https://doi.org/10.20965/jaciii.2017.p0031

Invited Paper:

A Review of Data Mining Techniques and Applications

Ratchakoon Pruengkarn, Kok Wai Wong, and Chun Che Fung

School of Engineering and Information Technology, Murdoch University
Perth, Australia
E-mail: {r.pruengkarn, k.wong, l.fung}@murdoch.edu.au
[Received October 24, 2016; accepted December 19, 2016]

Data mining is the analytics and knowledge discovery process of analyzing large volumes of data from various sources and transforming the data into useful information. Various disciplines have contributed to its development and is becoming increasingly important in the scientific and industrial world. This article presents a review of data mining techniques and applications from 1996 to 2016. Techniques are divided into two main categories: predictive methods and descriptive methods. Due to the huge number of publications available on this topic, only a selected number are used in this review to highlight the developments of the past 20 years. Applications are included to provide some insights into how each data mining technique has evolved over the last two decades. Recent research trends focus more on large data sets and big data. Recently there have also been more applications in area of health informatics with the advent of newer algorithms.

Keywords: data mining, data mining techniques, data mining application, big data

1. Introduction

Data mining can be viewed as a part of the natural evolution of information technology. It is used extensively to develop innovative solutions that address complex practical problems in today's society and to meet ever-increasing challenges. The main goal of data mining is to extract useful information or knowledge from large datasets, in which the data may be unstructured, complex, and from multiple sources. It is important to build an adaptive computer system capable of learning in order to handle diverse problems by combining theory from many fields, including computer science, engineering and neuroscience.

The purpose of this review is to re-examine various data mining concepts and algorithms and to map the development of this field over the past two decades. The literature included in this review comes from the period of 1996 to 2016 and was obtained from the online databases of IEEE Xplore, Journal of Advanced Computational Intelligence and Intelligent Informatics and Science Direct. The review covers two main categories: predictive methods and descriptive methods. Moreover, the new trend of large data sets and big data is included. The differences between data mining and big data will also be examined. However, due to the extensive number of publications available on data mining, only selected publications are used in this review to highlight the development of the field.

The structure of this paper is as follows. Section 2 provides a brief background of the data mining process. Section 3 presents a review of different data mining techniques. Section 4 examines the difference between data mining and big data. Conclusions are presented in Section 5.

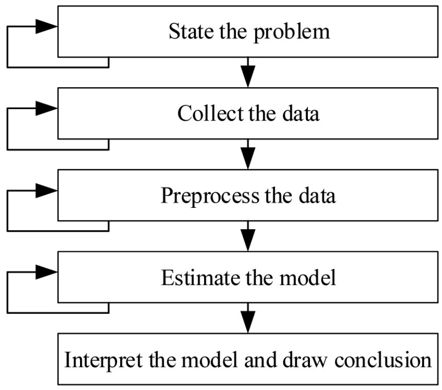

Fig. 1. The data mining process [2].

2. The Data Mining Process

Data mining is a step in Knowledge Discovery and Data mining (KDD), a process that consists of applying data analysis and discovery algorithms that produce a particular enumeration of patterns over the data under acceptable computational efficiency limitations [1]. In this paper, we focus on the data mining steps in the KDD process.

The general process of discovering knowledge from data consists of five steps: 1) state the problem, 2) collect the data, 3) preprocess the data, 4) estimate the model, and 5) interpret the model and draw conclusions as shown in **Fig. 1** [2].

2.1. State the Problem

It is important to state the problem of interest clearly. In this stage, there are initially several hypotheses for how to solve a problem. A set of variables is also specified. The data mining expert and the application expert should interact closely to state the problem as precisely as possible [2].

2.2. Collect the Data

This step concerns the collection and generation of data used for the data mining process. Designed experiments and observational approaches are two possible data collection processes. In the designed experimental approach, an expert controls the data generation process. When the observational approach is used, data are generated randomly. Understanding how data are collected is very important for understanding the nature of the data used. Therefore, a priori knowledge can be advantageous for modeling and interpreting results [2].

2.3. Process the Data

Data preprocessing is usually done after collecting the data from existing databases, data warehouses and data marts. Outlier detection and scaling, encoding, and selecting features are typical preprocessing tasks [2].

2.3.1. Outlier Detection and Removal

Outliers may be noise caused by measurement, coding, and recording errors, but they may also be important signals caused by hidden exceptional phenomena. If it is obvious that the observed outliers are noise, there are two main approaches for dealing with them: 1) detect and remove the outliers, and 2) develop a robust model [2].

2.3.2. Scaling, Encoding and Selecting Features

Due to the huge amount of data and the different ranges of data available today, scaling is necessary to organize the original data into similar weights or ranges to be applied to the data mining technique. Application-specific encoding methods can be used to achieve dimensionality reduction by providing a smaller number of informative features for data modeling [2].

2.4. Estimate the Model

In this stage, the data mining technique is used to construct and establish a data mining model. There are various techniques for developing an appropriate model for solving the identified problem. Some possible techniques, which constitute a major part of this review, are described in Section 3.

2.5. Interpret the Model and Draw Conclusions

The results from data mining models need to be interpretable and useful for human understanding and decision making. Data mining models are expected to present highly accurate results of a high-dimensional model [2].

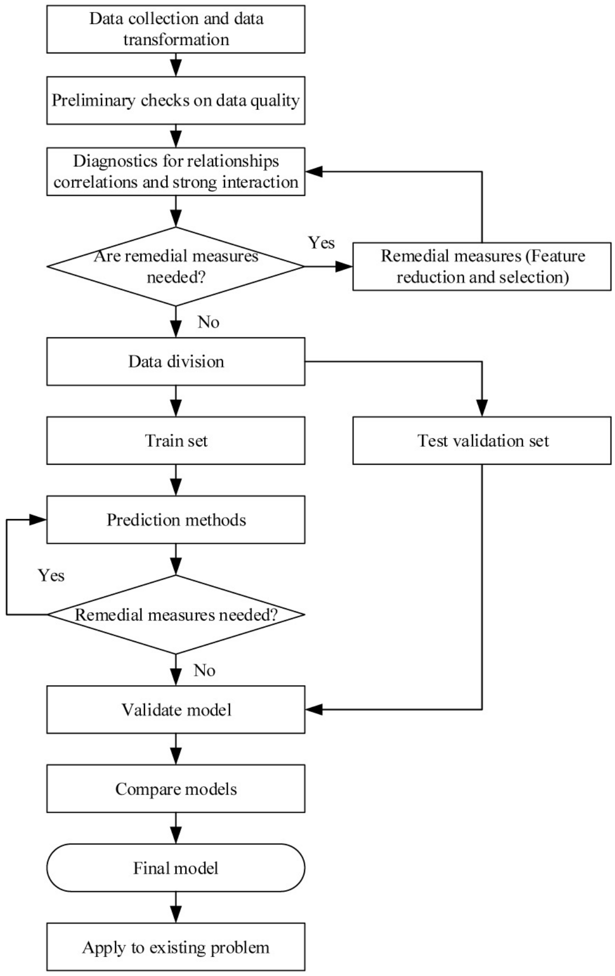

Fig. 2. The stages of predictive methods [3].

3. Data Mining Techniques and Their Applications

Data mining can be categorized into two main areas: predictive and descriptive models [3]. Predictive tasks consist of classification, regression, time series analysis and prediction. Descriptive tasks include clustering, summarization association rules, and sequence discovery. We examine each of these in the following subsections.

3.1. Predictive Methods

A predictive model predicts values for new or test data using known results found from training data [3]. Predictive data mining processes are as shown in **Fig. 2** [4]. Classification, regression, time series analysis and prediction are considered predictive data mining methods.

3.1.1. Classification Methods

Classification is the process of identifying which of a set of categories or classes an object belongs to, on the basis of relevant data properties [5]. It is referred to as supervised learning because classes are pre-determined by examining the data. Pattern recognition is a type of classification in which an input pattern is classified into one of several classes based on its similarity to predefined classes [3]. This section reviews six classification categories: Bayesian decision theory, rule-based classification, neural networks, support vector machines, lazy learners and other classification methods.

1) Methods based on Bayesian decision theory

The Bayesian classifier is a statistical classifier based on Bayes' theorem that classifies unknown instances using maximum likelihood. Bayes' theorem is presented in conjunction with the formula for conditional probability as follows [6].

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (1)$$

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} \quad (2)$$

$$\text{posterior} = \frac{\text{liklihood} \times \text{prior probability}}{\text{evidence}} \quad (3)$$

where

- $P(A)$ and $P(B)$ are the probabilities of events A and B, respectively, with A and B independent.
- $P(A|B)$ is the conditional probability of event A given that event B is true.
- $P(B|A)$ is the conditional probability of event B given that event A is true.

Bayesian decision theory is a fundamental statistical method in the pattern classification field. The objective of this theory is to minimize the probability of making a wrong decision. The Bayesian decision rule can be summarized in the following algorithm [5].

- Let D_k be the decision rule regarding the 'natural' state A_k.
- Given a measurement x, the error related to state A_k is defined as: $P(\text{error}|x) = 1 - P(A_k|x)$.
- The probability of failure is minimized.
- The Bayesian decision rule is given by the assertion:
- "Choose D_k if $P(A_k|x) > P(A_j|x), \forall j \neq k$".
- Equivalently, "Choose D_k if $P(x|A_k)P(A_k) > P(x|A_j)P(A_j), \forall j \neq k$".

One researcher derived parameters of a Bayesian network in order to assess the risk in GIS data. The Bayesian decision rule can also be used to discover rules and analyze data in a large database. It is a well-known technique for classification in many areas, such as medicine [7, 8] and renewable energy [9]. These have been highlighted in **Table 1**.

Table 1. Bayesian applications.

Year	Application	Description
1996	Geographic Information Systems [10]	Assessing the risk of desertification in burned forest areas
1999	Database system [11]	Decreasing computational time for mining large databases
2000	Knowledge discov-ery [12]	Analyzing ticket sales data
2004	Medical [13]	Extracting human limb regions
2006	Medical [14]	Building an automatic interracial spikes detection model
2008	Medical [15]	Assessment of individual risk of relapse or progression in patients diagnosed with brain tumors undergoing RT postoperatively
2016	Renewable energy [16]	Classification of clouds to determine atmospheric parameters before building solar plants

2) Rule-based Classification

Rule-based classification uses classification rules, such as a set of IF-THEN rules, to predict the class label. A rule takes the form $X \rightarrow Y$, where X represents the rule antecedent or condition, and Y represents the rule consequent or class label [5].

Table 2. Rule-based applications.

