gaussian noise [21].

The inference procedure for continuous variables is similar to the belief-set propagation of BNs for discrete variables. Still, the dispersion of probabilities means the dispersion of means and standard deviations. The most basic assumption is that the marginal distributions of all variables have a Gaussian distribution. The posterior probability of the set of variables in GBN can be described in Eq. (4) [21].

$$P(X|E) = N(\mu_X, \sigma_X) \tag{4}$$

μ_X and σ_X; E represents the mean or standard deviation of X when the probability of detection is known. When the propagation algorithm calculates the mean and standard deviation, it starts with the leaf node and returns to the parent and root nodes. This calculation is performed using Eq. (5).

$$\mu_i^- = \left({}^{1}/_{\beta_i}\right)\left[\mu_\lambda - \sum_{k \neq i} \beta_k \mu_k^+\right], \sigma_i^- = \left({}^{1}/_{\beta_i^2}\right)\left[\mu_\lambda - \sum_{k \neq i} \beta_k^2 \sigma_k^+\right] \tag{5}$$

Each node sends information to its child node j, and propagation continues to the leaves. The mean and standard deviation of the leaf nodes are calculated by Eq. (6) using the propagation parameters of the parent nodes.

$$\mu_j^+ = \frac{\sum_{k \neq j} \mu_k^- \Big/ \sigma_k + \mu_\pi \Big/ \sigma_\pi}{\sum_{k \neq j} 1 \Big/ \sigma_k^- + \mu_\pi \Big/ \sigma_\pi}, \ \sigma_j^+ = \left[\sum_{k \neq j} 1 \Big/ \sigma_k^- + 1 \Big/ \sigma_\pi\right]^{-1} \tag{6}$$

Each node integrates the information it receives from its children and parents with Eq. (7) and (8) [21].

$$\mu_\pi = \sum_i \beta_i \mu_i^+, \sigma_\pi = \sum_i \beta_i^2 \sigma_i^+ \tag{7}$$

$$\mu_\lambda = \sigma_\lambda \sum_j \mu_j^- \Big/ \sigma_j^-, \sigma_\lambda = \left[\sum_i 1 \Big/ \sigma_i^-\right]^{-1} \tag{8}$$

finally, for each node, the information's from the parent and child nodes are combined to obtain the mean and standard deviation in Eq. (9).

$$\mu_X = \frac{\sigma_\pi \mu_\lambda + \sigma_\lambda \mu_\pi}{\sigma_\pi + \sigma_\lambda}, \sigma_X = \frac{\sigma_\pi \sigma_\lambda}{\sigma_\pi + \sigma_\lambda} \tag{9}$$

The posterior probabilities obtained with GBNs have a Gaussian distribution (normal) [21]. The very successful PC algorithm is used to construct the DAG structure of GBN from continuous variables [26].

