Available online at http://ijim.srbiau.ac.ir/
Int. J. Industrial Mathematics (ISSN 2008-5621)
Vol. 10, No. 4, 2018 Article ID IJIM-1088, 10 pages
Research Article

Multi-Objective Model for Fair Pricing of Electricity Using the Parameters from the Iran Electricity Market Big Data Analysis

M. Salami *†, F. Movahedi Sobhani ‡, M. S. Ghazizadeh §

Received Date: 2017-06-09 Revised Date: 2017-12-12 Accepted Date: 2018-05-18

Abstract

Assessment of the electricity market shows that, electricity market data can be considered big data. this data has been analyzed by both conventional and modern data mining methods. The predicted variables of supply and demand are considered to be the input of a defined multi-objective for predicting electricity price, which is the result of the defined model. This shows the advantage of applying the new algorithms for big data mining.

Keywords : Big data; Electricity price; Electricity market; SVMGA algorithm; Multi-Objectives Model.

1 Introduction

THe global growth of data in different areas is substantial. Over 2.5 Exabyte of data are generated worldwide every day. This amount of data is rapidly expanding. Such extensive generation of data has resulted in a new concept called big data [1, 2]. There are different definitions of big data. For instance, Manyika believes that big data are those data which are too large for current software applications to analyze and store.According to the majority of researchers, the most important feature of big data is the volume [3, 4]. This massive volume of data cannot be analyzed with classic data mining methods, something which results in serious challenges on data mining quality and time [5]. These data are mostly non-structured and non-framed [7, 8]. Therefore, the institutes providing new platforms for the exploitation of big data can properly benefit from latent values of such data in decision-making and organizational processes [11, 12, 13]. Some of the most important applications include identifying procedures and making strategic and operational decisions. (It should be noted that making such decisions requires up-to-date and reliable information. However, necessary data are scattered in different information sources, including information systems, spreadsheet files, and even geographical data, so that it is complicated to access and analyze them.) Other applications include modeling patterns and risk management models. (The main risk management principle is to use up-to-date data in modeling, something which can only be possible through the use of big data analysis tools with regard to the massive volume and high speed of data generation.) Given the above-mentioned applications and the immediate need of organizations for data analysis to achieve agility, dynamism, and sustainability,

*Corresponding author. Mes.salami.eng@iauctb.ac.ir, Tel:+98(912)7252838.

†Department of Industrial Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran.

‡Department of Industrial Engineering, Science and Research Branch, Islamic Azad University, Tehran,Iran.

§Department of Electrical Engineering, Abbaspour School of Engineering, Shahid Beheshti University, Tehran, Iran.

it is essential to pay attention to big data and codify appropriate solutions for the exploitation of latent values in data [1, 2]. As a result, enterprises and institutions using data analysis in this area can gain competitive advantages and will benefit from more accurate prediction and foresight regarding different issues [14]. Irans electricity market was established in 2001. The mission of this market was to provide a system for the purchase and sale of electricity in supply and demand. This system was set up to determine the electricity price [18]. Electricity production, distribution, and transmission were managed unanimously in the old electricity industry structure. These sectors operate independently in the new structure [18]. In addition, the electricity market will act as an intermediary among the above-mentioned sectors. This market is regarded as one of the infrastructures required for privatization. Accordingly, the Energy Exchange began its work in early 2012, and subscribers are going to be able to purchase electricity directly from the Electricity Exchange in the foreseeable future [19]. The two most important variables of the electricity market are the price of electricity and the load of electricity at power plants. Using accurate and intelligent data mining methods to deal with energy production data and making correct decisions will result in huge savings in the costs of energy production and the accurate prediction of needs for energy as well as the presentation and codification of appropriate electricity price. Using proper data analysis to make correct decisions, the purchasers and buyers of this market gain advantages over their rivals. However, the analysis of data in this area is a serious challenge because of the features of big data [16, 17]. With an increase in new sources of data generation, a significant volume of data is created and stored every day, something which is observed to a greater or lesser degree in different industries. Many papers have investigated the application of big data in different applied areas, for example a multi-variate time series model for predicting electricity prices (Cuaresma et al. (2004)), paper on the provision of a neural network based model of effective parameters in predicting demand and electricity prices (Yamashita et al. (2004)), paper on the prediction of future electricity prices in the European countries using the neural network [22]. Another category of articles in this area is to present a variety of single or multi-purpose models with constraints designed to predict future electricity prices, such as, concerning modeling the price of electricity, taking into account randomized change [21], a model for forecasting future contracts with random parameters [9], or providing a model for epex forecasting taking into account renewable energy in the model (Florian et al. (2015)). Also studied the use of big data in health [11] and investigated the application of big data in the expansion of access to financial services in China [12]. This study investigates the main features of big data, such as volume and variety, in addition to the data stored in Irans electricity database to show that they are big data. Therefore, new algorithms of big data are required to analyze them, and classic methods are not efficient enough for analysis. Thus, a player of the energy industry that uses the analysis of big data in the purchase and sale of electricity will gain an advantage over rivals in making decisions [15] However, the conclusion points out that big data analysis methods should undoubtedly be developed. In other words, analysis should be done using algorithms resulting in a better quality of response and doing data mining tasks in shorter periods of time. moreover, we determine the minimum value of electricity and its distribution method from power plants, transfer companies and calculate the price of electricity. Therefore, the advantage of fair pricing analysis using data mining based on big data as well as the improvement by the proposed model is indicated. Improving runtime and increasing the accuracy of predictions, shows that the algorithms used for big data, require intelligent data mining methods.

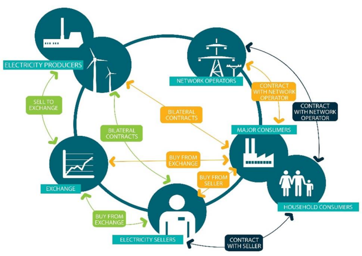

Figure 1: A Schema of Relationships among Different Players in the Electricity Market [23].

