



## Research Article

# Mining Patterns from Complex Energy Dataset: A Case Study of Pakistan's Energy Sector

Muhammad Shaheen

Faculty of Engineering and IT, Foundation University Rawalpindi Campus (FURC) New Lalazar, Pakistan.

**Abstract:** The role of energy in development of a country is vivacious. The sectors which when stand sustainable guarantees long-life prosperity of a nation include energy sector. The role of energy in social, economic and environmental factors defines its prominence in sustainability. To ensure sustainable energy development the consumption and production patterns may comply with each other and may also comply with the conditions of sustainability. This paper reviewed the current state of energy in Pakistan and proposed an automated intelligent geographic decision support system that ensures optimal energy development plan for the country in order to cope with the current energy crisis. The paper proposed data mining add-ins for its inclusion on GIS interface.

**Received:** January 10, 2020; **Accepted:** May 20, 2020; **Published:** June 20, 2020

\***Correspondence:** Muhammad Shaheen, Faculty of Engineering and IT, Foundation University Rawalpindi Campus (FURC) New Lalazar, Pakistan; **Email:** dr.shaheen@fui.edu.pk

**Citation:** Shaheen, M., 2020. Mining patterns from complex energy dataset: A case study of Pakistan's energy sector. *Journal of Engineering and Applied Sciences*, 39(1): 41-46.

**DOI:** <http://dx.doi.org/10.17582/journal.jeas/39.1.41.46>

**Keywords:** Data mining, Knowledge management applications, Sustainability, GIS

## Introduction

With the advent of modern technologies, the practices of energy usage and its production have been transformed in the last few decades. The energy development is wanted to be sustained to address the issues related to poverty and human welfare. The crisis of energy across the world make people realize the need for sustainable energy development a converse of which caused much of the anguish. There was a time when Pakistan's energy sector was amongst the world leading energy sectors but the scenario is now changed altogether. Pakistan like most of the other world countries is desperately looking for optimal energy utilization as well as for exploration of new energy reserves. Despite having huge energy reserves in the near past, the energy development in this region is not sustainable.

economic development of a nation. It is the primary need of every country although the primary source of energy in every country is different. Some of the world countries are self-sufficient or at least using oil/ natural gas as primary source while others are using renewable energy sources. Energy may be considered as primary catalyst of global economy. The world population largely depends upon the use of oil as fuel, use of animal power and use of bio-gas etc. Energy with all its dependency graphs and prominence is depleting with an exponential speed. Even the countries where there are huge energy reservoirs are becoming unable to change socio-economic dimensions of the country in a remarkable way. Energy sectors are dedicating bigger amounts of budgets to planning and development in spite of which large scale economic recessions are faced in parts of the world (Shaheen *et al.*, 2011).

Energy defines its prominence in social and

Sustainability is defined as "The development that

meets the needs of today without compromising the ability of future generations to meet their needs (Bruntland, 1987). The question is that what is sustainability in energy development? If energy consumption of a country or its demand complies with the prospection or production, the development is said to be sustainable (Shaheen *et al.*, 2013). An imbalance between two leads to unsustainable energy development and may cause culmination of energy resources. A study on use of crude oil in Pakistan forecasted depletion of oil reserves in Pakistan around 2020 (Shaheen and Khan, 2015).

There was a time when Pakistan's energy sector was amongst the world leading energy sectors but the scenario is changed altogether. The country like most of the other world countries is desperately looking for optimal energy utilization as well as for exploration of new energy reserves. Despite having huge energy reserves in past, the energy development in this region is not sustainable. Pakistan like many other countries of the world is investing much of its potential on exploration of new energy reserves as it seems to be the only solution to meet the energy crisis. The mitigation plan for energy crisis may not be optimized unless information technology with all its potentials will be utilized for:

- Analysis and optimization of existing energy network
- Prospection of new energy reservoirs.