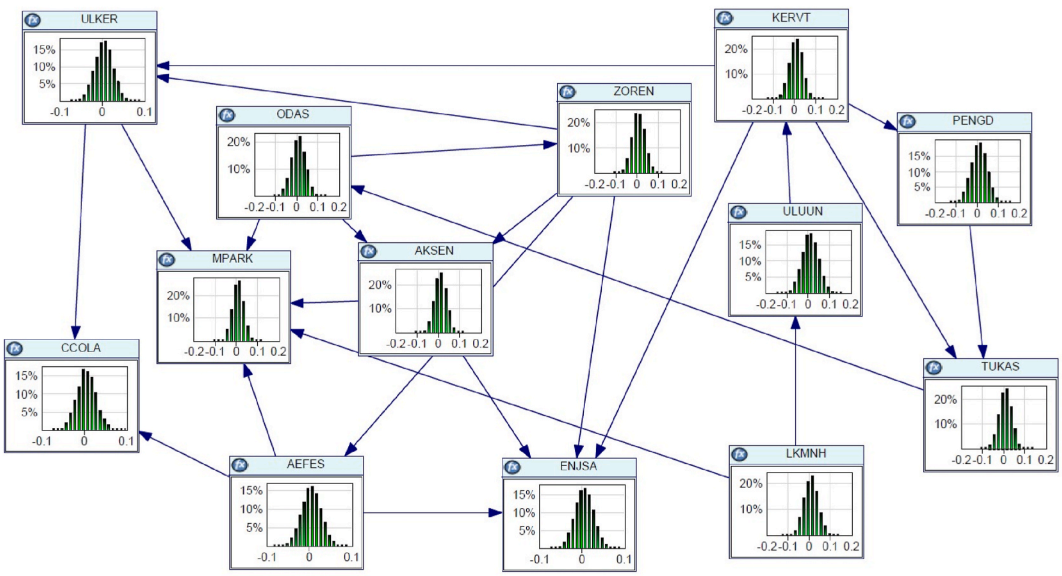

Fig. 4. GBN model of Stocks.

2.3. PC (Peter-Clark) algorithms

Firstly, the undirected graph of the BN, which is called the skeleton, is created, and then the direction of the edges is determined via the PC algorithm [26]. The procedure for determining the skeleton, starting from a fully connected undirected graph, further guarantees the conditional independence of each pair of nodes for a known subgraph of nodes. To do this, it is assumed that a procedure $I(X, Y|S)$ can determine whether X and Y are independent, given a subset of the variables S. When this measure is below a threshold determined by a certain confidence level, the margin between the pair of variables disappears. These tests are repeated for all pairs of variables in the graph [21].

In the second step, the alignment of the edges is determined using the conditional independence tests between variable triples. This involves searching for the substructures of the graph in the form $X - Y - Z$ such that there are no edges between $X - Y$. If X and Y are not independent, given Z, a V-structure is created by aligning the edges as $X\rightarrow Z\leftarrow Y$ [21]. When all V-structures are found, an attempt is made to orient other edges based on independence tests and avoid loops.

3. Results

GBN modeling is performed using the PC algorithm over a dataset of daily returns for 700 business days, covering the period from March 11, 2020, to December 26, 2022. A necessary and sufficient condition for obtaining the GBN structure is that the data set has a normal distribution [11]. All stages of this condition are discussed in Section 2. The GBN model of stocks traded in the healthcare, food, and energy sectors is constructed using the PC algorithm, as shown in Fig. 4.

Since the GBN is constructed using continuous variables, the distributions of the PDFs of each node are given in Fig. 4 instead of the evidence probabilities of the nodes. The PC algorithm is constraint-based, and the constraints of the algorithm are *maximum number of neighbors* = 8 and the significance level is set at $\alpha = 0.05$. The GBN model is constructed using the academic version of GeNIe 4.0 [27], a general development program for graphical decision analytic models, within the constraints set by the PC algorithm. The information about the parents and children of the GBN model of the stocks shown in Fig. 4 is listed in Table 2 for a better understanding of the model.

Table 2 shows that the stock LKMNH has no parent node; in other words, the stock LKMNH is a root node. Since no other nodes are without parents, the GBN model shown in Fig. 4 consists of a single root and is a rooted tree. CCOLA, MPARK, ENJSA, and ODAS have no child nodes; these stocks are referred to as leaf nodes in the child column of Table 2.

MPARK and ENJSA are leaf nodes with 5 and 4 parents, respectively. The number of parents of a node indicates the inner degree of that node, while the number of child nodes indicates the outer degree of that node [21]. MPARK and ENJSA, the nodes with the highest internal degree, are the most affected. Nodes with an internal degree of 1 are ULUUN, AEFES, KERVT, PENGD, ODAS, and ZOREN. The external degree of nodes ULUUN, TUKAS, and PENGD is 1.

The BN modeling representation includes the $G = (V, E)$ graph structure and CPT values based on DAG. The main advantage of BN models is the exploration and discovery of dependency structures with conditional independence tests. This is especially true for financial applications where low correlations between nodes are expected. In addition, in GBN models, linear posterior PDFs are computed in the DAG generated by conditional independences. In BN models, CPT values are calculated in the DAG generated by conditional independence, with variables having discrete values. The linear posterior PDFs shown in Table 3 are obtained from the GBN model graph structure of the stocks shown in Fig. 4.