2 Method

In the electricity market, a prediction of supply and demand needs to be used, which is applicable to all production data in all circumstances, even when demand and supply trends are removed from the routine and predictions are not accurate with time series. To be used. It is also very important to use a method that is an image of all the past data.This study investigates data that were previously obtained from Irans electricity market databanks to ascertain whether they are big and to show the necessity of using new methods for the analysis of big data. After determining the bigness of data, two data mining methods (classic and modern) are used to show that the nature of the electricity market data is big. Therefore, such data can provide better competitive advantages for the players of this market if big data algorithms are utilized. Next, the model of cost optimization uses the predicted parameters resulted from data mining to establish the appropriate distribution type of electricity and fair pricing.

3 Case Study: Seven Features of Big Data in Irans Electricity Market

3.1 Volume

The most important feature of big data is volume. Companies that can use a higher volume of data in their models in different areas will gain better results [5, 8] This big volume of data is obtained from various sources, such as sensors, webserver logs, audio streams, stock market interactions, webpage contents, and governmental documents in different industries [14] Regarding energy and the electricity market in particular, there has been an increase in the volume of data on parameters influencing the purchase and sale of electricity resulting from different sources, such as power plants, the stock market, the performance of power transmission and distribution companies, and regional power companies since the creation of this market in Iran in 2001. This increase is still growing. After changing the wholesale market into the retail market and implementing the smart grid across Iran in the coming years, it is predicted that this volume of data will increase very rapidly. According to the applied and effective data of different databases from 2002 to 2015, the volume of data reached from 7016 gigabytes to 10158 gigabytes (Fig 2). According to the experts of Niroo Research Institute, the double exponential smoothing prediction method was used to predict this trend until 2022. Based on the mean of results gained from experts predictions and the double exponential smoothing prediction method, this volume of data will reach 14766 gigabytes in 2018 and 21116 in 2020. Data growth in the electricity market is obvious and exponential. The predictions show that this growth will continue in the coming years (Table 2). These valuable data will influence the prediction of important variables, i.e., price and load. Given the time period in which data are generated in Irans electricity market, there are different rates and volumes of accumulated data. The objective of this market is to generate momentary data so that the process of changing parameters can be thoroughly monitored. The use of this procedure may result in a significant increase in the volume of generated data. According to Kaile Zhou et al. (2016), by making the power industry intelligent and using measuring tools for parameters and their relationships with the electricity market databases, over 1.82 terabytes of data will be generated on average [10]. All of these data will directly and indirectly influence the analyses of the electricity market and change the current database trend. However, making Irans electricity market intelligent is part of the next plan, and it has not been completely fulfilled. Even if making the power industry intelligent is neglected, there will be a significant volume of data [14].

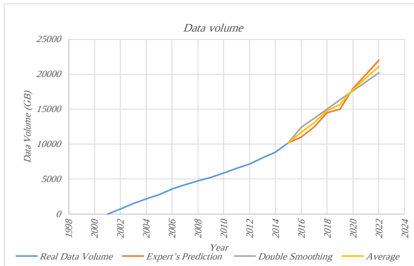

Figure 2: Data Growth Prediction until 2022.

Table 1: Predicting the Growth in the Volume of Irans Electricity Market Data between 2016-2022

Year	Experts Prediction (GB)	Smoothing Method of Prediction (GB)	Average (GB)
2016	11000	12433.2	11716.6
2017	12500	13733.3	13116.65
2018	14500	15033.3	14766.65
2019	15000	16333.4	15666.7
2020	18000	17633.5	17816.75
2021	20000	18933.6	19466.8
2022	22000	20233.6	21116.8

3.2 Velocity

The velocity of input data stream is another feature of big data that can be caused by feedback loops and cycles resulting from making instant decisions in a system. This rapid movement of data requires a powerful structure so that it can process and store input data quickly in order to provide the data necessary for a quick and accurate decision. Therefore, discussion on the velocity of creating data is the discussion on the creation of such structures [5, 8] In Irans electricity market, there are different velocities of data generation based on collection time periods. Some parameters, such as peak electricity consumption and the prediction of minutely load, are put into records; however, some other parameters, such as the average total power plan price or proposed sales stairs, are collected daily. Therefore, there will be different updating periods in the velocity of producing such data depending on their importance. The more we move towards data records in more limited periods of time, the greater the velocity and volume of data generation will be, and the features of generated data will be closer to the features of big data. Therefore, it will be easier to analyze the resultant data [14]. In the electricity market data analysis, the principle is to generate and record daily-collected data into shorter periods of time, such as each minute; therefore, this approach will result in an increase in the generation rate of data in this industry. Figure 3 indicates the differences in the rate of data generation with respect to the collection period. Charts with slower slopes have longer periods of time for data update, meaning one day. On the other hand, charts with steeper slopes include data generation in shorter periods of time, such as every minute. It should be noted that if the electricity market becomes more competitive, so that more players work on purchase and sales in this market, it will be necessary to have databases that can store data with higher rates of generation.

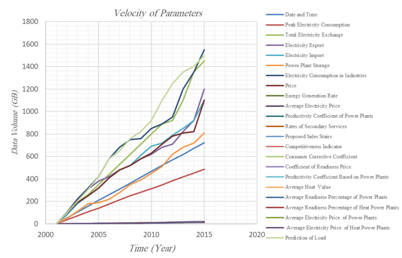

Figure 3: Rates of Different Recessive Parameters in Irans Electricity Market.