The amount of energy network and prospection data is enormously large. The data is not only larger in volume but also diverse in nature. With the emergence of remote sensing technologies, the analysis of various energy scenarios with geographical details became possible on desktops. But if the analysis will solely be made on remotely sensed data, the analyst will miss finer details related to the attribute values. Remote sensing also increased the sizes of datasets making them less invincible for non-automated analysis. Geographic Information System (GIS) is a computer system used to map, model, query and analyze large quantities of spatial and non-spatial data. It has layered architecture and it stores data related to space, geometry and position (spatial data) and data related to attributes (non-spatial data). GIS is an ideal information system for storing all energy data of Pakistan. GIS enormously increases the volume of data; statistical techniques may not be used for analysis. History can predict future, is the baseline for

the domain of data mining. It is an emerging domain that analyze the past data on the proposition that the size of dataset available for analysis is directly proportional to the accuracy of the obtained results. Different models of executing a data mining task where each model is comprised of some steps that start processing the raw data and ends up at trends and patterns (Shaheen *et al.*, 2010).

#### *Energy scenario In Pakistan*

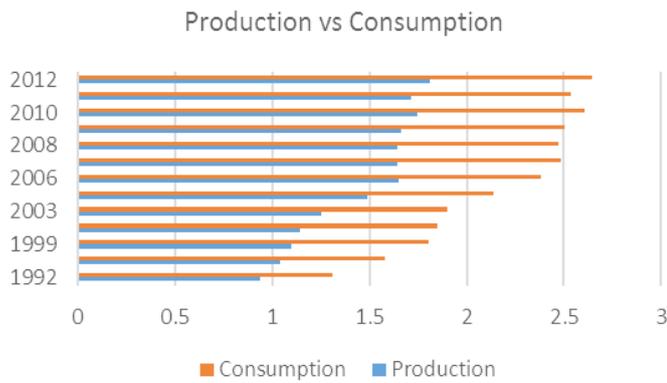
There was a time when Pakistan's energy sector was considered to be amongst world leading energy sectors. Because of development in industrial/ power sector and lack of energy development planning, Pakistan is facing severe energy crises and having shortage of electrical, natural gas/ oil and renewable energy development. In Pakistan, Water and Power development authority has been producing 6444 Mega Watts (MW) of electrical energy. Similarly, GENCO, PAEC, IPPs and KESC produces 4834, 462, 6185 and 1756 MW, respectively. A supply and demand analysis have shown that Pakistan has been facing a shortfall of 7539 MW of electrical energy which has been halted the businesses in execution. According to a study, the country is having total power potential of 54711 MW (Shah and Bhatti, 2009).

In a study (Shah and Bhatti, 2009) it is concluded that in order to create balance between demand and supply, lines losses control, improvement in power generation capacity, medium term solutions like renewable energy development, coal energy development and hydel power utilization should immediately be considered. All this will become possible only when the energy planning department would find an ease in network and energy scenario analysis and identification of hotspots.

Similarly, Pakistan was having larger potential reserves of natural gas at times but when the consumption of natural gas in Pakistan rose to 16,307,898 Tons Oil Equivalent (TOE) in 2011 from 10,067,245 TOE in 2003, the reserves seems to deplete. The supply of the same was 25,254,481 in 2003 and it is 30,241,540 at the current. The statistics for oil and other energy sources are also disappointing. The reason for non-sustainability in natural gas is UFG (Unaccounted for Gas) which is the most critical issue in pipeline companies, low exploration rate and congested network. UFG and congestion on network may be resolved by efficient network management while the

exploration rate can be increased if the planner is facilitated by the following:

- Prospection of new reserves/ sources is ensured in less cost/ time.
- Prospection of new reserves/sources is ensured at ease even at desktops.