Year	Application	Description
2008	Radiology [17]	Developing a computer-aided diagnostic scheme for lung nodule detection in order to assist radiologists in the detection of lung cancer from thin-section computed tomography (CT) images
2015	Stock trading [18]	Creating several rule pools with a rule-based evolutionary algorithm
2015	Computer security [19]	Using a class association rule-based system to detect network attacks
2016	Medical [20]	Classifying biological data to predict the accuracy of DNA variants and present human interpretable rules

The rule-based method has been used for classification in various fields, such as radiology, marketing, computer security and medicine. Other techniques including decision trees, evolutionary algorithms and genetic algorithms are used in rule extraction. Some applications of the rule-based approach are highlighted in **Table 2**.

IF-THEN rules can also be extracted from a decision tree. Decision Trees are used to assign object membership in different classes based on attributes [5]. The tree structure decision method denotes tests on an attribute with internal or nonleaf nodes. Each branch represents an outcome of a test, and leaf nodes or terminal nodes represent class labels. The definition of a decision tree is as follows [3] .

Suppose that for a database $D = \{t_1, \ldots, t_n\}$ [21], where $t_i = \langle t_{i1}, \ldots, t_{ih} \rangle$, the database schema contains the attributes $A = \{A_1, \ldots, A_h\}$ and a set of classes $C = \{C_1, \ldots, C_m\}$. A decision tree or classification tree associated with D has the following properties:

- Each internal node is labeled with an attribute, A_i.
- Each arc is labeled with a predicate that can be applied to the attribute associated with the parent.
- Each leaf node is labeled with a class, C_j.

Solving the classification problem using a decision tree is a two-step process:

1. Decision tree induction: construct a decision tree using training data.
2. For each $t_i \in D$, apply the decision tree to determine its class.

There are many advantages of using a decision tree for classification. First, it is easy to use and efficient. Second, it is easy to interpret and understand the rules. Third, the tree size is independent of the database size. Thus, a decision tree can work with a large database. Finally, a tree can be constructed for data with many attributes. On the other hand, a decision tree cannot handle continuous data well; the correct branches in the tree cannot be constructed with missing data; and the tree may overfit the training data. A tree pruning technique can be used to improve the performance of the tree in the classification phase by removing redundant comparisons or removing sub-trees. A final disadvantage is that the decision tree process ignores correlations among attributes. Some popular decision tree techniques are ID3, C4.5, C5.0 and CART [4].

Decision trees were first applied to develop and enhance rules. Later, they were used as a classifier to recognize patterns. Recent research often combines decision trees with other algorithms to improve the classification performance. Some of the applications of the decision tree are highlighted in **Table 3**.

From **Table 3**, it can be seen that decision trees have been combined with other theories, such as genetic programming and fuzzy theory, for rule generation. Additionally, they have been used as a classifier in many fields such as image recognition, art and geoinformatics.

A recent trend in the use decision trees is the application of other algorithms or hybrid classification algorithms, such as ant colony optimization and fuzzy theory [21, 27, 28], to construct a decision tree in order to overcome the problems derived from uncertainty, overlapping decision boundaries, and the possibility of overfitting.

3) Neural Networks

The artificial neural network (ANN) is an interconnected group of artificial neurons that compute output values from inputs. It learns from its past experience and errors in a non-linear parallel processing manner. The neuron is the basic calculating entity that computes with a number of inputs and delivers one output by comparison with a threshold. The computational processing is done by an internal structural arrangement consisting of hidden layers and weights, which utilize the back propagation and feed-forward mechanisms to deliver an output of the desired accuracy [29].

Table 3. Decision tree applications.

Year	Application	Description
1998	Rule generation [22]	Using decision trees to develop fuzzy rules for continuous-valued inputs and outputs
1999	Learning rules [23]	Combing decision trees with genetic programming to enhance learning rules in neural networks
2002	Identify I/O sets [7]	Applying C4.5 to the results from reinforcement learning incorporated with fuzzy inference in order to derive a proper action decision tree
2006	Pattern recognition [8]	Using decision rules for handwriting pattern classification
2008	Rainfall intensity [9]	Developing a heterogeneous hierarchical classifier to identify rain areas and rainfall intensity in those areas
2009	Product impression design [24]	Using questionnaire results to develop a decision tree for user impressions
2009	Image recognition [25]	Developing an ensemble decision tree and generating an on-line Random Forest for image recognition
2012	Artwork [26]	Using C4.5 to classify the authentication of works of art

Current neural network techniques aim to developing better training methods, better network architecture, and the ability to handle uncertainty and overfitting problems. It can also be observed that there is increasing integration of fuzzy logic into neural networks used in diverse engineering areas. **Table 4** presents some highlights of the applications of neural networks.

4) Support Vector Machines

The concept of support vector machine (SVM) is to find a hyperplane in the space of possible inputs or labeled training data based on the structural risk minimization (SRM) principle and statistical learning theory. The original optimal hyperplane is a linear classifier for binary classes. However, SVM can be used to as a non-linear classifier by applying a kernel function, such as a polynomial or radial basis function (RBF), and the multi-layer perceptron (MLP) to maximum-margin hyperplanes. Thus, the maximum-margin hyperplane is fit in the transformed feature space. SVM can be applied to regression problems using a loss function based on distance measures, such as quadratic, Laplace, Huber and ε-insensitive functions [36].

SVM was derived from statistical learning theory. It

Table 4. Neural networks applications.

Year	Application	Description
1998	Information retrieval [30]	Producing a word frequency measure from documents
1999	Speed control [31]	Controlling the speed of an elastic drive system
2003	Financial [32]	Developing some parametric to guide a neural network of a target pricing formula
2007	Medical [33]	Adjusting a neural network model using associative ontology to create relations between stimulus words and associated words
2011	Medical [34]	Using a neural network for diagnosis of liver cancer from medical images
2014	Weather forecasting [35]	Applying multi-layered back propagation to predict weather conditions

Table 5. Support vector machines applications.

Year	Application	Description
2002	Industrial [37]	Extracting decision rules for Polycythemia Vera by using an SVM as a classifier
2005	Marketing [38]	Evaluating a direct marketing mailing campaign
2007	Medical [39]	Classifying liver disease
2008	Unbalanced data [40]	Developing a weighted harmonic mean algorithm to improve the SVM classifier
2009	Human resource [41]	Using a weighted SVM to support decision making in human resource selection
2012	Recognition [42]	Combining an SVM and an online learning algorithm for hand shape recognition
2014	Financial [43]	Enhancing the prediction of stock market behavior with a one-against-one SVM

was created mainly as a linear classifier. The nonlinear classifier later created. SVM is closely related to the kernel methods, the large margin classifier, soft margin algorithms and reproducing kernels. Later, SVM was extended for multi-class and regression problems. Applications of support vector machines are highlighted in **Table 5**.

5) Lazy Learners

Eager learners, such as decision trees, Bayesian methods and support vector machines, construct the classification model before classifying the new or testing data. In contrast, lazy learners do less work when training and more work when making a classification or prediction. K-nearest neighbor classifiers and case-based reasoning classifiers are examples of lazy learners [44].

a) K-nearest Neighbor Classifiers

Table 6. *K*-nearest neighbor applications.

Year	Application	Description
2000	Medical fraud [45]	Classifying prescriptions written by General Practitioners as appropriate or inappropriate
2011	Network security [46]	Detecting large-scale attacks such as DoS in real time
2016	Big data [47]	Using k-NN in the mapping phase from different training data splits for big data
2016	Mechanical [48]	Identifying different gear crack levels

The K-nearest neighbor classifier compares a given testing instance with k similar training instances. The distance between a testing instance and a training instance can be computed by some distance metric such as Euclidean distance. The optimal value for k can be determined by beginning with $k = 1$ and estimating the error rate of the classifier. The number of k that gives the minimum error rate may be selected [44].

Most applications of K-nearest neighbors involve classifying and detecting. Furthermore, it can be used to develop hybrid systems for big data and real-world time series problems. **Table 6** provides some highlights of applications of K-nearest neighbors.

b) Case-based Reasoning Classifiers

Case-based reasoning (CBR) classifiers store training data as cases for problem-solving unlike the nearest-neighbor classifier, which stores training instances as points. A case-based reasoner first checks if a training cases exists when given a new case to classify. If a training case is found, then the solution for that case is returned. However, if no case is found, the case-based reasoner searches for neighboring training cases that are similar to the new case. In order to determine a solution for the new case, the reasoner tries to combine the solutions from the neighboring training cases. The challenge of developing case-based reasoning classifiers is finding a good similarity metric, suitable methods for combining solutions, the selection of features to index training cases, and the development of efficient indexing techniques [44]. **Table 7** highlights some developments in case-based reasoning.

6) Other Types of Classification

a) Classification using Genetic Algorithms

Genetic algorithms (GA) are an example of evolutionary computing methods and optimization-type algorithms. Biological evolution is the basis for evolutionary computing algorithms in that the fittest evolution is produced over time [3]. An initial population, a set of individuals or chromosomes, consisting of randomly generated rules is created. Each rule can be represented by a string of bits. Given two parents from the population, genetic operators such as crossover and mutation are then used to generate new children or offspring. New offspring

Table 7. Case-based reasoning applications.

Year	Application	Description
1997	Legal [49]	Using case rules to interpret court judgments
2002	Medical [50]	Using the importance of features for diagnosis of lung pathologies
2005	Paper machine [51]	Combing linguistic equations and fuzzy logic in case-based reasoning to monitor the paper web break tendency in modern paper machines
2010	Web service [52]	Applying case-based reasoning to store planning and related information to make planning much faster when users have similar needs
2015	Biomedical [53]	Exposing the ontological application in the field of clinical decision support. Case-based reasoning memorizes and restores experience data for solving similar problems with the help of the matching approach and defined interfaces of ontologies
2016	Bankruptcy [54]	Developing new case-based reasoning for avoiding bankruptcy in borderline situations

Table 8. Genetic algorithms applications.

Year	Application	Description
1998	Placement time [55]	Applying a GA to searching for tape feeder arrangement to optimize placement time for chip mounting
2006	Work scheduling [56]	Finding the optimal solution for the 24-7 schedule in a restaurant
2007	Algorithm [57]	Using a partitioning method to adapt GA in order to find an effective solution
2009	Medical [58]	Matching lung images in many subareas
2016	Recognition [59]	Improving performance of human action recognition with GA and convolutional neural networks

consist of the fittest rules in the current population and are generated by switching subsequences of a string. In crossover, substrings from pairs of rules are swapped to form new pairs of rules, whereas in mutation, randomly selected bits in a rule's string are inverted. Finally, the fitness of a rule is assessed by its accuracy when classifying a set of training instances [44].