3.3 Variety

Big data are data with various structures. Such data include structured, semi-structured and non-structured data [5, 8] Most information is directly obtained from humans; therefore, errors are inevitable. This variety influences the integrity of data because more various data are prone to more errors [4]. Regarding the electricity market, data are obtained from main resources, i.e., electricity-generating power plants, distribution companies, transmission companies, and regional electricity companies regarded as buyers. Each of these sectors and their subsidiaries have a role in the generation of this volume of data. With the expansion of public and private companies operating in the electricity market in recent years, data generation sources have been increasing, which has increased the volume of data. In-

Table 2: Comparing the Number of Data Generation Sources in Each 10 Years

Year	Power Plants	Distribution Companies	Regional Power Companies
2015	46	16	9
2016	90	37	16

creasing data generation sources in the electricity market has resulted in the more accurate prediction of parameters, i.e., load and price. The growth of parameters and sources can be seen in Figure 3. It is predicted that this increasing trend will continue in the number of parameters and sources. It is also predicted that other parameters will be included for more accurate analysis and conclusions. The number of data generation sources changed from 71 to 143 between 2007 and 2015. In this same period, the number of parameters changed from 8 to 22. However, with the development plan of Irans electricity market, it is predicted that organizations involved in the purchase and sale of electricity (public and private) will enter the industry. For instance, there have been more regional electricity companies operating as the purchasers of the electricity market in recent years, and the majority of these companies are present in 2015. Therefore, the number of data generation sources is increasing. Moreover, the creation of intelligent electricity networks in Irans market will significantly increase data sources in the foreseeable future. With the implementation of the intelligent electricity market, new parameters can be defined in electricity consumption, and every household consumer of electricity will become a data generation source [15]. Table 2 indicates the growth and dispersion of power plants, distribution companies, and regional electricity companies in Irans electricity market. Figure 5 shows the number of parameters and sources. The growth of participation increased the volume and variety of generated data.

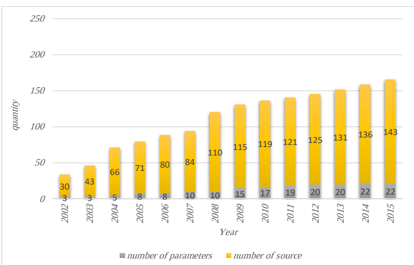

Figure 4: The Growth of Parameters and Sources of Data Generation in Electricity Market

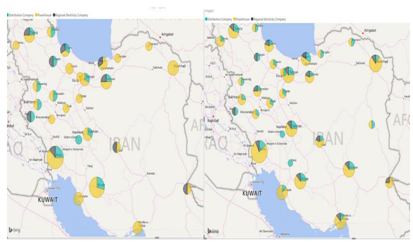

Figure 5: The Dispersion of Players in Irans Electricity Market: a) 2005 b) 2015.

3.4 Veracity

Given that data are obtained from different sources, they may not be reliable. Moreover, it is not easy to ignore the great volume of information[7]. Some investigations introduced this challenge as the maintenance of all the features of big data to guarantee quality and accuracy However, the second definition is true about the drivers of big data to generate a group of data that can indicate the features of main data. In the expanding volume of electricity market data, each parameter has a level of accuracy over the recession time; therefore, they have different levels of errors depending on what sources they come from[1, 2]. Figure 6 shows the average prediction error of the two important variables of electricity market (load and price) between 2002 and 2015. As observed, the values of errors in these two variables can be regarded as indicators for the accuracy of electricity market data with fluctuations in different years. These fluctuations are caused by errors in parameters that can predict the two variables. Increasing the volume of data increases the accuracy of predictions and the probability of errors in collected data used to predict variables, something which is a feature of big data. As Figure 5 shows, the error percentage has been decreasing in recent years; however, a percentage of errors has always been observed in the most important data of the electricity market, i.e., the predictions of load and price.

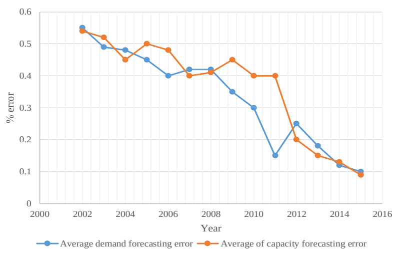

Figure 6: The Average Error in the Predictions of Load and Price.

3.5 Validity

Assuming that data are accurate, they may not be suitable for some applications. In other words, they may not be valid enough for use in other applications[7]. The coefficients of correlation between parameters and variables (load and price) can be investigated to determine how much data influence the predictions. Table 3 indicates the correlations of some parameters with effective variables (load and price). It can be concluded that the parameters having the strongest correlation with the electricity market variables have the greatest impact on predictions, and there are also parameters that have weaker correlations with load and price. Therefore, parameters show different levels of validity in the prediction of important variables. However, maintaining all parameters, even those having the slightest correlation with main variables, results in the maintenance of big data features. Thus, such parameters are always recorded and investigated for the integrity of data. The implementation of data mining methods in Irans electricity market is meant to investigate the effectiveness of the parameters that have indirect impacts on the electricity market. There are parameters such as the rates of secondary services that have much weaker correlations than the important parameters, but investigating their impacts will make data mining methods more accurate. Table 3 indicates the interval of the coefficients of correlation between some parameters, selected as samples, and the main variables of the electricity market.

Irans electricity market databases are recording the parameters, which include a wide range of correlation with predicted variables Figure 7 indicates the minimum and maximum coefficients of correlation between parameters selected in Table 3 and the two important variables of the electricity market as well as the expansion of range.

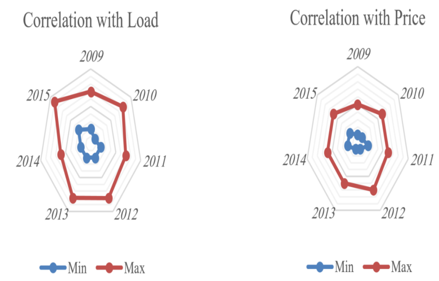

Figure 7: Comparing the Ranges of Correlation between Selected Parameters and Variables (Load and Price).

3.6 Volatility

The values of different data can change at different speeds over time. In an ordinary electronic commerce system, data do not fluctuate very fast. It is possible that data can maintain their values for one year. However, there are great fluctuations in other applications, such as currency and stock analysis, and data change their values very quickly. However, it is important to store information for a long time to analyze the changes and fluctuations in data. Increasing the storage period will definitely have high implementation costs, which should be taken into account. In Table 4, the moving average method and the exponential smoothing method were used based on real information to predict load and price [17]. The coefficient of the smoothing method was 0.8, while that of the moving average was 3 periods of prediction. Therefore, the latter is more important than the former when using older data. Comparing the error resulting from these two methods with respect to the low error of the exponential smoothing method for the prediction of both parameters, it is obvious that historic data kept their values. More accurate predictions were made by using historic data; however, the fluctuations in errors can be caused by different factors, such as the accuracy of assumed parameters.