**Figure 1:** Comparison of energy consumption and production in Pakistan.

### Use of data mining for energy development

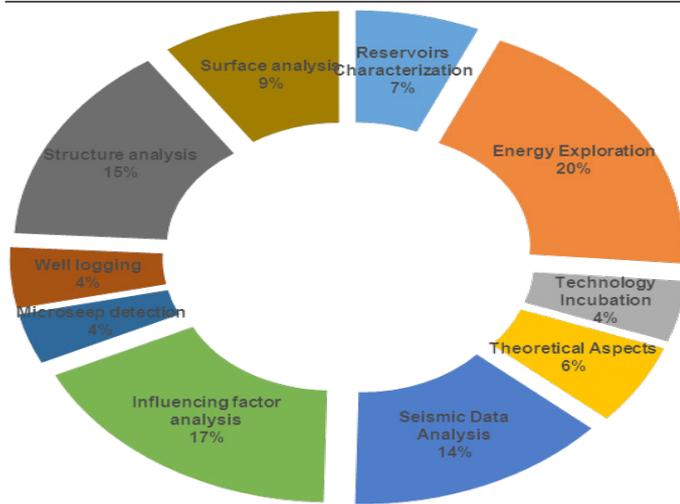
As stated earlier, the datasets of energy sector are much larger in size especially when remote sensing is used as data collection tool. Data mining is an emerging domain that extracts knowledge from data by historical analysis of data. The domain currently also has emerged the concepts of spatial data mining which accommodated remotely sensed data hence making both spatial and non-spatial data evaluable through data mining. The process of knowledge discovery in numeral, ordinal and text databases after multiple iterations have now matured and can be applied onto diverse formats of data which are not limited to graphs, text, web logs and spatio temporal data. Data mining does not bring any change in the databases on which it is applied but apply some techniques picked from other analysis domains like statistics, artificial intelligence and cognitive science. There are substantial databases created for different purposes in the energy development. Data mining can do extravaganza on these databases. Some efforts for use of data mining in energy applications can be cited from the literature and can be traced back to 1975. A brief review of applications of clustering, support vector machines, artificial neural networks, remote sensing and other techniques is presented in (Shaheen *et al.*, 2010). A formal review of extant literature on use of data mining in different domains of energy development is given in Figure 2.

The diffusion of information technology in energy

sector of Pakistan is not representing state of the art for which the use of data mining is still beyond the expectation. There exists few software that provide analytical functions to analyze energy networks and remotely sensed data. Some are in practice for network analysis which are provided through GIS platform developed through ESRI's ArcGIS software. However, it is known that the volume of spatial data is much larger and it becomes sparser when combined with non-spatial/ attribute data for which data mining, machine learning and optimization techniques may perform in an optimal manner. Classification of energy development in a country (Shaheen *et al.*, 2011), derivation of associations among different energy network objects (Shaheen *et al.*, 2013), contextual data mining of energy scenarios, suitable energy site selection (Shaheen and Khan, 2015) are few methods of spatial and non-spatial data mining proposed by the author of this paper, which can be incorporated in the overall framework of energy development. Graphet Inc. (2000) is a company working on energy data mining and analysis. The company has been extracting useful usage patterns, sustained modes of operations and prediction of energy usage. A software product has also been developed to support the innovation. In this product, different algorithms of prediction, classification, rules mining were coded to analyze and predict different values related to energy usage patterns. Existing information systems for energy development provide very few object-oriented data mining techniques, spatial data analysis techniques, automatic extraction of knowledge and energy crisis management.

A spatial data mining based holistic framework for assessment and improvement of energy development is need of the energy sector in Pakistan. The system should be capable of answering the following and other similar questions by (1). Analyzing and optimizing existing energy network, (2). Prospecting new energy reservoirs:

1. What are energy bottlenecks and hotspots?
2. What should be the optimal energy network design to avoid energy losses and to ensure optimal energy utilization?
3. What are different patterns in failure and losses in energy network?
4. What changes should be made in existing energy network to increase energy utilization with minimum losses?