GAs have been introduced in optimization problems. Several adaptive natural mechanisms such as ant colony optimization, particle swarm optimization, and memetic algorithms have now emerged to capitalize on its important strengths. However, a simple GA cannot compete with more sophisticated optimization algorithms that are tailored towards a specific problem and exploit its particular features. Thus many hybrid approaches have been developed in order to improve the performance of GAs. **Table 8** highlights some applications of GAs.

b) Rough Set Approach

Rough set (RS) theory is used to discover structural relationships within imprecise or noisy data in the classification domain [44]. The RS is defined by two topological notions, namely interior and closure, known as approximations. First, the RS transforms the given dataset into an information table that consists of the objects' attributes and their values. Second, all of the data tables form equivalence classes within the given training data. Third, the discernibility matrix, in which elements are distinguished between equivalence classes, is constructed. RS can be used to approximately or roughly define such classes. The lower approximation and upper approximation are then approximated for a set of a given class X. The lower approximation consists of all data objects that can be classified as member of X with certainty, whereas the upper approximation consists of all data objects that can be described as not being elements of X. The boundary region of X is considered to be those objects that cannot be classified with any certainty as either inside or outside of X. Finally, decision rules can be generated for each class, and a decision table is used to represent the rules [5].

Rough set theory is used in feature selection for data representation; it rejects the irrelevant features. The concept of dynamic reducts was proposed to find a robust and generalized feature subset. Many methods of calculating reducts that are based on genetic algorithms and feature weighting and instance selection and rough set theory have been developed. An emerging technique for clustering tasks is the simple extension of rough set theory to cluster analysis and mapping clustering datasets to decision tables. In addition, rough sets are used to develop efficient heuristic searches for relevant tolerance relations that allow for extraction of objects in the data. Finally, they are used in rule induction. Particle swarm optimization and Boolean reasoning are used in reduction algorithms [60]. Currently researchers are working on integrating rough set theory with other contemporary techniques, genetic algorithms and evolutionary methods [61]. **Table 9** highlights applications of rough sets.

c) Fuzzy Set Approach

Fuzzy set theory is used to deal with vague or inexact facts. Elements can belong to more than one fuzzy set.

Early fuzzy set research focused on handling approximate information and uncertainty to generate decisions in scientific applications. It has further been used to develop formalized tools to deal with the intrinsic imprecision in the shape of membership functions and the transition between crisp extremes through fuzziness to a wide variety of problems. Applications include developing standards and type II fuzzy sets, improving the completeness and consistency of available data, and supporting the develop-

Table 9. Rough set applications.

Year	Application	Description
2000	Feature selection [62]	Using rough sets to provide domain knowledge expressed for data analysis
2004	Apply theory [63]	Applying rough set theory to select attributes as nodes of a decision tree
2008	Utility program [64]	Defining the numerical values in utility programs
2009	Pattern classification [65]	Improving overall performance of unsupervised ANN
2011	Prediction model [66]	Analyzing time series data of trick-wise price fluctuations
2014	Imperfect data [67]	Handling imprecision and uncertainty in data

Table 10. Fuzzy set applications.

Year	Application	Description
2000	System architecture [68]	Using type-2 fuzzy sets to analyses system architecture hypergraphs
2001	Transportation [69]	Integrating fuzzy sets and fuzzy DEA to forecast short-term production efficiency of a city bus company
2006	Language [70]	Dealing with word meanings in ambiguous language
2016	Linguistic [71]	Applying fuzzy rules to cope with asymmetrical fuzzy sets

ment of knowledge-based systems using fuzzy sets [44]. Some highlights of fuzzy set applications are presented in **Table 10**.

3.1.2. Regression Methods

Regression analysis is a statistical methodology that is most often used for numerical prediction although other methods exist for this. Regression also encompasses the identification of distribution trends based on the available data. It can be used to solve problems such as forecasting by fitting a set of points to a curve. Linear regression is a simple regression method that represents a relationship between the input data and output data as follows [72].

$$y = c_0 + c_1 x_1 + \cdots + c_n x_n \quad \ldots\ldots\ldots \quad (4)$$

where $c_0, c_1, \ldots, c_n$ are the regression coefficients and $x_0, x_1, \ldots, x_n$ are the input parameters. y is the output parameter, which depends on the relationship of the input parameters.

Multiple linear regression is used when there are n input variables, called predictors or regressors; one output variable, called the response; and $n+1$ constants which are chosen during the modeling process to match the input examples. Popular regression methods, such as nonparametric regression, robust regression, ridge regression and deep learning, are described as follows.

Table 11. Univariate time series applications.

Year	Application	Description
1996	Economic [74]	Examining long-run neutrality in 10 different countries
1998	Biomedicine [75]	Modeling and forecasting monthly patient volume
2004	Energy [76]	Forecasting hourly electricity spot prices
2006	Weather [77]	Forecasting weather density
2010	Rainfall [78]	Forecasting the summer monsoon rainfall over India
2015	Clinic [79]	Investigating individual client progress and determining whether symptoms have improved

1) Time Series Analysis

A time series is a collection of observations made chronologically. The values are obtained at evenly spaced time points, e.g. daily, weekly or hourly. Thus, time series data is characterized as being numerical and continuous. It is also always considered as a whole, instead of an individual, numerical field. There are two main goals of time series analysis: 1) identifying the nature of the phenomenon represented by the sequence of observations and 2) forecasting the possible evolution as time goes on. The increasing use of time series data has initiated a great deal of research and development in the field of data mining [73]. Univariate and multivariate time series are both described here.

A univariate time series is a sequence of measurements of the same variable collected over time. Most often the measurements are made at regular time intervals. There are four major movements in characterizing time-series data:

- Trends or long-term movements: A trend can indicate the general direction of attribute values over a long interval of time. This can be represented by trend curve or trend line.
- Cyclic movements or cyclic variations: These refer to cyclic behavior that may or may not be periodic. They do not necessarily follow exactly similar patterns after equal intervals of time.
- Seasonal movements or seasonal variations: These detected patterns may be calendar-related, such as relating to the time of year, month or day.
- Irregular or random movements: These are occasional detected motions of time series.

Economic researchers have used univariate time series to examine the economy of a long-run neutrality proposition. They can also be used for forecasting in biomedicine, energy and environmental applications as highlighted in **Table 11**.

The multivariate time series approach involves measuring the same quantity or time series depending on

Table 12. Multivariate time series applications.

Year	Application	Description
2000	Environmental [81]	Finding vegetation and climate variation trends over the United States
2002	Chemistry [82]	Calculating the hydraulic parameters of a stream-aquifer system using pumping test data
2005	Neuroscience [83]	Analyzing multichannel EEGs/MEGs
2007	Medical [84]	Learning and building a behavioral profile of a person from their daily activities
2015	Agricultural [85]	Forecasting short-term irrigation demands
2016	Traffic [80]	Forecasting traffic systems

some fundamental quantity that leads to a multivariate series [80]. The aim is to discover hidden patterns in multi-dimensional data. However, most complex dynamic data cannot be described with linear models. Thus, this has led to the development of non-linear models and stochastic volatility models. Multivariate time series are well known in variety of applications, including in environmental, science, medical and agricultural domains as highlighted in **Table 12**.

2) Nonparametric Regression

Nonparametric regression is used to model the relationship between the response variable and the associated predictor variables, which need to be constructed according to information derived from the data. Methods of nonparametric regression analysis have been rendered possible by advances in statistics and computing [86].

Nonparametric regression can be used for developing a kernel to deliver interpretability and reasonability to data. Highlights of the applications of nonparametric methods are shown in **Table 13**.

3) Robust Regression

Robust regression analysis searches for the relationship between one or more independent variables and a dependent variable using regression methods such as least squares [72]. These methods provide relatively insensitive, consistent and highly efficient estimators when there are slight violations of the standard assumptions in the assumed statistical model and for the rational consideration of outliers in regression analysis [93].

In dairy science and chemistry, robust regression can provide reasonable analysis of outliers and handle high leverage points by decreasing the weight given to specific data values that are in disagreement with the majority of the sample. In data analysis research area, it has been combined with the forward search algorithm and EM algorithm to handle missing values. In addition, robust regression uses some outlier identification to help in performing the analysis. **Table 14** highlights some of the applications of robust regression.

Table 13. Nonparametric applications.

Year	Application	Description
2001	Econometrics [87]	Developing several kernel-based consistency tests of a hypothesis of additivity in nonparametric regression
2004	Biosystem [88]	Constructing a gene network from time series microarray gene expression data
2011	Statistical Planning [89]	Developing the theory of nonparametric regression for the classical case of responses missing at random
2014	Energy [90]	Creating online condition monitoring of a wind power system
2015	Geoscience [91]	Estimating spatial-temporal mountain glaciers from satellite images
2016	Economic [92]	Predicting microeconomics and financials in the direction of price changes

Table 14. Robust regression applications.

Year	Applications	Description
1999	Dairy science [94]	Developing conversion equations for production, type and health traits
2000	Data analysis [95]	Developing methodology to avoid multiple imputations for missing data
2003	Biochemistry [96]	Studying the binding of the methylphenazinium cation with double-stranded DNA
2006	Chemistry [97]	Detecting outlier data in analytical chemistry
2015	Neuroimage [98]	Analyzing large neuroimaging cohorts, detecting imaging genetics and avoiding false positives in a large-scale analysis of brain-behavior relationships
2016	Data analysis with improving outliers handling [99]	Proposing a new robust regression to deal with case-wise and cell-wise outliers

4) Ridge Regression

Ridge regression, also known as Tikhonoy regularization, is the most commonly used method for regularization of ill-posed problems [86]. It is a technique for analyzing multiple regression data that suffer from multicollinearity.

In policy capturing applications, human judgment is combined with ridge regression, which improves the judgment of policy models [100]. Least squares, ridge regression and partial least squares regression are used to vary calibration data size in chemometrics research [101]. Additionally, ridge regression can be used in generalized

Table 15. Ridge regression applications.

Year	Applications	Description
1996	Behaviour [100]	Capturing human judgment
2002	Chemometrics [101]	Comparing calibration of near infrared data
2005	Epidemiology [102]	Generating linear models for controlling collinearity among metabolites
2007	Hydrology [103]	Applying ridge regression in a feature space to forecast the hydrologic time series
2016	Computational [104]	Classification of microarray gene expression

linear models in order to control for collinearity among metabolites before analyses are done [102]. This technique has also been applied to forecasting in time series data [103], in featured spaces and in classification in microarray medical data [104]. The highlights of its applications are shown in **Table 15**.

5) Nonlinear Regression

Nonlinear regression occurs when the predictors in the linear regression function are modified by some function, such as the square or square root function [86].

Kernel regression is one method of non-parametric regression that can estimate data without the assumptions required in a physical model. It is an estimation technique that is used to fit data between sample data and input data using the standardized Euclidean distance [105]. Kernel regression is a superset of local weighted regression methods, such as K-nearest neighbors, radial basis functions, neural networks, and support vector machines. To perform kernel regression, a set of weighted functions called 'kernels' is placed locally at each observational point. The kernel assigns a weight to each location based on its distance from the observational point [106].