Figure 8 shows the predictions calculated by using the exponential smoothing method with 0.6

Table 3: Coefficient of Correlations between Parameters and Variables (Price and Load).

Year	2009		2010		2011		2012	
parameters	Price	Load	Price	Load	Price	Load	Price	Load
Power Plant Storages	0.6	0.8	0.5	0.6	0.5	0.8	0.7	0.6
Energy Production Rate	0.3	0.5	0.2	0.2	0.3	0.6	0.4	0.6
Productivity Coefficient of Power Plants	0.5	0.3	0.3	0.2	0.6	0.4	0.4	0.3
Rate of Secondary Services	0.4	0.5	0.3	0.3	0.2	0.2	0.4	0.5
Proposed Sales Stairs	0.4	0.9	0.2	0.6	0.2	0.8	0.4	0.8
Corrective Coefficients of Consumers	0.3	0.4	0.3	0.2	0.4	0.3	0.5	0.3
Coefficient of Productivity for Power Plants	0.2	0.4	0.3	0.2	0.1	0.3	0.1	0.4
Average Heat Value	0.4	0.6	0.6	0.4	0.3	0.4	0.5	0.5
Coefficient of Readiness Fee	0.4	0.4	0.3	0.2	0.3	0.3	0.6	0.5
Electricity Import	0.2	0.5	0.4	0.3	0.3	0.3	0.3	0.3
Electricity Export	0.4	0.4	0.2	0.2	0.3	0.5	0.5	0.2
Coefficient of Productivity based on Power Plant	0.5	0.6	0.3	0.4	0.4	0.5	0.6	0.5

and 0.4 coefficients. The bigger coefficient includes smaller errors, which means that considering previous data is important to a certain extent. Therefore, there will be fewer prediction errors.

3.7 Value

Value refers how many data are important in making decisions in terms of information. In other words, it ascertains whether the expenditures spent on maintaining and processing information are worth making decisions or not. Data can usually be moved between different layers. Higher layers have higher values of data. Therefore, some organizations can afford the high maintenance expenditures pertaining to higher layers. The initial investment is in creating a hardware and software structure for the storage of information and data; however, there will be much lower costs required to upgrade these structures afterwards. In Irans electricity market, the income from creating hardware and soft-

Tabel 3. Continue

Year	2013		2014		2015	
parameters	Price	Load	Price	Load	Price	Load
Power Plant Storages	0.5	0.5	0.5	0.4	0.2	0.7
Energy Production Rate	0.4	0.4	0.3	0.2	0.3	0.3
Productivity Coefficient of Power Plants	0.6	0.2	0.6	0.3	0.2	0.2
Rate of Secondary Services	0.4	0.2	0.3	0.1	0.1	0.2
Proposed Sales Stairs	0.4	0.7	0.5	0.8	0.3	0.8
Corrective Coefficients of Consumers	0.2	0.3	0.2	0.3	0.4	0.5
Coefficient of Productivity for Power Plants	0.3	0.2	0.1	0.3	0.3	0.5
Average Heat Value	0.3	0.5	0.1	0.7	0.5	0.5
Coefficient of Readiness Fee	0.5	0.4	0.2	0.3	0.3	0.3
Electricity Import	0.3	0.4	0.1	0.6	0.1	0.3
Electricity Export	0.5	0.3	0.1	0.4	0.4	0.3
Coefficient of Productivity based on Power Plant	0.3	0.3	0.3	0.5	0.4	0.4

ware databases come from appropriate decisions made in this industry. The income of predicting a price unit appropriately is equal to the price of one megawatt per hour (4.20), and the savings of predicting the required electricity appropriately in one megawatt per hour is equal to 11.80. Since the creation of Irans electricity market and the increase in the volume of collected data, the prediction error has increased with the definition of different parameters. Decreasing the prediction error in the consumed electricity load and predicting the accurate electricity price have resulted in an increase in the income of producers and consumers [18]. Therefore, when the volume of data becomes bigger, big data methods will be required to provide more savings. In the same way, the profit of accurate predictions has been increasing. According to Figure 9, the break-even

Table 4: Calculating the Average Error with Moving Average Method and Exponential Smoothing Method (specified weights = 0.8).

Prediction Parameters	Values	2005	2006	2007	2008	2009	2010
50.5cmPrediction of Price	Real Value	144000	145100	155544	156652	168320	169500
	Moving Average	144550	145100	148214	150324	153859	154000
	Error Percentage	3	1	5	4	8	9
	Exponential Smoothing	143820	144848	153409	156004	160066	168000
	Error Percentage	0.12	0.17	1.37	0.41	4.72	0.9
50.5cmPrediction of Load	Real Value	31250	31680	32200	32450	32560	32000
	Moving Average	32831	33006	33172	33311	33455	32500
	Error Percentage	5	4	3	3	3	1
	Exponential Smoothing	32771	32960	33129	33274	33419	33450
	Error Percentage	0.18	0.14	0.13	0.11	0.11	0.95

Table 4. Continue

Prediction Parameters	Values	2011	2012	2013	2014	2015	Average
50.5cmPrediction of Price	Real Value	168000	169000	178000	188000	189000	-
	Moving Average	156215	158042	160536	163587	166128	-
	Error Percentage	7	6	10	13	12	6.9
	Exponential Smoothing	167520	168704	176141	185628	188326	-
	Error Percentage	0.28	0.17	1.04	1.26	0.35	0.9
50.5cmPrediction of Load	Real Value	33000	33200	33500	33550	34920	-
	Moving Average	33634	33792	33990	34295	34100	-
	Error Percentage	2	2	1	2	2	2.7
	Exponential Smoothing	33591	33751	33943	34224	34124	-
	Error Percentage	0.13	0.12	0.14	0.21	0.07	0.14

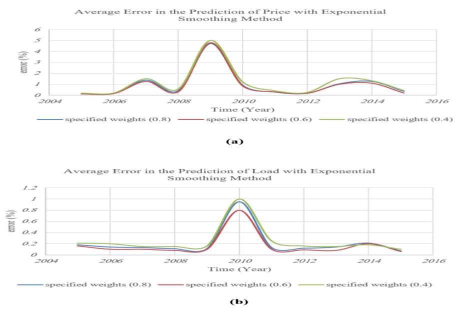

Figure 8: a) Average Error in the Prediction of Price with Exponential Smoothing Method b) Average Error in the Prediction of Load with Exponential Smoothing Method.