The posterior PDFs in Table 3 represent the propagation of price changes in stocks over time in the period under investigation. For example, since the LKMNH stock has no parent at t = 121, the calculation starts by generating a random value from the posterior PDFs in Table 3. Then, the end-of-day closing values of the stocks at t = 121 are estimated using the posterior PDFs based on the parent-child relationship in Table 3.

Based on a 700-day research period, our research can be estimated for a possible business day (e.g., 120th day, September 4, 2020) by simultaneously running the posterior PDFs presented in Table 3 with the obtained GBN models. One of the most important considerations is to run the posterior PDFs together. The random values to be obtained from the normal distribution, depending on the

Table 2
Parents and children of stocks.

Sector	Stocks	Parent	Child
Food & Beverage	**ULKER**	KERVT, ZOREN	CCOLA, MPARK
	ULUUN	LKMNH	KERVT
	TUKAS	KERVT, PENGD	ODAS
	AEFES	ZOREN	CCOLA, ENJSA, MPARK
	CCOLA	ULKER, AEFES	
	KERVT	ULUUN	ENJSA, PENGD, TUKAS, ULKER
	PENGD	KERVT	TUKAS
Healthcare	**LKMNH**		MPARK, ULUUN
	MPARK	ULKER, AEFES, LKMNH, ODAS, AKSEN	
Energy	**ENJSA**	AEFES, KERVT, AKSEN, ZOREN	
	ODAS	TUKAS	AKSEN, MPARK, ZOREN
	AKSEN	ODAS, ZOREN	
	ZOREN	ODAS	AEFES, AKSEN, ENJSA, ULKER

Table 3
The posterior probability density estimation functions of stocks.

Sector	Stocks	Functions
Food & Beverage	**ULKER**	$ULKER = 0.206911 * KERVT + 0.191251 * ZOREN + Normal(0.000195145, 0.0204869)$
	ULUUN	$ULUUN = 0.235237 * LKMNH + Normal(0.00449469, 0.0397044)$
	TUKAS	$TUKAS = 0.355931 * KERVT + 0.110216 * PENGD + Normal(0.00087602, 0.0282039)$
	AEFES	$AEFES = 0.28504 * ZOREN + Normal(0.00135835, 0.0224898)$
	CCOLA	$CCOLA = 0.225405 * ULKER + 0.351159 * AEFES + Normal(0.00152911, 0.0206697)$
	KERVT	$KERVT = 0.284235 * ULUUN + Normal(0.0015836, 0.0293089)$
	PENGD	$PENGD = 0.680582 * KERVT + Normal(0.00112535, 0.0323557)$
Healthcare	**LKMNH**	$LKMNH = Normal(0.00361286, 0.0345702)$
	MPARK	$MPARK = 0.192566 * ULKER + 0.205706 * AEFES + 0.29696 * LKMNH + 0.0683476 * ODAS + 0.117322 * AKSEN + Normal(0.0001279, 0.02492)$
Energy	**ENJSA**	$ENJSA = 0.148784 * AEFES + 0.113066 * KERVT + 0.134633 * AKSEN + 0.184162 * ZOREN + Normal(0.000741325, 0.0201912)$
	ODAS	$ODAS = 0.324198 * TUKAS + Normal(0.00295516, 0.0335774)$
	AKSEN	$AKSEN = 0.232151 * ODAS + 0.296668 * ZOREN + Normal(0.00340639, 0.0261252)$
	ZOREN	$ZOREN = 0.450508 * ODAS + Normal(0.00148704, 0.0271039)$

mean and standard deviation of the stocks, will affect the prediction performance of the BN. The error needs to be minimized using MCMC.

Using the linear posterior PDFs listed in the “Functions” column of Table 3, the daily stock price changes are estimated using the MCMC simulation with 1000 repetitions. To evaluate the estimation performance, the mean, standard deviation, mean squared error (MSE), and root mean square error (RMSE) of the simulation are calculated and presented in Table 4.