A multivariate kernel regression specifies how the response variable, y, depends on a vector of explanatory variables, denoted by X. The objective is to find a nonlinear relation between a pair of random variables X and y as follows.

$$E(y|X) = m(X) + \varepsilon \quad (5)$$

and

$$y = m(X) + \varepsilon \quad (6)$$

where m is a weight function or kernel

The Nadaraya-Watson kernel regression below is derived based on those equations.

$$\hat{y} = \hat{m}_h(x) = \frac{\sum_{i=1}^{n} K_h(x - X_i) y_i}{\sum_{i=1}^{n} K_h(x - X_i)} \quad (7)$$

where K is a kernel with a bandwidth h.

Table 16. Nonlinear regression applications.

Year	Applications	Description
1998	Aquaculture [107]	Evaluating the effects of dietary phytase enzyme treatment of plant protein diets for rainbow trout
2003	Economics [108]	Using regression analysis to analyze performance of a university's dial-up modem pool under various time limit policies and customer behavior patterns
2006	Soil [109]	Estimating the impact of puddling, tillage and residue management on wheat
2011	Circulatory systems [110]	Diagnosing the failure determination of target systems in chemical plant operation
2014	Environment [111]	Predicting the compressive strength of concrete
2016	Thermal Biology [112]	Calculating and comparing characteristics and thresholds of metabolic rate and evaporative water loss in Australian rodents

Even though any function that satisfies several mathematical requirements can be a kernel, the Gaussian kernel, called the bandwidth, kernel radius or metric window, is usually used:

$$K_h(x - X_i) = e^{-\frac{(x-X)^2}{2\sigma^2}} \quad (8)$$

Table 16 presents highlights of nonlinear regression applications. One-way analysis of variance and mean differences are used to analyze the digestibility trails of rainbow trout in aquaculture research. The Find Laws algorithm of PolyAnalyst 4.3 is used to generate nonlinear prediction models for analyzing the performance of a university's dial-up modem pool. In soil science, predictions of wheat seedling emergence and growth are made using nonlinear regression models.

6) Deep Learning

The architecture of the original neural networks consists of an input layer, hidden layer and output layer, where the output from one layer is used as the input to the next. There may be a single hidden layer or more than one in multiple layer networks (MLPs). Backpropagation was the first computationally efficient model of how neural networks could learn multiple layers of representation. It required labeled training data, and it did not work well in deep networks. However, multilayer neural networks can overcome the limitations of backpropagation learning [113]. MLPs can be used to learn complex functions to solve non-linear problems. In this case, the network is composed of one or multiple hidden layers which have some type of sigmoid activation function, such as the long-sigmoid and hyperbolic tangent. Adding

Table 17. Deep learning applications.

Year	Applications	Description
2012	Neural Networks [119]	Recognizing German traffic signs
2014	Robotics [120]	Integrating sensory-motor time-series data and the self-organization of multimodal fused representations based on a deep learning approach
2015	Dialogues recognition [121]	Recognizing Chinese dialogues
2016	Human actions [122]	Recognizing human actions in video

multiple hidden layers to MLPs is also called a deep network [114]. Therefore, deep learning methods are representation learning methods with multiple levels of representation, obtained by composing non-linear modules that each transforms the representation at one level into a representation at a higher level. With the composition of enough such transformations, very complex functions can be learned [115]. Deep neural networks have been successfully used for regression and classification problems [116].

Deep neural networks attempt to disentangle intricate aspects of input by creating multiple levels of representation. They are inspired by the functioning of the visual cortex, in which each layer receives input from a layer below it, transforms its representations and then propagates it to the layer above it [117].

Deep learning trends focus on developing better training methodologies and enhanced network architectures. Current research in this field works on large and complex architectures with large amounts of data. A challenge is to run deep learning in highly efficient manner and on cheap mobile devices without any extra hardware [118]. Useful applications like machine translation, speech recognition, facial recognition and human movement recognition have utilized deep learning methods. Highlights of deep learning applications are shown in **Table 17**.

3.2. Descriptive Methods

Descriptive methods attempt to describe all of the data under analysis or the process used to generate the information.

3.2.1. Clustering Methods

Clustering is also called data segmentation. It is similar to classification in that data are processed into classes or clusters except that in clustering the classes are not predefined. The objects within a cluster share significant similarities but are very dissimilar to objects in other clusters [44].

1) K-means clustering

Table 18. *K*-mean clustering applications.

Year	Application	Description
1998	Image Processing [124]	Segmentation of three-dimensional image data based on a novel combination of adaptive *K*-mean clustering and knowledge-based morphological operations
2005	Physics [125]	Partitioning parallel hierarchical *N*-body interactions from fluid flow at high Reynolds numbers to gravitational astrophysics and molecular dynamics
2009	Food security [126]	Using clustering analysis to highlight the linkages between food insecurity and poverty
2010	Medical [127]	Converting the input grey-level brain image to a color space image and operation on the image labeled by clustering
2015	Biomedical [128]	Clustering vectors of observations at different time instants and then estimating the initial innervation pulse train (IPT)

K-means is one of the simplest unsupervised learning algorithms. The *K*-mean algorithm is used to distinguish between similar and dissimilar data points to represent a meaningful grouping [123]. First, k points at random cluster centers are selected. Instances are then assigned to their closest cluster center according to some similarity distance function. Next, the centroid or mean of all instances in each cluster is computed. Finally, the data is clustered into k groups where k is predefined [5].

In the early stage, research focused on the introduction of *K*-means in clustering areas. After that, all modifications and improvements were based on the basic *K*-mean clustering. An adaptive *K*-mean clustering and knowledge based morphological operation has been proposed to develop a robust algorithm for segmentation of 3D images [124]. The number of cluster is specified by partition geometry. Weighted *k*-means partitions minimize the sum of clusters to create well-localized domains and reduce the computational cost of N-body problems [125]. *K*-means have also been used to group data into classes that form linkages between food insecurity and poverty [126]. In medical research areas, such as detecting MRI brain images, this method is applied by converting the input grey-level image into a color space image and operating on the image labeled by a cluster index [127]. Combining *K*-mean clustering with the convolution kernel compensation method has been proposed to demonstrate the reliability and capability in multichannel surface EMG decomposition [128]. Some examples of applications are highlighted in **Table 18**.

There are some limitations to *K*-means, such as handling of outliers, number of clusters, empty clusters and non-globular shapes and sizes, which still require more

Table 19. Fuzzy c-mean applications.

Year	Application	Description
1999	Noise reduction and 3D ultrasonic images [131]	Reconstruction of images from data acquired via ultrasonic sensors giving noise-reduced, enhanced images.
2001	Physics and chemistry of earth [132]	Mapping soil data for detection of polluted sites
2008	Medical engineering [133]	Providing an efficient classification approach for automated segregation of brain MR images
2010	Computer in simulations [134]	Improving the detection of false alarms
2011	Engineering [135]	Providing scientific warning methods
2015	Mathematical modelling [136]	Reducing thermal errors of machine tools in the design of thermal prediction models
2016	Expert system [137]	Separating data sets by providing a consistent clustering algorithm, which can be used for N-dimensional data as well as circular data

work. Setting an appropriate initial number of clusters is also a challenging task if the nature of the problem is unknown

2) Fuzzy c-mean clustering

Fuzzy c-mean clustering (FCM) is based on the principle that each data input may belong to more than one cluster with different membership values ranging from $[0,1]$, such that the sum of membership values for each data point must be one. The total sum of all membership values in a cluster must also be smaller than the number of data points [129].

Since FCM was proposed and since it has been claimed to solve the partition factor of classes, it has been used widely. However, its convergence speed is slower than that of hard clustering. Thus, its modification and integration with other techniques have been developed for hierarchical clustering, spatial information, and weighting and context information. Fuzzy minimax models and fuzzy variants of evolutionary algorithms are also broadly studied. In addition, the modification of type-2 fuzzy clustering has been developed for MRI image segmentation [130]. Applications of FCM are highlighted in **Table 19**.

3.2.2. Association Rules

Association rules are used to identify the relationship among items in a given dataset. Rule support and confidence are two measures of the relevancy of rules. Typically, association rules are considered interesting if they satisfy both a minimum support threshold and a minimum confidence threshold. Such thresholds can be set by users or domain experts [44]. For a set of items $I = \{I_1, I_2, \ldots, I_m\}$ and a database of transactions $D = \{t_1, t_2, \ldots, t_n\}$, where $t_{i1} = \{I_1, I_{i2}, \ldots, I_{ik}\}$ and $I_{ij} \in I$, an association rule is an implication of the form $X \Rightarrow Y$ where $X, Y \subset I$ are sets of items called itemsets and $X \cap Y = \varphi$. The support for an association rule $X \Rightarrow Y$ is the percentage of transactions in the database that contain $X \cup Y$. The confidence or strength of an association rule $X \Rightarrow Y$ is the ratio of the number of transactions that contain $X \cup Y$ to the number of transaction that contain X.

Table 20. Apriori algorithm applications.

Year	Application	Description
2004	CRM [138]	Finding fuzzy association rules from quantitative transactions in a supermarket
2006	Database [139]	Extracting rules from dense databases and dependent pairs of the sets of attributes in the database
2007	Database [140]	Applying the RPF-Apriori algorithm to find fuzzy association rules from fuzzy item data
2007	Grid technologies [141]	Implementing the apriori algorithm using the Grid service infrastructure to improve response time of the calculus power of various geographic distributed resources
2012	Web-based [142]	Detecting web-based intrusion systems to identify a variety of attacks and improve the overall performance of detection systems
2016	Medical [143]	Applying association analysis to medical records data to analyze symptoms and drug combination data

1) The Apriori Algorithm

Apriori is an algorithm for mining frequent itemsets for Boolean association rules using prior knowledge of frequent itemset properties. Apriori employs an iterative approach called level-wise search where k itemsets are used to explore $(k+1)$ itemsets. If an itemset is frequent, then all of its subsets must also be frequent; otherwise, the subset is pruned [86].

The original algorithm has been optimized through means such as bringing in random sampling, parallel thoughts, adding reference points, declining rules, and changing the storing framework to improve the efficiency of algorithm rules. The weighted apriori algorithm uses single minimum support for selecting the frequent itemsets. In addition, apriori algorithms such as dynamic hashing and pruning, partition, and distributed max miners are used in parallel and distributed data mining. **Table 20** shows some of the applications of apriori algorithms.

2) Multidimensional Association Rules

Multidimensional association rules consider more than

Table 21. Multidimensional association rules applications.

Year	Application	Description
2001	Weather [144]	Using 2- and *m*-dimensional contexts for mining inter-transactional association rules to improve predictive capability
2006	Data warehouse [145]	Defining authorization rules for users and objects and assigning sensitive information rules to the main elements of a multidimensional model
2009	Web design [146]	Extracting useful information from web users' access paths
2010	Financial [147]	Designing an efficient personal financial services system by using registered information and historical financial product information of a customer to generate a recommendation set
2014	Education [148]	Appling the algorithm to students' synthesis marks evaluation system and analyzing the related factors that influence students' synthesis marks to help educators improve teaching methods and the quality of talent training

one-dimensional attributes at the same time.