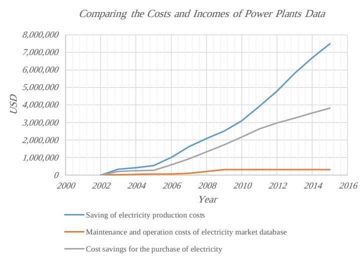

Figure 9: The Growth of Parameters and Data Generation Sources

point of creating and maintaining the electricity market data was observed in the first year of implementation (2002). Then it started to produce income. The maintenance of previous data will be fruitful with the current structure, although there are still some problems in the analysis of data.

4 Optimization and fair pricing of electricity

4.1 Definition of multi-objective model

The multi-objective model shall be used to minimize the costs for electricity consumers and suppliers. The determined indexes, sets, parameters and variables are as follows:

Table 5: Indexes, sets, parameters and variables of the model.

index	Description
$i \in I$	Index of each plant
$j \in J$	Index of each electricity transmission companies
$k \in K$	Index of electricity distribution companies
$t \in T$	Index of each period
Sets	Sets
I	Set of electric power plants
J	Set of electricity transmission companies
K	Set of electricity distribution companies
T	Set of period
Parameters	Description
N	The total number of power plants
M	The total number of transmission companies
L	The total number of distribution companies
d_{kt}	Unit k Demand in period t
q_{maxi}	The maximum capacity of unit i
q_{mini}	The minimum capacity of unit i
q'_{maxj}	The maximum power transmission capacity of unit j
q'_{minj}	The minimum power transmission capacity of unit j
A_t	The maximum allowable number of power plants to supply electricity transmission company in period t
B_t	The maximum allowable number of electricity transmission companies to supply electricity distribution companies in the period t
$C_i^{(t)}$	The cost of generating a unit load in plant i in period t

1- Minimization of total operating costs is in 1:

$$\min Q = \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{t=1}^{T} u_{it}[C_{it}(q_{ijt}) + S_{it}] + \sum_{j=1}^{M}\sum_{k=1}^{L}\sum_{t=1}^{T} u_{jt}[C'_{jt}(q'_{jkt}) + S'_{it}] \ (1)$$

2- Minimization of the total cost of electricity power plants in 2:

$$\min R = \sum_{i=1}^{N}\sum_{t=1}^{T} P_{it} \ (2)$$

3- Minimization of the cost of electricity not supplied by the electricity distribution companies in 3 and 4:

$$\min S = \sum_{j=1}^{M} \omega'_{jkt} q'_{jkt} \forall k \in K, \forall t \in T \ (3)$$

$$\min U = \varphi Q + \varphi' R + \varphi'' S \ (4)$$

4- Indicates the minimum required spinning reserve in time t :

$$\sum_{j=1}^{M} \omega'_{jkt} q'_{jkt} \geq d_{kt} + R_{kt} \forall k \in K, \forall t \in T \ (5)$$

5- These constraints include the minimum and maximum of production and the minimum on and off times of production and transfer units, which is of the operational constraints of production units in 6 to 11:

$$u_{it} q_{mini} \leq \sum_{j=1}^{M} \omega_{ijt} q_{ijt} \leq u_{it} q_{maxi} \forall i \in I, \forall t \in T \ (6)$$

$$(X_{on.i}^{t-1} - T_i^{up})(u_{i(t+1)} - u_{it}) \geq 0 \forall i \in I, \forall t \in T \ (7)$$

Table 5. Contunue.

C''^{t}_{jk}	Load transmission cost by company j from Unit i to Unit k in period t
S_{it}	Fixed generation cost in Unit i in period t
α_{it}	the profits Percent generation in Unit i in period t
α_{it}	the profits Percent generation in Unit i in period t
$S'_j t$	Fixed cost of transmission by Unit j in period t
$R_{kt} R_{kt}$	k Units Spinning reserve required in period t
T_i^{up}	Minimum time that the Unit i should be in operation
T_j^{up}	Minimum time that Unit j should be in the circuit transmission
T_i^{down}	Minimum time that the Unit i should be shutdown
T_j^{down}	Minimum time that Unit j shouldnt be in the circuit transmission
Variables	Description
$X_{on.i}^{t}$	operation duration time of Unit i in period t
$X_{off.i}^{t}$	Shutdown duration time of Unit i in period t
$X'^{t}_{on.j}$	Operation duration time of Unit j in period t
$X'^{t}_{off.j}$	Shutdown duration time of Unit j in period t
A'_{kt}	Affordability of unit k in period t
q_{ijt}	Amount of loud that generated in Unit i and transferred by Unit j in period t
q'_{jkt}	Amount of loud that generated in Unit j to Unit k in period t
P_{it}	The price of electricity for Unit i in period t
u_{it} u_{it}	If unit i in period t is generating is 1 else is 0
u'_{jt} u'_{jt}	If unit j in period t is in circuit is 1 else is 0
ω_{ijt}	If Unit i supply Unit j in period t is 1 else 0
ω'_{jkt}	If Unit j supply Unit k in period t is 1 else 0

$(X_{off.i}^{t-1} - T_i^{down})(u_{i(t+1)} - u_{it}) \geq 0 \forall i \in I, \forall t \in T$ (8)

$u_{\prime jt} q_{\prime minj} \leq \sum_{i=1}^{N} \omega_{\prime jkt} q_{\prime jkt} \leq u_{\prime jt} q_{\prime maxj} \forall k \in K, \forall t \in T$ (9)

$(X_{\prime on.j}^{t} - T_j^{up})(u_{\prime j(t+1)} - u_{\prime jt}) \geq 0 \forall j \in J, \forall t \in T$ (10)

$(X_{\prime off.j}^{t} - T_j^{down})(u_{\prime j(t+1)} - u_{\prime jt}) \geq 0 \forall j \in J, \forall t \in T$ (11)