Comparing the simulation results in Table 4 with the descriptive statistics in Table 1, the results are very similar. One can measure the proximity of these results by looking at the MSE and RMSE values in Table 4. While the lowest MSE (1.06e-03) and RMSE values (3.22e-02) are provided by the estimate of the change in ULKER stocks, ULUUN has the highest MSE (3.43e-03) and RMSE values (5.83e-02) observed in the estimation of stocks. The results of the other stocks are lower than the error rates in the estimation of the ULKER and ULUUN stocks. The results drawn using the estimators obtained with GBN are quite successful because of the MCMC simulation.

Another advantage of GBN is scenario applications based on queries. Queries of continuous variables are scenario applications interpreted as increase or decrease. Scenarios are created for ULKER stock with the best estimation ability and the least error in PDFs. The conditions for increase and decrease are queried as 1 and -1, respectively, and the possible scenarios are shown in Figs. 5 and 6.

When the decreases and increases of the ULKER stock are known, the other 12 stocks also show a decreasing and increasing trend simultaneously. It can be concluded that the stocks analyzed in BIST show similar trends and that the sectors acted together during the pandemic.

4. Conclusion

The mathematical relational structure of 13 different stocks traded in the healthcare, food, and energy sectors on BIST since the beginning of the pandemic has been analyzed within the framework of probabilistic dependencies via BNs. Continuous values of stocks consisting of percentage return rates at the end of the business day are modeled with GBNs, and the posterior PDFs appropriate for the graph structure of the network are obtained. The MCMC simulation is performed using posterior PDFs, and the prediction performances are observed. It is observed how other stocks behave with the changes in the ULKER stock for possible increase or decrease conditions.

As a result of making the data discrete, which is often done in modeling continuous variables with BNs, the loss of information obtained from the data is prevented. On the contrary, the biggest difficulty encountered in obtaining a GBN model with continuous variables is the realization of the data on variables with a Gaussian distribution.

In portfolio analysis, it is believed that a BN model with continuous variables can make very successful predictions by supporting

Table 4
MCMC simulation results of the prediction functions for stocks.

	ULKER	ULUUN	TUKAS	AEFES	CCOLA	KERVT	PENGD
Mean	−0.0099	0.0067	−0.0094	−0.0066	−0.0110	−0.0057	−0.0133
St. Dev.	0.0204	0.0411	0.0268	0.0225	0.0202	0.0286	0.0307
MSE	**0.0010**	**0.0034**	0.0018	0.0011	0.0011	0.0019	0.0026
RMSE	**0.0322**	**0.0583**	0.0424	0.0338	0.0330	0.0430	0.0513
	LKMNH	**MPARK**	**ENJSA**	**ODAS**	**AKSEN**	**ZOREN**	
Mean	0.0035	−0.0135	−0.0135	−0.0102	−0.0199	−0.0273	
St. Dev.	0.0348	0.0255	0.0201	0.0337	0.0266	0.0263	
MSE	0.0024	0.0017	0.0011	0.0025	0.0020	0.0024	
RMSE	0.0493	0.0414	0.0339	0.0498	0.0451	0.0493	