Efficient mining algorithms have been developed for 2- and *m*-dimensional contexts to extend the traditional association rule framework from intra-transactional associations to inter-transactional associations in order to enforce its predictive capability in applications such as weather forecasting [144]. Multidimensional association is also used in the problem of modeling security in data warehouses by defining authorization rules for users and objects, and assigning sensitive information rules to the main elements of a multidimensional model [145]. It can also be used to analyze website promotion by discovering user access patterns from web log data [146]. For the complexity and diversity of personal financial services, multidimensional association rule mining is used to generate data cubes and recommendation sets before integrating recommendation weights in the recommendation process [147]. Similarity, an improved algorithm of association rule mining based on multidimensional sets can be used to find the maximally frequent item sets for each dimensional subset and to prune the database. This leads to the reduction of workload for the subsequent mining [148]. Some of the applications of multidimensional association rules are highlighted in **Table 21**.

3) Quantitative Association Rules

Quantitative association rules (QARM) refer to a special type of association rules in the form of $X \Rightarrow Y$, with X and Y consisting of a set of numerical and/or categorical attributes. They differ from general association rules, in which both the left- and right-hand sides of the rule should be categorical (nominal or discrete) attributes, because at least one attribute of the quantitative association rule (left or right) must involve a numerical attribute [149].

Table 22. Quantitative applications.

Year	Application	Description
2005	Dense regions [151]	Capturing the characteristics of quantitative attributes
2011	Biological [152]	Defining the interrelationship between genes
2014	Satellites [153]	Mining all quantitative association rules to improve satellite on-orbit performance analysis

QARM research can be categorized into six approaches [150]: 1) Partitioning: the majority of the research, working to solve the reasonable range or focusing on interval detection; 2) Clustering: variation of density to generate intervals for mining, reduction of the support-confidence conflict and elimination of useless or redundant rules; 3) Statistical: researchers consider the distribution of continuous data via standard statistical measures to analyse distributions or to determine reasonable intervals for frequent itemset generation; 4) Fuzzy: focusing on interval detection of attributes, which may not be precise or meaningful for analysts, to unleash non-trivial knowledge with ease; a challenge is to optimize the selection of partition points without suffering from the Support-Confidence conflict if the thresholds are not wisely chosen; 5) Evolutionary: the use of multi-objective evolutionary algorithms to optimize QAR mining so as to reduce the cost of mining and optimize the number of positive and negative QARs generated without compromising rule interestingness; 6) Other: presentation of particular techniques to improve QAR mining; min-support and min-confidence thresholds to find relevant rules are prone to Support-Confidence conflict. Highlights of QAR applications are shown in **Table 22**.

3.2.3. Sequence Discovery

Sequential pattern discovery is used to identify patterns of ordered events within a database [154]. It can also take advantage of the elapsed time between transactions that make up the event.

Recently, research in healthcare has focused on discovering, classifying and visualizing frequent patterns among patient paths. In education, sequential pattern discovery is usually used to extract patterns from student interaction data. To mine web usage, this method usually analyses how websites are being used and sometimes focuses on transaction identification from web usage data. Sequential pattern discovery is additionally used in text mining to discover trends for text categorization, document classification and authorship identification. This technique is applied in bioinformatics in order to predict rules for organization of certain elements in genes, protein function prediction, and gene expression analysis. **Table 23** presents some of the applications of sequence discovery.

Table 23. Sequence discovery applications.

Year	Application	Description
1998	Biology [155]	Discovering microbial genome sequences
2002	Biochemistry [156]	Discovering a common signature of Diacylglyceride kinases, sphingosine kinases, NAD kinases
2008	Biomedical [157]	Extracting protein sequences from the protein databank and classifying proteins in the appropriate fold category
2014	Biomedical [154]	Identifying temporal relationships between drugs and predicting the next medications to be prescribed for a patient

4. Data Mining for Big Data

Big data is a collection of massive and complex datasets and data volumes that include the huge quantities of data, data management capabilities, social media analytics and real-time data. There are three main characteristics of big data—the 3Vs—as follow [158, 159].

a) Volume:

Big data is a large amount of archived data that organizations are trying to harness to improve decision-making across the enterprise. Data volumes are increasing at an unpredictable rate. The size could be in the petabytes.

b) Variety

Big data comes from different types of data and data sources. It can include structured data, semi-structured data and unstructured data. It is also not always easy to manage big data into a relational database. This increases the complexity of both storing and analyzing big data.

c) Velocity

Velocity refers to the speed of data generation. More data is created and provided to users as a constant stream of real-time queries for meaningful information to be served up on demand, rather than in batches.

Organizations also require the two additional characteristics that follow to make big data analysis work [160, 161].

d) Value

Value refers to meaningful insights that deliver predictive analytics for future trends and patterns from deep, complex analysis based on machine learning, statistical modeling, and graph algorithms. These analytics go beyond the results of traditional business intelligence querying and reporting and give an organization a compelling advantage.

e) Veracity

The quality of the data can vary greatly. It is possible that the data includes defects, such as conflicts, missing data and outliers. Thus, the accuracy of data analysis depends on the data veracity.

Table 24. Differences between data mining and big data.

Areas	Data Mining	Big Data
Concept	This refers to techniques to discover the relevant information and interesting patterns in datasets; the datasets could be small files or large volumes of data.	This is a term referring to large scale storage and processing of large datasets; the dataset grows faster than a simple database and previous data handling architectures.
Value	It is tools and techniques that employ valuable results for decision maker.	It is an asset.
Processing	Processing refers to the operation that involves relatively sophisticated search operations.	Processing varies depending on the capabilities of the organization managing it and applications to process and analyses the data.
Task	Not all data mining tasks deal with big data.	All big data tasks involve data mining.

Data mining is a technique for discovering interesting patterns as well as descriptive, understandable models from large-scale data. Data mining and big data relate to the use of large data sets to handle the collection or reporting of data that serves businesses or other recipients. However, there are difference between data mining and big data as shown in **Table 24** [159, 162].

5. Conclusions

This paper presents a review summary of the data mining techniques and applications appearing from 1996 to 2016. The purpose of data mining techniques is to discover meaningful information or knowledge from previously collected data and build an adaptive computer system to handle specific issues. It is used in many fields such as biology, medicine, computer science and engineering. Some research has used data mining techniques to improve system performance.

Predictive methods and descriptive methods are the two major categories of data mining techniques. Predictive methods, such as decision trees, time series and regression methods, have been popular in science, medicine and engineering.

Descriptive methods are often used in finance, marketing and database technologies. K-mean and fuzzy c-mean techniques are well-known methods for clustering in computer science, engineering and mathematical modeling. A lot of medical research uses sequence discovery to find relationships between genes, proteins and symptoms. Apriori algorithms are normally used to find associations between attributes in a database; however, it can be combined with other techniques such as grid technologies to improve system performance.

Big data has been a hot issue recently. This paper provides a comparison between data mining and big data. It can be seen that big data refers to a large volume of data that is complex and unstructured, while data mining is a tool or technique to deal with big data in order to retrieve meaningful knowledge from it.

References:

[1] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Database," AI Magazine, Vol.17, pp. 37-54, 1996.

[2] M. Kantardzic, "Data Mining: Concepts, Models, Methods and Algorithms," John Wiley & Sons, Inc., 2002.

[3] M. H. Dunham, "Data mining introductory and advanced topics," Pearson Education, Inc., 2003.

[4] G. C. Nsofor, "A Comparative Analysis of Predictive Data-Mining Techniques," Master of Science, The University of Tennessee, Knoxville, 2006.

[5] F. Gorunescu, "Data Mining Concepts, Models and Techniques," Vol.12, Springer-Verlag Berlin Heidelberg, 2011.

[6] M. M. Mottalib, M. Rokonuzzaman, and M. T. Habib, "Fabric defect classification with geometric features using Bayesian classifier," Int. Conf. on Advances in Electrical Engineering 2015 (ICAEE), 2015.

[7] H. Yamagishi, H. Kawakami, T. Horiuchi, and a. O. Katai, "Reinforcement Leaning of Fuzzy Control Rules with Context-Specitic Segmentation of Actions," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.6, pp. 19-24, 2002.

[8] N. Takagi, "An Application of Binary Decision Trees to Pattern Recognition," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.10, No.5, pp. 682-687, 2006.

[9] S. T. Chen, S. L. Dou, and W. J. Chen, "A Data Mining Approach to Rainfall Intensity Classification Using TRMM/TMI Data," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.12, pp. 516-522, 2008.

[10] A. Stassopoulou, M. Petrou, and J. Kittler, "Bayesian and neural networks for geographic information processing," Pattern Recognition Letters, Vol.17, pp. 1325-1330, 1996.

[11] S. Lavington, N. Dewhurst, and A. F. E. Wilkins, "Interfacing knowledge discovery algorithms to large database management systems," Information and Software Technology, Vol.41, pp. 605-617, 1999.

[12] A. Rauber and J. Paralic, "Cluster Analysis as a First Step in the Knowledge Discovery Process," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.4, pp. 258-262, 2000.

[13] T. Tamaki, "Human Limb Extraction Based on Motion Estimation Using Optical Flow and Image Registration," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.8, pp. 150-155, 2004.

[14] P. Valenti, E. Cazamajou, M. Scarpettini, A. Aizemberg, W. Silva, and S. Kochen, "Automatic detection of interictal spikes using data mining models," Neuroscience Methods, Vol.150, pp. 105-110, 2006.

[15] J. Kazmierska and J. Malicki, "Application of the Naïve Bayesian Classifier to optimize treatment decisions," Radiotherapy and Oncology, Vol.86, pp. 211-216, 2008.

[16] J. Alonso-Montesinos, M. Martínez-Durbán, J. d. Sagrado, I. M. d. Águila, and F. J. Batlles, "The application of Bayesian network classifiers to cloud classification in satellite images," Renewable Energy, Vol.97, pp. 155-161, 2016.

[17] Q. Li, F. Li, and K. Doi, "Computerized Detection of Lung Nodules in Thin-Section CT Images by Use of Selective Enhancement Filters and an Automated Rule-Based Classifier," Academic Radiology, Vol.15, pp. 165-175, 2008.

[18] S. Mabu, M. Obayashi, and T. Kuremoto, "Ensemble learning of rule-based evolutionary algorithm using multi-layer perceptron for supporting decisions in stock trading problems," Applied Soft Computing, Vol.36, pp. 357-367, 2015.

[19] S. Mabu, W. Li, and K. Hirasawa, "A Class Association Rule Based Classifier Using Probability Density Functions for Intrusion Detection Systems," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.19, pp. 555-566, 2015.

[20] D. M. Farid, M. A. Al-Mamun, B. Manderick, and A. Nowe, "An adaptive rule-based classifier for mining big biological data," Expert Systems with Applications, Vol.64, pp. 305-316, 2016.