6- Constraints on payments to ensure that all winner units cover their proposed costs of no load, installation and production of energy in 12: $A'_{kt} \geq \sum_{j=1}^{M} (\sum_{i=1}^{N} C_{it}(\omega_{ijt} q_{ijt}) + S_{it} + C_{\prime jt}(\omega_{\prime jkt} q_{\prime jkt}) + S_{\prime jt} \forall k \in K, \forall t \in T$ (12)

7- Constraint of the balance of supply chain: total delivered load amount of transfer companies is equal to the amount delivered to distribution companies in 13: $\sum_{i=1}^{N} \omega_{ijt} q_{ijt} \geq \sum_{k=1}^{L} \omega_{\prime jkt} q_{\prime jkt} \forall j \in J, \forall t \in T$ (13)

8- Restriction on the relationship between profit, cost and revenue is indicated by the following equation in 14:

$P_{it} = (1 + \alpha)(C_{it} \sum_{j=1}^{M} (\omega_{ijt} q_{ijt}) + S_{it}) + (1 + \beta)(C_{\prime jt} \sum_{j=1}^{M} (\omega_{\prime jkt} q_{\prime jkt}) + S_{\prime it}) \forall i \in I, \forall t \in T$ (14)

9- Limiting the maximum link of distribution and transfer companies as well as the maximum link of transfer companies to power plants in 15 and 16:

$\sum_{i=1}^{N} \sum_{j=1}^{M} \omega_{ijt} \geq A_{kt} \forall k \in K, \forall t \in T$ (15)

$\sum_{j=1}^{M} \omega_{\prime jkt} \geq B_{kt} \forall k \in K, \forall t \in T$ (16)

10- If the power plant (i) is operating within circuit in time t, it should be the same in the next periods. Moreover, if the transmission company of (j) operates within circuit at time t, then it should be the same in the next two periods in 17 and 18 :

$u_{it} \geq u_{i(t+1)} + u_{i(t+2)} \forall t \in 1, ..., T$ (17)

$u_{\prime jt} \geq u_{\prime j(t+1)} + u_{\prime j(t+2)} \forall t \in 1, ..., T$ (18)

4.2 Classic and Modern Data Mining Methods for the Electricity Market Big Data

After explaining the above-mentioned model and its associated limitations, it is necessary to conduct data mining in order to estimate the input parameters which include demand estimation and minimum and maximum capacity of power plant to supply electricity. In other words, the input parameters of the model are calculated as the estimated variables of past data clustering. According to previous investigations of the features of big data, it is obvious that the data of this field are big. Therefore, modern data analysis methods are required. Different methods have been proposed for the analysis of big data. For instance, there are some methods in which time and space are not important, such as Map Reduce and Graph Lab methods. On the other hand, there are some methods in which time limitations are considered, such as STORM, SAMOA, or SVMGA [20]. However, there are classic data mining methods that have severe limitations in the analysis of big data. In this section, a classic data mining method, K-means, is compared with SVMGA. In the electricity market, price prediction and load prediction are two important variables. Therefore, the data obtained from Irans electricity market are used in separate years to cluster data. Table 5 shows the results of clustering data in 2015 by using SVMGA with R version 3.1.0 software [20]. After defining the status of each of the parameters included in the table 5, it is necessary to determine the distance between points and cluster centers to identify the nearest cluster. Other target parameters are determined by the cluster after identifying the number of the appropriate cluster. An example of parameter determination is given in the table 6. These values are in fact input parameters of minimizing model.

4.3 solve the model

After obtaining the parameters from the clustering of big data on the electricity market database and the other given parameters, Multi-objective model will be solved by considering 26 plants, 31transmission companies and 56 distribution companies. The maximum allowable power plants to supply are 3 and the maximum number of transmission companies are 5 units. The permitted Profit for power plants is 30 percent and for transmission companies is 25 percent. the optimization of price is as follows. This is a fair price based on a minimum of three objectives have been achieved

5 Results

Data mining using big data techniques for the analysis of the big data of electricity market provides more accurate and faster prediction of the parameters such as customer demand and production capacity of power plants. Given the largeness of data, it is necessary to employ new techniques of big data to improve the response quality and speed-up data mining. Graph 10 presents an average improvement by 6 percent in the price estimating quality and graph 11 presents the improvements made by the algorithm and Since there is a large volume of data since 2004, the analysis cannot be handled by the traditional methods of data mining. After data mining and estimations of demand and minimum and maximum electricity generating capacity of power plants and other required parameters, the multi-objective model is implemented by the director of the Electricity Market Department to minimize the costs. When the electricity distribu-

Table 6: Clustering Results Using SVMGA Based on Data in 2015.

Cluster No.	Price	Load	Elec. Con. in Peak	Elec. Con. (Last Year)	Tot. Power Exch.	Elec. Exp.	Power Emp.	Powerhouse Storage	Industries Elec. Exch.
1	3228	1112	353	951	1456	0	21989	25194	30910
2	3823	9044	343	1102	1445	32625	32633	32843	31693
3	3605	10197	391	882	1273	0	30132	25400	38451
4	3775	9842	364	1362	1726	32670	34104	33359	33581
5	2862	7643	370	935	1305	0	21165	26266	31697
6	3654	10147	414	789	1203	32613	33877	31098	41100
7	2981	6985	365	717	1082	0	30785	31219	34791
8	3682	14148	519	757	1276	32497	33679	28656	44478
9	4147	14327	387	717	1104	0	31165	33891	34863
10	4200	16011	581	703	1284	32429	30690	27639	38185
11	3905	17203	393	699	1092	0	26289	33092	42823
12	3703	14187	570	866	1436	29726	32964	32207	33732
13	3608	12959	371	876	1247	0	30300	27018	39073
14	4030	14611	630	864	1494	31233	31437	30336	39095
15	2936	4031	360	677	1037	0	21842	28788	31412
16	4017	13494	536	845	1381	32058	33090	30450	41255
17	3054	9730	364	650	1014	0	25002	28114	41987
18	3002	551	411	1071	1627	32247	34157	29990	33796
19	3958	12640	390	798	1188	0	31340	33136	37168
20	2965	7704	452	812	1406	30720	31273	25471	33492
21	3400	7523	368	618	986	0	22014	26900	33780
22	3690	12542	600	644	1244	31908	33893	30148	41079
23	3826	11465	397	620	1017	0	30966	26900	39751
24	3814	12197	588	1044	1632	31836	34126	25487	30304
25	3852	10884	448	803	1251	0	31022	30444	34831
26	3948	11496	624	912	1536	29424	34381	26322	41256
27	3770	11317	416	845	1261	0	31047	32851	37487
28	4203	13557	580	862	1442	30365	31489	33715	44158
29	3482	627	427	825	1252	27600	27461	27271	31093
30	3668	12228	680	882	1562	31511	32990	29817	44364
31	3685	11689	370	535	905	0	30318	25725	32453
32	3522	2073	188	1591	1779	26486	43068	30312	36619
33	3242	1599	295	1526	1821	45046	42426	26804	38923
34	3390	2973	525	1605	2130	44800	43165	25860	41696
35	3317	2546	649	1492	2141	45144	43153	32541	4275
36	3290	1681	376	883	1259	39459	38540	33767	37201
37	3456	1879	601	1510	2111	45095	45135	28563	44481
38	3609	1181	606	1515	2121	44647	45197	25087	41812
39	3501	692	351	1054	1405	42310	43002	25905	31383
40	3540	1295	359	847	1206	29330	30087	26821	30166