**Figure 2:** Percent contribution of data mining in different activities in energy sector.

## Materials and Methods

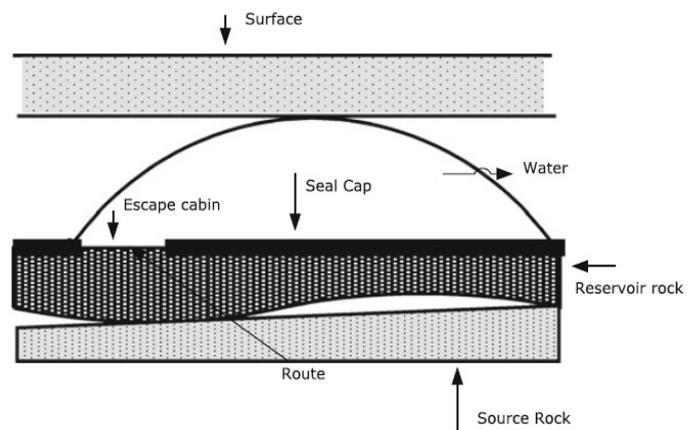
### Indicators of sustainable energy development

Sustainable development of an energy sector can now be determined with the help of a list of variables called sustainability indicators. These variables which help us to assess the development either sustainable or unsustainable are developed by the Commission on Sustainable Development (CSD) at the World Summit in South Africa in 2001. There were a few indicators (58 in number) that were proposed by UNDESA's indicators were also added to the debate. (UNDESA, 2001) A total of forty-one (41) indicators were finalized by the International Atomic Energy Agency (IAEA) in the year 2005 (Vera et al, 2005). IAEA added another important dimension of sustainability which was institutional aspect. So, the sustainability was now considered to be a balance between (i) economic; (ii) social; (iii) environmental; and (iv) institutional development of a country. The list of sustainable energy indicators in addition to the above were finalized by a joint effort made by IEA, Eurostate, UNDESA and the European Environment Agency (EEA). Different studies were made to analyze the datasets made up of these indicators by using statistical methods and tools to reach on developing an optimal energy development plan for energy sectors of different nations. A brief database, descriptions and data types of these indicators i.e. IAEA's indicators are given in (Shaheen et al., 2010).

### Data mining framework for energy management

Energy management and development should be established by following set of formal procedures and should be addressed by an interdisciplinary stance.

The exploration of energy reserves (most of the times oil and gas reserves) is done by following a series of steps specialized in its domain and need specialized human and physical resource. The prediction or prospection of energy reservoir involves huge investments of money, time and effort. The process also demands rarely available expertise of the field. The steps involved in energy reservoir prospection are not limited to manifestations of energy reserve over the chemicals and objects on the surface of earth, then its details are collected through magnetographic and seismic surveys and finally a wildcat is prospected which may or may not contain bigger energy reservoir or an empty empty vessel at all (Shaheen et al., 2011).



**Figure 3:** Energy accumulation in source rock.

About data gathering, energy data may be collected from different sources and methods. As stated earlier, remotely sensed data is collected through satellite images of particular vicinity, data related to sustainability indicators and other attributes can be collected from concerned sector and private exploration companies. The energy data collection and its compaction are nevertheless an important activity for energy optimization.

Data pre-processing is the process of making data ready for evaluation by using data mining. A number of techniques like principal component analysis (PCA), rough sets and missing values calculation are available to accurately preprocess the available data to improve its quality and to reduce redundant dimensions (Smith, 2002).

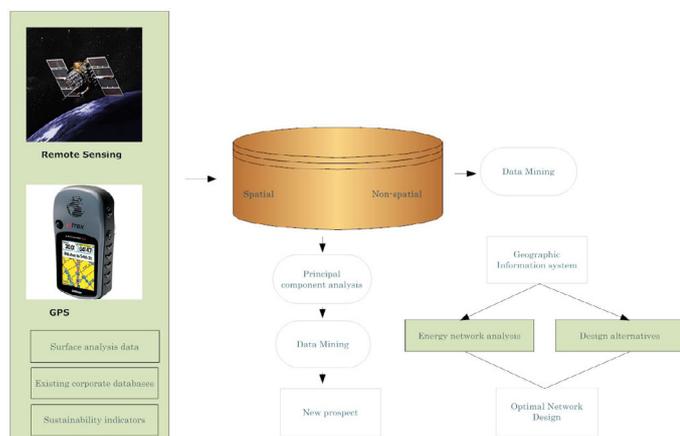
### Analysis and optimization of existing energy network