[21] V. S. R. P. V. Kamadi, A. R. Allam, S. M. Thummala, and V. N. R. P., "A computational intelligence technique for the effective diagnosis of diabetic patients using principal component analysis (PCA) and modified fuzzy SLIQ decision tree approach," Applied Soft Computing, Vol.49, pp. 137-145, 2016.

[22] L. O. Hall and P. Lande, "Generation of Fuzzy Rules from Decision Trees," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.2, pp. 128-133, 1998.

[23] O. Matsumoto, K. J. Mackin, and E. Tazaki, "Emergence of Learning Rule in Neural Networks Using Genetic Programming Combined with Decision Trees," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.3, pp. 223-233, 1999.

[24] M. T. a. N. Muranaka, "Product-Impression Analysis Using Fuzzy C4.5 Decision Tree," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.13, pp. 731-737, 2009.

[25] H. E. Osman, "Averaging Forest for Online Vision," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.13, pp. 400-406, 2009.

[26] K. M. Lee, K. S. Hwang, K. M. Lee, S. K. Han, W. H. Jung, and S. Lee, "Supervised Learning-Based Feature Selection for Mondrian Paintings Style Authentication," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.16, pp. 894-899, 2012.

[27] J. Kozak and U. Boryczka, "Collective data mining in the ant colony decision tree approach," Information Sciences, Vol.372, pp. 126-147, 2016.

[28] K. Kim, "A hybrid classification algorithm by subspace partitioning through semi-supervised decision tree," Pattern Recognition, Vol.60, pp. 157-163, 2016.

[29] M. Puri, A. Solanki, T. Padawer, S. M. Tipparaju, W. A. Moreno, and Y. Pathak, "Chapter 1 - Introduction to Artificial Neural Network (ANN) as a Predictive Tool for Drug Design, Discovery, Delivery, and Disposition: Basic Concepts and Modeling," Artificial Neural Network for Drug Design, Delivery and Disposition, pp. 3-13, 2016.

[30] P. Baranyi, L. T. Koczy, and T. D. Gedeon, "Improved Fuzzy and Neural Network Algorithms for Word Frequency Prediction in Document Filtering," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.2, pp. 88-95, 1998.

[31] R. Dhaouadi and K. Nouri, "Neural Network-Based Speed Control of A Two-Mass-Model System," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.3, pp. 427-430, 1999.

[32] J. Lu and H. Ohta, "A Powerful Neural Network Method with Digital-contract Hints for Pricing Complex Options," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.7, pp. 139-146, 2003.

[33] N. Watanabe and S. Ishizaki, "Neural Network Model for Word Sense Disambiguation Using Up/Down State and Morphoelectrotonic Transform," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.11, pp. 780-786, 2007.

[34] T. Kondo, J. Ueno, and S. Takao, "Medical Image Diagnosis of Liver Cancer Using a Neural Network and Artificial Intelligence," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.15, pp. 714-722, 2011.

[35] K. G. Abistado, C. N. Arellano, and E. A. Maravillas, "Weather Forecasting Using Artificial Neural Network and Bayesian Network," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.18, pp. 812-817, 2014.

[36] S. R. Gunn, "Support Vector Machines for Classification and Regression," University of Southampton, 1998.

[37] M. Kantardzic, B. Djulbegovic, and H. Hamdan, "A data-mining approach to improving Polycythemia Vera diagnosis," Computers & Industrial Engineering, Vol.43, pp. 765-773, 2002.

[38] S. F. Crone, S. Lessmann, and R. Stahlbock, "The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing," European J. of Operational Research, Vol.173, pp. 781-800, 2006.

[39] C. C. Lee, S. H. Chen, and Y. C. Chiang, "Classification of Liver Disease from CT Images Using a Support Vector Machine," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.11, pp. 396-402, 2007.

[40] B. Li, J. Hu, and K. Hirasawa, "Support Vector Machine Classifier with WHM Offset for Unbalanced Data," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.12, pp. 94-101, 2008.

[41] B. L. Qiangwei Wang, and Jinglu Hu, "Human Resource Selection Based on Performance ClassificationUsingWeighted Support Vector Machine," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.13, pp. 407-415, 2009.

[42] K. Shimada, R. Muto, and T. Endo, "A Combined Method Based on SVM and Online Learning with HOG for Hand Shape Recognition," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.16, pp. 687-695, 2012.

[43] "Stock Market Trend Prediction Based on Text Mining of Corporate Web and Time Series Data," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.18, pp. 22-31, 2014.

[44] J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques," ELSEVIER SCIENCE & TECHNOLOGY, 2011.

[45] H. He, S. Hawkins, W. Graco, and X. Yao, "Application of Genetic Algorithm and K-Nearest Neighbour Method in Real World Medical Fraud Detection Problem," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.4, pp. 130-137, 2000.

[46] M. Y. Su, "Real-time anomaly detection systems for Denial-of-Service attacks by weighted k-nearest-neighbor classifiers," Expert Systems with Applications, Vol.38, pp. 3492-3498, 2011.

[47] J. Maillo, S. Ramírez, I. Triguero, and F. Herrera, "kNN-IS: An Iterative Spark-based design of the k-Nearest Neighbors classifier for big data," Knowledge-Based Systems, pp. 1-13, 2016.

[48] D. Wang, "K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: Revisited," Mechanical Systems and Signal Processing, Vol.70-71, pp. 201-208, 2016.

[49] K. Hirota, H. Yoshino, M. Q. Xu, Y. Zhu, X. Y. Li, and D. Horie, "An Application of Fuzzy Theory to the Case-Based Reasoning of the CISG," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.1, pp. 86-93, 1997.

[50] N. H. Phuong, N. B. Tu, and A. Ohsato, "Developing Case-based Reasoning System for Medical Consultation Using the Importance of Features," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.6, pp. 41-50, 2002.

[51] T. Ahola and K. Leiviskä, "Case-Based Reasoning in Web Break Sensitivity Evaluation in a Paper Machine," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.9, pp. 556-561, 2005.

[52] C. H. L. Lee, A. Liu, and H. H. Huang, "Using Planning and Case-Based Reasoning for Service Composition," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.14, pp. 540-548, 2010.

[53] Y. Shen, J. Colloc, A. Jacquet-Andrieu, and K. Lei, "Emerging medical informatics with case-based reasoning for aiding clinical decision in multi-agent system," J. of Biomedical Informatics, Vol.56, pp. 307-317, 2015.

[54] F. Sartori, A. Mazzucchelli, and A. D. Gregorio, "Bankruptcy forecasting using case-based reasoning: The CRePERIE approach," Expert Systems with Applications," Vol.64, pp. 400-411, 2016.

[55] T. Miyajima and T. Koitabashi, "Placement Time Optimization of Chip Mounter by Genetic Algorithms – Search for Optimal Tape Feeder Arrangement –," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.2, pp. 160-163, 1998.

[56] M. Watanabe, H. Nobuhara, K. Kawamoto, F. Dong, and K. Hirota, "Chain Restaurant Work Scheduling Based on Genetic Algorithm with Fuzzy Logic," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.10, pp. 50-59, 2006.

[57] C. W. Han and H. Nobuhara, "Advanced Genetic Algorithms Based on Adaptive Partitioning Method," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.11, pp. 677-680, 2007.

[58] T. Inaba, L. He, K. Suzuki, K. Murakami, and Y. Chao, "A Genetic-Algorithm-Based Temporal Subtraction for Chest Radiographs," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.13, pp. 289-296, 2009.

[59] E. P. Ijjina and K. M. Chalavadi, "Human action recognition using genetic algorithms and convolutional neural networks," Pattern Recognition, Vol.59, pp. 199-212, 2016.

[60] P. Mahajan, R. Kandwal, and R. Vijay, "Rough Set Approach in Machine Learning: A Review.," J. of Computer Applications, Vol.56, pp. 1-13, 2012.

[61] P. Pattaraintakorn and N. Cercone, "Integrating rough set theory and medical applications," Applied Mathematics Letters, Vol.21, pp. 400-403, 2008.

[62] I. Jagielska, "Using Rough Sets for Practical Feature Selection in a Rough Sets/Neural Network Framework for Knowledge Discovery," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.4, pp. 31-38, 2000.

[63] W. Jin-Mao, W. Shu-Qin, and W. Ming-Yang, "Novel Approach to Decision-Tree Construction," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.8, p. JACIIIVol.8 No.3pp. , 2004.

[64] H. Sakai, K. Koba, and M. Nakata, "Rough Sets Based Rule Generation from Data with Categorical and Numerical Values," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.12, pp. 426-434, 2008.

[65] A. Kothari and A. Keskar, "Rough Set Approach for Overall Performance Improvement of an Unsupervised ANN-Based Pattern Classifier," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.13, pp. 434-440, 2009.

[66] Y. Matsumoto and J. Watada, "Rough Sets Based Prediction Model of Tick-Wise Price Fluctuations," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.15, pp. 449-453, 2011.

[67] D. V. Nguyen, K. Yamada, and M. Unehara, "Rough Set Approach with Imperfect Data Based on Dempster-Shafer Theory," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.18, pp. 280-288, 2014.

[68] S. Park and H. Lee-Kwang, "Type-2 Fuzzy Hypergraphs Using Type-2 Fuzzy Sets," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.4, pp. 362-367, 2000.

[69] G. H. Tzeng, C. M. Feng, and C. C. Kang, "Application of Fuzzy Set Theory and DEA Model to Evaluating Production Efficiency for Taipei City Bus Company," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.5, pp. 128-138, 2001.

[70] H. Sekiya, T. Kondo, M. Hashimoto, and T. Takagi, "Dynamic Sense Representation Using Conceptual Fuzzy Sets," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.10, pp. 859-867, 2006.

[71] K. Uehara and K. Hirota, "Multi-Level Control of Fuzzy-Constraint Propagation via Evaluations with Linguistic Truth Values in Generalized-Mean-Based Inference," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.20, pp. 355-377, 2016.

[72] R. A. Yaffee, "Robust Regression Analysis:Some Popular Statistical Package Options," Statistics, Social Science, and Mapping Group Academic Computing Services Information Technology Services, 2002.

[73] T. C. Fu, "A review on time series data mining," Engineering Applications of Artificial Intelligence, Vol.24, pp. 164-181, 2011.

[74] A. Serletis and D. Krause, "Empirical evidence on the long-run neutrality hypothesis using low-frequency international data," Economics Letters, Vol.50, pp. 323-327, 1996.

[75] R. E. Abdel-Aal and A. M. Mangoud, "Modeling and forecasting monthly patient volume at a primary health care clinic using univariate time-series analysis," Computer Methods and Programs in Biomedicine, Vol.56, pp. 235-247, 1998.