tion method is determined in the supply chain, the electricity will be fairly priced in the model. The following graph 12 presents the implementation of meta-heuristic genetic-based algorithm to optimize objective functions and changes in the 26 values for fair prices: The price is suggested by the director of the Iranian Electricity Market Department for the new period and is a benchmark for the market exchanges. Because of high volume of exchanges in the electricity market, improved predictions of demand and electricity generating capacity as two important input parameters for the problem and also other parameters required for the model lead to accurate prediction of electricity price, hence incredible savings in this market. Figure 13 presents a graph of revenues (the product of multiplying the volume of electricity generation by power plants and the price per unit) for 5, 6 and 7 percent improvements in parameter estimations based on data mining.An

Table 7: Model parameters Results Using SVMGA Based on Data in 2015.

Cluster No.	i	j	k	$q_{max(i)}$	$q_{min(i)}$	$q'_{max(j)}$	$q'_{min(j)}$	d_{kt}	$ClusterNo.$
i	j	k	$q_{max(i)}$	$q_{min(i)}$	$q'_{max(j)}$	$q'_{min(j)}$	d_{kt}		
1	1	1	1	560	542	1.4	1.1	4.1	1
1	14	1	534	521	1.3	0.8	4.2		
1	1	2	1	532	520	1.6	1.3	4.8	1
1	15	1	546	540	1.4	1	4.2		
1	1	3	1	542	504	1.6	1.2	3.6	1
1	16	1	569	540	1.4	1	4.2		
1	1	4	1	536	480	2.1	1.8	2.3	1
1	17	1	598	534	0.8	3.5	3.6		
1	1	5	1	534	523	2.3	2	3.5	1
1	18	1	534	501	0.8	3.6	3.1		
1	1	6	1	562	502	2.4	2	3.5	1
1	19	1	546	509	1.8	3.5	3.1		
1	1	7	1	524	492	3.4	1.2	3.1	1
1	20	1	522	497	2.1	3.7	3.5		
1	1	8	1	528	503	3.6	2.3	3.5	1
1	21	1	534	489	2.1	3.6	3.9		
1	1	9	1	549	496	3.8	2	3.6	1
1	22	1	565	490	1.8	3.6	4.1		
1	1	10	1	489	480	3.4	2.1	4.1	1
1	23	1	513	478	1.8	3.2	4.3		
1	1	11	1	495	490	1.5	0.8	4.1	1
1	24	1	567	508	1.2	3.5	4.6		
1	1	12	1	509	498	1.2	0.8	4.2	1
1	25	1	555	534	1.2	3.1	4.4		
1	1	13	1	521	502	1.1	1	4.6	1
1	26	1	554	503	1.2	3.3	4.2		

Table 8: Model parameters Results Using SVMGA Based on Data in 2015.

i	P_{it}	i	P_{it}	i	P_{it}
1	11.3765	10	9.9975	19	11.0547
2	10.5757	11	9.7175	20	10.5775
3	11.4	12	8.56	21	10.639
4	10.579	13	8.061	22	11.675
5	10.59	14	8.614	23	11.640
6	10.814	15	11.141	24	11.163
7	10.692	16	11.08	25	11.08
8	10.94	17	11.14	26	10.80
9	10.80	18	11.33	-	-

accurate prediction of parameters of electricity demand and generating capacity of power plants is effective in accurately predicting the electricity price in the next period, and since high volume of investments are involved in the market over the years, the slightest improvement can make great revenue (in dollar) for sellers and significantly reduce costs for purchasers. The promising future in the electricity market is for the actors who carry out a more efficient data mining in this competitive field. Data generated in the electricity market undoubtedly has the characteristics of the big data. Characteristics that have been examined case by case in the electricity market of

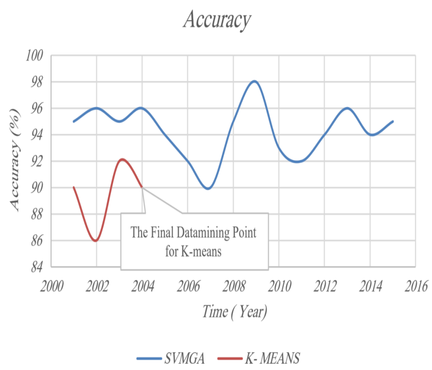

Figure 10: The Average Accuracy of Implementing Algorithm to Predict Load and Price