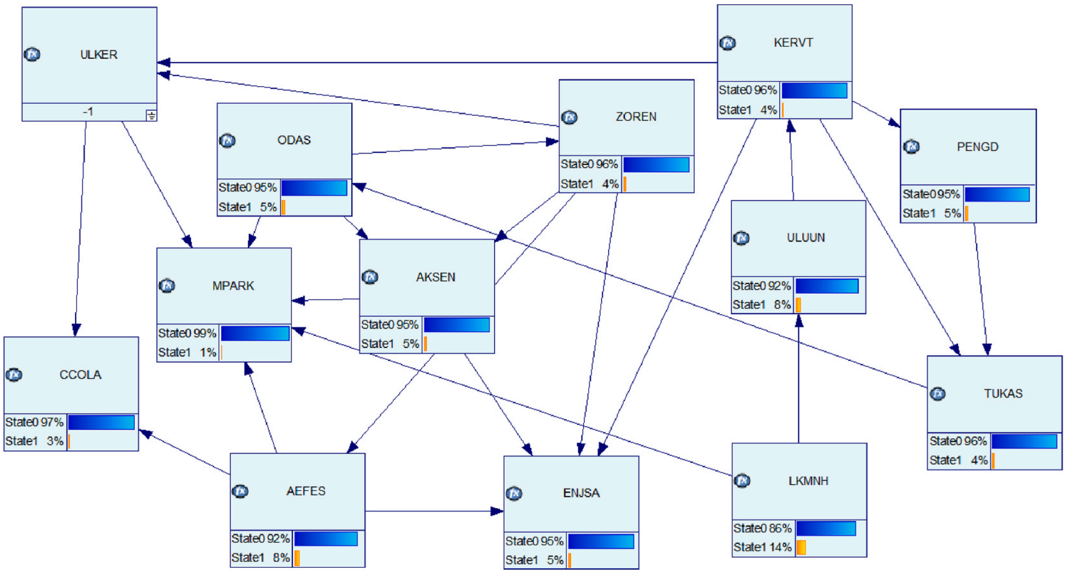

Fig. 5. Propagation of information in GBN in the event of a decline in ULKER stock.

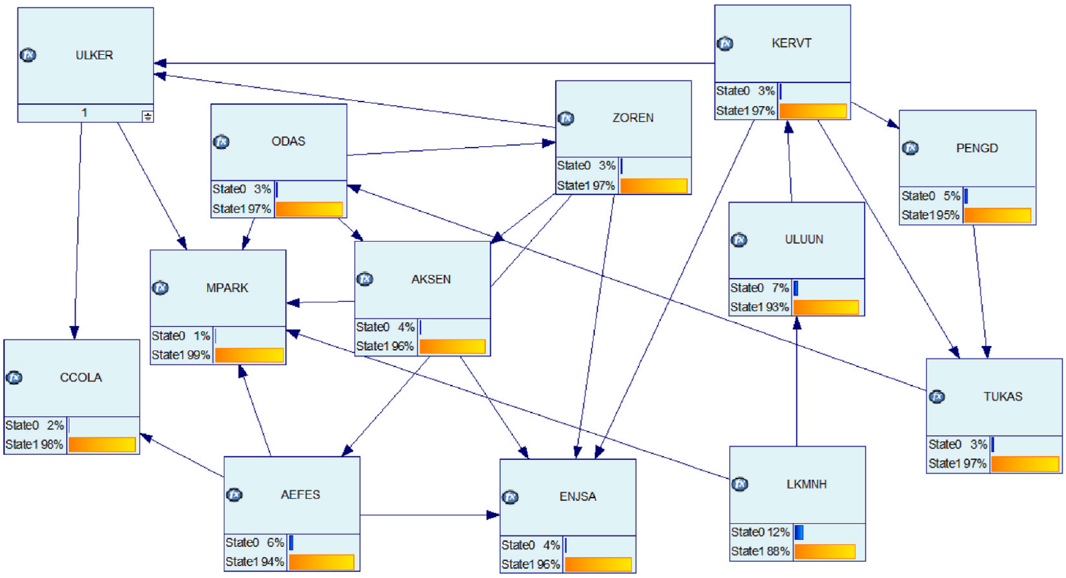

Fig. 6. Propagation of information in the GBN in the event of an increase in the ULKER stock.

MCMC simulation while performing risk analysis on potential investments made with investment instruments in the market in a basket created by the investor. By using BNs and financial market knowledge, managers can make tangible decisions based on scientific and objective tools. In summary, the dynamic nature of BNs not only defined the current situation for the pandemic period but also enabled the Turkish economy to simulate any external influence in real-time.

In future studies, a comparison between continuous and discrete variables can be made with BN models to observe the effects of discretization methods. By discretizing the data, the extent of loss of information obtained from the a priori probabilities of the nodes can be observed.

Additional information

No additional information is available for this paper.

CRediT authorship contribution statement

Ersin Sener: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Ibrahim Demir:** Supervision, Resources, Project administration, Methodology, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix 1. Definitions of the variables and data sources

Data are the end-of-day closing values of stocks and can be accessed at tr.investing.com. The data used in the article are available in open access. Related access links are given below.