[76] J. C. Cuaresma, J. Hlouskova, S. Kossmeier, and M. Obersteiner, "Forecasting electricity spot-prices using linear univariate time-series models," Applied Energy, Vol.77, pp. 87-106, 2004.

[77] J. W. Taylor and R. Buizza, "Density forecasting for weather derivative pricing," Int. J. of Forecasting, Vol.22, pp. 29-42, 2006.

[78] S. Chattopadhyay and G. Chattopadhyay, "Univariate modelling of summer-monsoon rainfall time series: Comparison between ARIMA and ARNN," Comptes Rendus Geoscience, Vol.342, pp. 100-107, 2010.

[79] M. Maric, E. d. Haan, S. M. Hogendoorn, L. H. Wolters, and H. M. Huizenga, "Evaluating Statistical and Clinical Significance of Intervention Effects in Single-Case Experimental Designs: An SPSS Method to Analyze Univariate Data," Behavior Therapy, Vol.46, pp. 230-241, 2015.

[80] Y. Yin and P. Shang, "Forecasting traffic time series with multivariate predicting method," Applied Mathematics and Computation, Vol.291, pp. 266-278, 2016.

[81] Z. Li and M. Kafatos, "Interannual Variability of Vegetation in the United States and Its Relation to El Niño/Southern Oscillation," Remote Sensing of Environment, Vol.71, pp. 239-247, 2000.

[82] X. Chen and X. Chen, "Sensitivity analysis and determination of streambed leakance and aquifer hydraulic properties," J. of Hydrology, Vol.284, pp. 270-284, 2003.

[83] A. Müller, H. Osterhage, R. Sowa, R. G. Andrzejak, F. Mormann, and K. Lehnertz, "A distributed computing system for multivariate time series analyses of multichannel neurophysiological data," J. of Neuroscience Methods, Vol.152, pp. 190-201, 2006.

[84] F. Duchêne, C. Garbay, and Vincent Rialle, "Learning recurrent behaviors from heterogeneous multivariate time-series," Artificial Intelligence in Medicine, Vol.39, pp. 25-47, 2007.

[85] K. C. Perera, A. W. Western, B. George, and B. Nawarathna, "Multivariate time series modeling of short-term system scale irrigation demand," J. of Hydrology, Vol.531, pp. 1003-1019, 2015.

[86] A. Karahoca, D. Karahoca, and M. Sanver, "Survey of Data Mining and Applications (Review from 1996 to Now)," Data Mining Applications in Engineering and Medicine, InTech, 2012.

[87] P. L. Gozalo and O. B. Linton, "Testing additivity in generalized nonparametric regression models with estimated parameters," J. of Econometrics, Vol.104, pp. 1-48, 2001.

[88] S. Kim, S. Imoto, and S. Miyano, "Dynamic Bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data," Biosystems, Vol.75, pp. 57-65, 2004.

[89] S. Efromovich, "Nonparametric regression with responses missing at random," J. of Statistical Planning and Inference, Vol.141, pp. 3744-3752, 2011.

[90] N. Yampikulsakul, E. Byon, S. Huang, S. Sheng, and M. You, "Condition Monitoring of Wind Power System With Nonparametric Regression Analysis," IEEE Trans. on Energy Conversion, Vol.29, pp. 288-299, 2014.

[91] N. N. Kachouie, T. Gerke, P. Huybers, and A. Schwartzman, "Nonparametric Regression for Estimation of Spatiotemporal Mountain Glacier Retreat From Satellite Images," IEEE Trans. on Geoscience and Remote Sensing, Vol.53, pp. 1135-1149, 2015.

[92] Z. Pan, Y. Wang, and C. Wu, "A nonparametric approach to test for predictability," Economics Letters, Vol.148, pp. 10-16, 2016.

[93] T. H. Soukissian and F. E. Karathanasi, "On the use of robust regression methods in wind speed assessment," Renewable Energy, Vol.99, pp. 1287-1298, 2016.

[94] K. A. Weigel and S. W. Lin, "Development of Int. Conversion Equations Using Robust Regression Methodology," J. of Dairy Science, Vol.82, pp. 2023-2029, 1999.

[95] A. C. Atkinson and T. C. Cheng, "On robust linear regression with incomplete data," Computational Statistics & Data Analysis, Vol.33, pp. 361-380, 2000.

[96] D. E. Booth and K. Lee, "Robust regression-based analysis of drug-nucleic acid binding," Analytical Biochemistry, Vol.319, pp. 258-262, 2003.

[97] M. C. Ortiz, L. A. Sarabia, and A. Herrero, "Robust regression techniques: A useful alternative for the detection of outlier data in chemical analysis," Talanta, Vol.70, pp. 499-512, 2006.

[98] V. Fritsch, B. D. Mota, G. V. Eva Loth, T. Banaschewski, G. J. Barker, A. L. W. Bokde, et al., "Robust regression for large-scale neuroimaging studies," NeuroImage, Vol.111, pp. 431-441, 2015.

[99] Andy Leung, H. Zhang, and R. Zamar, "Robust regression estimation and inference in the presence of cellwise and casewise contamination," Computational Statistics & Data Analysis, Vol.99, pp. 1-11, 2016.

[100] R. J. Holzworth, "Policy Capturing with Ridge Regression," Organizational Behavior and Human Decision Processes, Vol.68, pp. 171-179, 1996.

[101] J. Huang, D. Brennan, L. Sattler, J. Alderman, B. Lane, and C. O'Mathuna, "A comparison of calibration methods based on calibration data size and robustness," Chemometrics and Intelligent Laboratory Systems, Vol.62, pp. 25-35, 2002.

[102] L. M. Grosso, E. W. Triche, K. Belanger, N. L. Benowitz, T. R. Holford, and M. B. Bracken, "Association of caffeine metabolites in umbilical cord blood with IUGR and preterm delivery: A prospective cohort study of 1609 pregnancies," Annals of Epidemiology, Vol.115, pp. 659-660, 2005.

[103] X. Yu and S. Y. Liong, "Forecasting of hydrologic time series with ridge regression in feature space," J. of Hydrology, Vol.332, pp. 290-302, 2007.

[104] P. Mohapatra, S. Chakravarty, and P. K. Dash, "Microarray medical data classification using kernel ridge regression and modified cat swarm optimization based gene selection system," Swarm and Evolutionary Computation, Vol.20, pp. 144-160, 2016.

[105] K. Jin, S. Lee, S. Lee, and G. Heo, "Development of nuclear forensic models using kernel regression," Progress in Nuclear Energy, Vol.94, pp. 55-63, 2017.

[106] G. Y. Heo, "Condition monitoring using empirical models: technical review and prospects for nuclear applications," Nuclear Engineering and Technology, Vol.40, pp. 49-68, 2008.

[107] D. Lanari and C. T. E D'Agaro, "Use of nonlinear regression to evaluate the effects of phytase enzyme treatment of plant protein diets for rainbow trout (Oncorhynchus mykiss)," Aquaculture, Vol.161, pp. 345-356, 1998.

[108] P. F. Schikora and M. R. Godfrey, "Efficacy of end-user neural network and data mining software for predicting complex system performance," Int. J. of Production Economics, Vol.84, pp. 231-253.

[109] M. Mohanty, D. K. Painuli, A. K. Misra, K. K. Bandyopadhyaya, and P. K. Ghosh, "Estimating impact of puddling, tillage and residue management on wheat (Triticum aestivum, L.) seedling emergence and growth in a rice–wheat system using nonlinear regression models," Soil and Tillage Research, Vol.87, pp. 119-130, 2006.

[110] D. Kimura, M. Nii, T. Yamaguchi, Y. Takahashi, and T. Yumoto, "Fuzzy Nonlinear Regression Analysis Using Fuzzified Neural Networks for Fault Diagnosis of Chemical Plants," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.15, pp. 336-344, 2011.

[111] N. Deshpande, S. Londhe, and S. Kulkarni, "Modeling compressive strength of recycled aggregate concrete by Artificial Neural Network, Model Tree and Non-linear Regression," Int. J. of Sustainable Built Environment, Vol.3, pp. 187-198, 2014.

[112] S. Tomlinson, "Novel approaches to the calculation and comparison of thermoregulatory parameters: Non-linear regression of metabolic rate and evaporative water loss in Australian rodents," J. of Thermal Biology, Vol.57, pp. 54-65, 2016.

[113] G. E. Hinton, "Learning multiple layers of representation," TRENDS in Cognitive Sciences, Vol.11, pp. 428-434, 2007.

[114] S. Raschka, "What is the Difference Between Deep Learning and "Regular" Machine Learning?" Available: http://www.kdnuggets.com/2016/06/difference-between-deep-learning-regular-machine-learning.html

[115] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, Vol.521, pp. 436-444, 2015.

[116] E. d. l. Rosa and W. Yu, "Randomized algorithms for nonlinear system identification with deep learning modification," Information Sciences, Vol.364-365, pp. 197-212, 2016.

[117] Soniya, S. Paul, and L. Singh, "A review on advances in deep learning," IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI), 2015.

[118] S. Curtis, "What To Expect from Deep Learning in 2016 and Beyond," http://www.kdnuggets.com/2016/01/deep-learning-2016-beyond.html

[119] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification," Neural Networks, Vol.32, pp. 333-338, 2012.

[120] K. Noda, H. Arie, Y. Suga, and T. Ogata, "Multimodal integration learning of robot behavior using deep neural networks," Robotics and Autonomous Systems, Vol.62, pp. 721-736, 2014.

[121] Y. Zhou, Q. Hu, J. Liu, and Y. Jia, "Combining heterogeneous deep neural networks with conditional random fields for Chinese dialogue act recognition," Neurocomputing, Vol.168, pp. 408-417, 2016.

[122] E. P. Ijjina and C. K. Mohan, "Hybrid deep neural network model for human action recognition," Applied Soft Computing, Vol.46, pp. 936-952, 2016.

[123] R. Kumari, Sheetanshu, M. K. Singh, R. Jha, and N. K. Singh, "Anomaly detection in network traffic using *K*-mean clustering," The 3rd Int. Conf. on Recent Advances in Information Technology 2016 (RAIT), 2016.

[124] C. W. Chen, J. Luo, and K. J. Parker, "Image segmentation via adaptive *K*-mean clustering and knowledge-based morphological operations with biomedical applications," IEEE Trans. on Image Processing, Vol.7, pp. 1673-1683, 1998.

[125] Y. M. Marzouk and Ahmed F. Ghoniem, "*K*-means clustering for optimal partitioning and dynamic load balancing of parallel hierarchical N-body simulations," J. of Computational Physics, Vol.207, pp. 493-528, 2005.

[126] S. C. Babu and P. Sanyal, "Chapter 13 – Classifying households on food security and poverty dimensions – application of *K*-mean cluster analysis," Food Security, Poverty and Nutrition Policy Analysi, pp. 265-277, Academic Press, 2009.