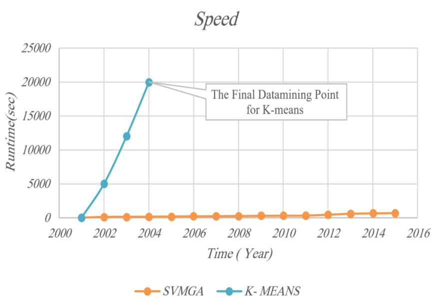

Figure 11: The Speed of Algorithm Implementation

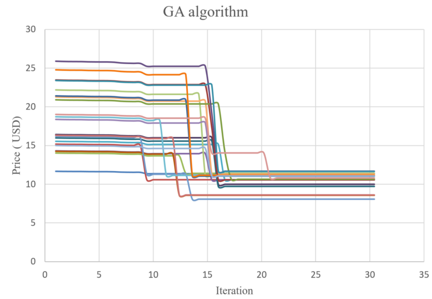

Figure 12: GA Operation

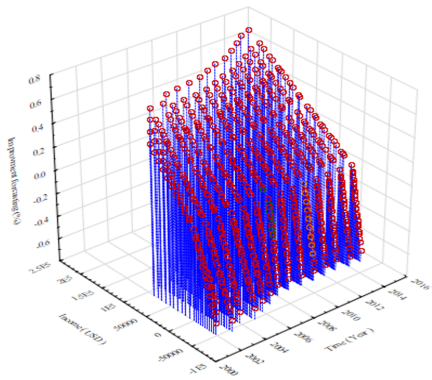

Figure 13: Income Based on Time and Improvement in the Production Accuracy (0.05%, 0.06%, and 0.07%) with Big Data Algorithms (SVMGA)

Iran.

References

[1] P. Andr, G. Maugis, Big data uncertainties, *Journal of Forensic and Legal Medicine* 57 (2018) 7-11

[2] L. M.Caesarius, J. Hohenthal, Searching for big data: How incumbents explore a possible adoption of big data technologies, *Scandinavian Journal of Management* 34 (2018) 129-140.

[3] M. DeLisi,The big data potential of epidemiological studies for criminology and forensics, *Journal of Forensic and Legal Medicine* 57 (2018) 24-70

[4] J. Eckroth, A course on big data analytics, *Journal of Parallel and Distributed Computing* 118 (2018) 166-176.

[5] Q. Zhang, L. T. Yang, Z. Chen, P. Li, A survey on deep learning for big data, *Information Fusion* 42 (2018) 146-157.

[6] B. Mokhtar, M. Eltoweissy, Big data and semantics management system for computer networks, *Ad Hoc Networkss* 57 (2017) 32-51.

[7] C. K. Emani , N. Cullot , C. Nicolle, Understandable Big Data: A survey, *Computer Science Review* 17 (2015) 70-81.

[8] A. Shalaginov, K.Franke ,Big data analytics by automated generation of fuzzy rules for Network Forensics Readiness, *Applied Soft Computing* 52 (2017) 359-375.

[9] Suren Islyaev, Paresh Date , European Journal of Operational Research, Electricity futures price models, *Calibration and forecasting* 247(2015) 144-154.

[10] K. Zhou, S.Yang, Understanding household energy consumption behavior: The contribution of energy big data analytics, *Renewable and Sustainable Energy Reviews* 56 (2016) 810-819.

[11] S. Suresh, Big Data and Predictive Analytics : Applications in the Care of Children, *Pe-*

diatric Clinics of North America 63 (2016) 357-366.

[12] N. Kshetri, Big datas role in expanding access to financial services in China, *International Journal of Information Management* 36 (2016) 297-308.

[13] A. Chluski, L. Ziora, The Role of Big Data Solutions in the Management of Organizations Review, *International Conference on Communication, Management and Information Technology (ICCMIT 2015), Procedia Computer Science* 65 (2015) 1006-1012.

[14] K. Zhoua, C. Fua, S. Yanga, Big data driven smart energy management: From big data to big insights, *Renewable and Sustainable Energy Reviews* 56 (2016) 215-255.

[15] P. D. Diamantoulakisb, V. M. Kapinasb, G. K. Karagiannidis, Big Data Analytics for Dynamic Energy Management in Smart Grids, *Big Data Research* 2 (2015) 94-101.

[16] V. Kekatos, Y. Zhang ; G. B. Giannakis, Electricity Market Forecasting via Low-Rank Multi-Kernel Learning, *IEEE Journal of Selected Topics in Signal Processing* 8 (6) (2004) 113-124.

[17] V. Kekatos , Y. Zhang , G. B. Giannakis, Low-rank kernel learning for electricity market inference, *Asilomar Conference on Signals, Systems and Computers* (2013).

[18] Y. C. Liang, X. Lu, W. D. Li, S. Wang,Cyber Physical System and Big Data enabled energy efficient machining optimisation, *Journal of Cleaner Production* 187 (2018) 46-62.

[19] S. Wang, Y. C. Liang, W. D. Li, X. T. Cai, Big Data enabled Intelligent Immune System for energy efficient manufacturing management, *Journal of Cleaner Production* 195 (2018) 507-520.

[20] O. Y. Al-Jarraha, P. D.Yoob, S. Muhaidatc, G. K. Karagiannidis, K.Tahaa, Efficient machine learning for big data: A Review, *Big Data Research* 2 (2015) 87-93.

[21] Svetlana Borovkovaa, Maren Diane Schmeckb,Electricity price modeling with stochastic time change, *Energy Economics* 63 (2017)51-65.

[22] H. T. Pao , Forecasting electricity market pricing using artificial neural networks, *Energy Conversion and Management* 48 (2007) 907-912.

[23] http://elering.ee/information-for-small-scale-consumers.

Mebaholdin Salami is student at Islamic azad university Central Tehran Branch. He also studied industrial engineering in Master's degree and bachelor degree in Iran. Now he is working on big data specially in energy demand price forecasting. In addition, he is interesting in multi-objective model and metaheuristic solving method. He really attend to combine big data with other topics in industrial and management.

Farzad Movahedi Sobhani received his Ph.D. in Industrial Engineering from Department of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran. He is currently a faculty member at the Department of Industrial Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran. His interests include business process management, knowledge management, and data mining.

Mohammad Sadegh Ghazizadeh is Associate professor at Shahid Beheashti University (Power and Water University of Technology), he interested in power system operation and control, Electricity Markets and smart grid. Price Forecasting is one of interesting topic that Discuss in electricity market and he focus on it.