Stocks Codes	Stocks	Source
ULKER	ULKER BISCUIT INDUSTRY INC.	https://tr.investing.com/equities/ulker-biskuvi-historical-data
ULUUN	ULUSOY FLOUR INDUSTRY AND TRADE INC.	https://tr.investing.com/equities/ulusoy-un-sanayi-ve-ticaret-as-historical-data
TUKAS	TUKAS FOOD INDUSTRY AND TRADE INC.	https://tr.investing.com/equities/tukas-historical-data
AEFES	ANADOLU EFES BREWERY AND MALT INDUSTRY INC.	https://tr.investing.com/equities/anadolu-efes-historical-data
CCOLA	COCA-COLA BEVERAGE INC.	https://tr.investing.com/equities/coca-cola-icecek-historical-data
KERVT	KEREVITAS FOOD INDUSTRY AND TRADE INC.	https://tr.investing.com/equities/kerevitas-gida-historical-data
PENGD	PENGUEN FOOD INDUSTRY INC.	https://tr.investing.com/equities/penguen-gida-historical-data
LKMNH	LOKMAN HEKIM ENGURSAG HEALTH TURISM EDUCATION SERVICES AND CONSTRUCTION INC.	https://tr.investing.com/equities/lokman-hekim-saglik-historical-data
MPARK	MLP HEALTH SERVICES INC.	https://tr.investing.com/equities/mlp-saglik-historical-data
ENJSA	ENERJISA ENERGY INC.	https://tr.investing.com/equities/enerjisa-enerji-historical-data
ODAS	ODAS ELECTRICITY GENERATION INDUSTRY AND TRADE INC.	https://tr.investing.com/equities/odas-elektrik-historical-data
AKSEN	AKSA ELECTRICITY GENERATION INC.	https://tr.investing.com/equities/aksa-enerji-uretim-historical-data
ZOREN	ZORLU ENERGY ELECTRICITY GENERATION INC.	https://tr.investing.com/equities/zorlu-enerji-historical-data

References

[1] J. Pearl, Fusion, propagation, and structuring in belief networks, Artif. Intell. 29 (3) (Sep. 1986) 241–288, https://doi.org/10.1016/0004-3702(86)90072-X.
[2] M. Scanagatta, A. Salmerón, F. Stella, A survey on Bayesian network structure learning from data, Progress in Artificial Intelligence 8 (4) (Dec. 2019) 425–439, https://doi.org/10.1007/s13748-019-00194-y.
[3] L. Sun, P.P. Shenoy, Using Bayesian networks for bankruptcy prediction: some methodological issues, Eur. J. Oper. Res. 180 (2) (Jul. 2007) 738–753, https://doi.org/10.1016/j.ejor.2006.04.019.
[4] Y. Zuo, E. Kita, Stock price forecast using Bayesian network, Expert Syst. Appl. 39 (8) (Jun. 2012) 6729–6737, https://doi.org/10.1016/j.eswa.2011.12.035.
[5] T. Adrian, M.K. Brunnermeier, CoVaR," American Economic Review 106 (7) (Jul. 2016) 1705–1741, https://doi.org/10.1257/aer.20120555.
[6] Z. Civan, G.G. Simsek, E.C. Akay, Identifying the systemically important banks of Turkey with the CoVaR method, Heliyon 6 (9) (2020), https://doi.org/10.1016/j.heliyon.2020.e04790.
[7] T. Beuzen, L. Marshall, K.D. Splinter, A Comparison of Methods for Discretizing Continuous Variables in Bayesian Networks, vol. 108, Environmental Modelling & Software, Oct. 2018, pp. 61–66, https://doi.org/10.1016/j.envsoft.2018.07.007.
[8] Y. Zuo, E. Kita, Up/down analysis of stock index by using Bayesian network, Eng. Manag. Res. 1 (2) (Jul. 2012) 46, https://doi.org/10.5539/emr.v1n2p46.
[9] F. Katsumi, E.S. Gomi, A Bayesian network model to improve stock market trend following strategies, in: Anais Do I Brazilian Workshop on Artificial Intelligence in Finance (BWAIF 2022), Sociedade Brasileira de Computação, Jul. 2022, pp. 81–92, https://doi.org/10.5753/bwaif.2022.223291.
[10] Y. Liu, H. Feng, K. Guo, The dynamic relationship between macroeconomy and stock market in China: evidence from Bayesian network, Complexity 2021 (Dec. 2021) 1–12, https://doi.org/10.1155/2021/2574267.
[11] S. Yan, J. Feng, N. Zhang, X. Zeng, Bayesian network model of China's financial risk under COVID-19 based on grey clustering, J. Grey Syst. 34 (3) (2022) 21–35. Online]. Available: http://jgs.nuaa.edu.cn/EN/abstract/article_10105.shtml. (Accessed 10 January 2023).