[127] L. H. Juang and M. N. Wu, "MRI brain lesion image detection based on color-converted *K*-means clustering segmentation," Measurement, Vol.43, pp. 941-949, 2010.

[128] Y. Ning, X. Zhu, S. Zhu, and Y. Zhang, "Surface EMG Decomposition Based on *K*-means Clustering and Convolution Kernel Compensation," IEEE J. of Biomedical and Health Informatics, Vol.19, pp. 471-477, 2015.

[129] O. Kesemen, Ö. Tezel, and E. Özkul, "Fuzzy c-means clustering algorithm for directional data (FCM4DD)," Expert Systems with Applications, Vol.58, pp. 76-82, 2016.

[130] J. Nayak, B. Naik, and H. S. Behera, "Fuzzy C-Means (FCM) ClusteringAlgorithm: A Decade Reviewfrom 2000 to 2014," Computational Intelligence in Data Mining, Vol.2, pp. 133-149, 2015.

[131] M. Gil, E. G. Sarabia, J. R. Llata, and J. P. Oria, "Fuzzy c-means clustering for noise reduction, enhancement and reconstruction of 3D ultrasonic images," 7th IEEE Int. Conf. on Emerging Technologies and Factory Automation, 1999.

[132] M. Hanesch, R. S. a, and M. J. D. b, "The application of fuzzy C-means cluster analysis and non-linear mapping to a soil data set for the detection of polluted sites," Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy, Vol.26, pp. 885-891, 2001.

[133] M. Maitra and A. Chatterjee, "Hybrid multiresolution Slantlet transform and fuzzy c-means clustering approach for normal-pathological brain MR image segregation," Medical Engineering & Physics, Vol.30, pp. 615-623, 2008.

[134] Z. Xue, Y. Shang, and A. Feng, "Semi-supervised outlier detection based on fuzzy rough C-means clustering," Mathematics and Computers in Simulation, Vol.80, pp. 1911-1921, 2010.

[135] M. L. a, M. F. b, and G. Haiyan, "The E-commerce Risk Early-Warning Model Based on the Unascertained C-means Clustering," Procedia Engineering, Vol.15, pp. 4740-4744, 2011.

[136] A. M. Abdulshahed, A. P. Longstaff, S. Fletcher, and A. Myers, "Thermal error modelling of machine tools based on ANFIS with fuzzy c-means clustering using a thermal imaging camera," Applied Mathematical Modelling, Vol.39, pp. 1837-1852, 2015.

[137] O. Kesemen, Ö. Tezel, and E. Özkul, "Fuzzy c-means clustering algorithm for directional data (FCM4DD)," Expert Systems with Applications, Vol.58, pp. 76-82, 2016.

[138] T. P. Hong, C. S. Kuo, and S. L. Wang, "A fuzzy AprioriTid mining algorithm with reduced computational time," Applied Soft Computing, Vol.5, pp. 1-10, 2004.

[139] K. H. Kaoru Shimada, and Jinglu Hu, "Alternate Genetic Network Programming with Association Rules Acquisition Mechanisms Between Attribute Families," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.10, pp. 954-963, 2006.

[140] B. C. Chien, M. H. Zhong, and J. J. Wang, "Mining Fuzzy Association Rules on Has-A and Is-A Hierarchical Structures," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.11, pp. 423-432, 2007.

[141] C. Aflori and M. Craus, "Grid implementation of the Apriori algorithm," Advances in Engineering Software, Vol.38, pp. 295-300, 2006.

[142] L. Hanguang and N. Yu, "Intrusion Detection Technology Research Based on Apriori Algorithm," Physics Procedia, Vol.24, pp. 1615-1620, 2012.

[143] W. Renli, D. Yueming, and D. Liming, "The application of Apriori-BSO algorithms in medical records data mining," IEEE Information Technology, Networking, Electronic and Automation Control Conf. 2016, 2016.

[144] L. Feng, T. Dillon, and J. Liu, "Inter-transactional association rules for multi-dimensional contexts for prediction and their application to studying meteorological data," Data & Knowledge Engineering, Vol.37, pp. 85-115, 2001.

[145] E. Fernández-Medina, J. Trujillo, R. Villarroel, and M. Piattini, "Access control and audit model for the multidimensional modeling of data warehouses," Decision Support Systems, Vol.42, pp. 1270-1289, 2006.

[146] J. Yan, L. Zhang, F. Zhao, W. Liu, and K. Li, "Multidimensional Association Analysis of Web Users' Access Path for Website Design and Promotion," Sixth Int. Conf. on Fuzzy Systems and Knowledge Discovery, 2009.

[147] G. Liu, H. Jiang, R. Geng, and H. Li, "Application of multidimensional association rules in personal financial services," Int. Conf. on Computer Design and Applications (ICCDA), 2010.

[148] Y. L. Y. Zou, X. Qin, and S. Ma, "Research and application of association rule mining algorithm based on multidimensional sets," 5th IEEE Int. Conf. on Software Engineering and Service Science (ICSESS), Beijing, 2014.

[149] X. Zhu, "Quantitative Association Rules," Encyclopedia of Database Systems, pp. 2240-2244, 2009.

[150] D. Adhikary and S. Roy, "Trends in Quantitative Association Rule Mining," IEEE 2nd Int. Conf. on Recent Trends in Information Systems (ReTIS), 2015.

[151] W. Lian, D. W. Cheung, and S. M. Yiu, "An efficient algorithm for finding dense regions for mining quantitative association rules," Computers & Mathematics with Applications, Vol.50, pp. 471-490, 2005.

[152] M. Martínez-Ballesteros, I. Nepomuceno-Chamorro, and J. C. Riquelme, "Inferring gene-gene associations from Quantitative Association Rules," 11th Int. Conf. on Intelligent Systems Design and Applications (ISDA), 2011.

[153] X. Dong and D. Pi, "An Effective Method for Mining Quantitative Association Rules with Clustering Partition in Satellite Telemetry Data," Second Int. Conf. on Advanced Cloud and Big Data (CBD), 2014.

[154] A. P. Wright, A. T. Wright, A. B. McCoy, and D. F. Sittig, "The use of sequential pattern mining to predict next prescribed medications," J. of Biomedical Informatics, Vol.53, pp. 73-80, 2015.

[155] M. B. Schmid, "Novel approaches to the discovery of antimicrobial agents," Novel approaches to the discovery of antimicrobial agents, Vol.2, pp. 529-534, 1998.

[156] G. Labesse, D. Douguet, L. Assairi, and A. M. Gilles, "Diacylglyceride kinases, sphingosine kinases and NAD kinases: distant relatives of 6-phosphofructokinases," Trends in Biochemical Sciences, Vol.27, pp. 273-275, 2002.

[157] T. P. Exarchos, C. Papaloukas, C. Lampros, and D. I. Fotiadis, "Mining sequential patterns for protein fold recognition," J. of Biomedical Informatics, Vol.41, pp. 165-179, 2008.

[158] Shilpa and M. Kaur, "BIG Data and Methodology-A review," Int. J. of Advanced Research in Computer Science and Software Engineering, Vol.3, pp. 991-995, 2013.

[159] B. Thakur and M. Mann, "Data Mining for Big Data: A Review," Int. J. of Advanced Research in Computer Science and Software Engineering, Vol.4, pp. 469-473, 2014.

[160] Ishwarappa and J. Anuradha, "A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology," Int. Conf. on Computer, Communication and Convergence (ICCC-2015), 2015.

[161] I. I. Center, "Vision Paper Distributed Data Mining and Big Data," Intel Corporation, pp. 1-14, 2012.

[162] V. Shobaba, S. Maheshwair, and M. Savithri, "Study on Big data with Data Mining," Int. J. of Advanced Research in Computer and Communication Engineering, Vol.4, pp. 381-383, 2015.

Name:
Ratchakoon Pruengkarn

Affiliation:
Candidate Ph.D. Student, School of Engineering and Information Technology

Address:
90 South St., Murdoch, WA, Australia

Brief Biographical History:
1996-2000 BS.C. Computer Science, Chiang Mai University
2001-2005 MS.C. Computer Science, Kasetsart University
2007-2014 Lecturer, Dhurakij Pundit University

Main Works:
- "Using Misclassification data to Improve Classification Performance," pp. 1-5, 2015.
- "Improving Classification Performance with Complementary Fuzzy-Based Neural Networks," pp. 1-5, 2015.

Membership in Academic Societies:
- The Institute of Electrical and Electronics Engineers (IEEE)

Name:
Kok Wai Wong

Affiliation:
Associate Professor, School of Engineering and Information Technology

Address:
90 South St., Murdoch, WA, Australia

Brief Biographical History:
1995 B.Eng. (Hons), Curtin University
2000 Ph.D., Curtin University
2001 Lecturer, Murdoch University
2003 Assistant Professor, Nanyang Technological University
2007 Associate Professor, Murdoch University

Main Works:
- P. Jeatrakul, K. Wong, and L. Fung, "Data cleaning for classification using misclassification analysis," J. Adv. Comput. Intell. Intell. Inform. (JACIII), Vol.14, No.3, pp. 297-302, 2010.
- K. W. Wong, D. Tikk, T. D. Gedeon, and L. T. Kóczy, "Fuzzy Rule Interpolation for Multidimensional Input Spaces with Applications," IEEE Trans. of Fuzzy Systems, Vol.13, No.6, pp. 809-819, 2005.
- K. W. Wong, T. D. Gedeon, and L. T. Kóczy, "Construction of Fuzzy Signature from Data: An Example of SARs Preclinical Diagnosis System," Proc. of IEEE Int. Conf. on Fuzzy Systems, FUZZ-IEEE 2004, pp. 1353, 2004.

Membership in Academic Societies:
- The Institute of Electrical and Electronics Engineers (IEEE), Senior Member
- Australia Computer Society
- Asia Pacific Neural Network Society

Name:
Chun Che Fung

Affiliation:
Emeritus Professor, School of Engineering and Information Technology

Address:
90 South St., Murdoch, WA, Australia

Brief Biographical History:
1981 B.Sc. (1st Hon, Maritime Tech.), University of Wales
1982 M.Eng. (System Test Engineering), University of Wales
1982-1988 Lecturer, Singapore Polytechnic
1989-2003 Lecturer and Senior Lecturer, Curtin University
1994 Ph.D., University of Western Australia
2003-2015 Associate Professor, Murdoch University
2015 Emeritus Professor, Murdoch University

Main Works:
- "Simulated annealing based economic dispatch algorithm," IEEE Proc. C, Vol.140, No.6, pp. 509-515, 1993.
- "How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach," Computers & Education Vol.55, No.4, pp. 1424-1442, 2010.
- "Distinguishing games and simulation games from simulators," Computers in Entertainment (CIE), Vol.4, No.2, pp. 9, 2006.

Membership in Academic Societies:
- The Institute of Electrical and Electronics Engineers (IEEE), Senior Member