[12] J. Olbryś, Forecasting portfolio return based on Bayesian network model, in: W. Milo, P. Wdowiński, G. Szafrański (Eds.), Financial Markets: Principles of Modelling, Forecasting, and Decision-Making, FindEcon Monograph Series: Advances in Financial Market Analysis, first ed., University of Lodz, Poland, 2009, pp. 157–171.
[13] İ. Demir, SPSS Ile Istatistik Rehberi, first ed., Efe Akademi, Istanbul, 2020.
[14] E. Sener, H.A. Karaboga, I. Demir, Bayesian network model of Turkish financial market from year-to-September 30th of 2016, Sigma J. Eng. Nat. Sci. 37 (4) (Dec. 2019) 1497–1511.
[15] P.A. Aguilera, A. Fernández, R. Fernández, R. Rumí, A. Salmerón, Bayesian networks in environmental modelling, Environ. Model. Software 26 (12) (Dec. 2011) 1376–1388, https://doi.org/10.1016/j.envsoft.2011.06.004.
[16] S. Kumar Sharma, S. Sharma, Developing a Bayesian network model for supply Chain risk assessment, Supply Chain Forum Int. J. 16 (4) (Jan. 2015) 50–72, https://doi.org/10.1080/16258312.2015.11728693.
[17] Z. Liao, J. Duan, P. van Beek, On identifying significant edges for structure learning in Bayesian networks, Proceedings of the Canadian Conference on Artificial Intelligence 57 (3) (May 2022) 207–217, https://doi.org/10.21428/594757db.139d8e3e.
[18] J. Hu, B. Xiong, Z. Zhang, J. Wang, A Continuous Bayesian Network Regression Model for Estimating Seismic Liquefaction-induced Settlement of the Free-field Ground, Earthq Eng Struct Dyn, Dec. 2022, pp. 1–22, https://doi.org/10.1002/eqe.3804.
[19] F. Akdeniz, Olasılık Ve İstatistik, first ed., Akademisyen Kitabevi, Ankara, 2014.
[20] C. Croarkin, P. Tobias, NIST/SEMATECH e-handbook of statistical methods 1 (2014), https://doi.org/10.18434/M32189.
[21] L.E. Sucar, Bayesian networks: representation and inference, in: Advances in Computer Vision and Pattern Recognition, 2015, pp. 101–136, https://doi.org/10.1007/978-1-4471-6699-3_7.
[22] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Second, Morgan Kaufmann Publishers, Inc., San Francisco, California, 1988.
[23] Y.-C. Chen, T.A. Wheeler, M.J. Kochenderfer, Learning discrete Bayesian networks from continuous data, J. Artif. Intell. Res. 59 (Jun. 2017) 103–132, https://doi.org/10.1613/jair.5371.
[24] J. Suzuki, Bayesian network structure learning for discrete and continuous variables, in: 2012 2nd International Conference on Uncertainty Reasoning and Knowledge Engineering, IEEE, Aug. 2012, pp. 141–144, https://doi.org/10.1109/URKE.2012.6319529.
[25] D. Koller, N. Friedman, Probabilistic Graphical Models: Principles and Techniques, no. 3, MIT, Cambridge, Massachusetts, 2009.
[26] P. Spirtes, C. Glymour, R. Scheines, Causation, Prediction, and Search, The MIT Press, 2001, https://doi.org/10.7551/mitpress/1754.001.0001.
[27] M.J. Druzdzel, GeNIe: a development environment for graphical decision-analytic models, J. Am. Med. Inf. Assoc. (1999) 1206. PMCID: PMC2232868.