Energy crisis is caused not only by depletion of energy reserves but also by poor energy network design. Identifying bottlenecks and other critical energy network design factors and proposing design

alternatives can be considerable in optimization of existing energy networks. A Geographic Information System (GIS) with accurate GPS coordinates is a better option to do the same. Various energy parameters may be stacked in the form of layers on such GIS. Energy sustainability indicators might be placed on these geographic maps in the form of non-spatial data. Since the data of sustainability indicators is stored in the form of non-spatial or attribute data in the database, all the data mining techniques can be applied on this database.

*Prospection of new energy reservoirs*

Prospection of new energy reservoirs is based upon number of factors. Data mining is a science of discovering hidden knowledge from huge dataset so it may potentially be applied for the cause. The mining processes otherwise involved in energy reservoir prospection will reduce enormously if data mining would produce the desired accuracy in the prospection. The data sets used in prospecting are from diverse sources. Remote sensing, in last decade, proved its worth in predicting such reserves on the basis of surface indicators. These indicators manifest in different ways for which can be used to predict energy. These manifestations or indicators can become part of a database which can become part of energy data mining framework. The framework might also include data related to remotely sensed satellite images and the sustainability indicators given in previous section. These three data sets are included to set a new method for determining an initial go through for onward prospection. The dependency of initial decision on geological, geophysical survey will change to dependency on the patterns extracted from remotely sensed satellite image. The framework is given in detail in [Figure 4](#).



**Figure 4:** Data mining framework for sustainable energy development.

**Results and Discussion**

Different algorithms of spatial data mining/ machine learning are applied on spatial/ non-spatial data extracted from satellite imagery or otherwise. There are certain data mining algorithms suitable for application on diverse datasets of energy sector like PNARM ([Wu et al., 2004](#)), CBPNARM ([Shaheen et al., 2013](#)), multilevel mining ([Sharma et al., 2005](#)) and labeled clustering ([Shaheen et al., 2013](#)) etc. The data collected from diverse sources is integrated on to one platform for further evaluation and analysis. Such integration made data classification simpler and helped build a precise prediction model for energy prospection. The final GIS equipped with both the capabilities of network analysis and new prospection (data mining add-ins) is developed as a new product but could also be implemented with existing information systems.

Data mining being, an emerging discipline of IT, an explorer of knowledge from large databases, an imperative forecaster and a data analyzer can be applied on energy datasets to resolve energy crisis in a country. Existing energy networks are meant to be mapped on GIS to find goods and bads of existing network for which data miner can be integrated as a plug-in. Data mining can further reduce the human effort involved in initial prospecting of new energy reservoirs. Energy sector in Pakistan needs to adapt IT based solution for long term sustainability of energy reserves. The journey needs acceleration to quickly adapt and diffuse primary IT equipment's and techniques in the initial phase and adapt state of the art knowledge discovery techniques in lateral phase.

**Conclusions and Recommendations**

A geographic information system equipped with the capabilities of data analysis which are incorporated through data mining algorithms plug-ins is developed for sustainable energy development in energy sector of Pakistan. In order to ensure sustainable energy development, the system is built to prospect new energy reservoirs and to optimize the existing energy network design. The testing of system is yet to be reported.

In future, the system can become part of a larger electronic government system where data mining can

once again define its prominence in quick adaptability to a larger framework.

## Novelty Statement

In this paper, we proposed a data mining based framework for sustainable energy development. Pakistan's energy sector is considered as a case study for the implementation of the proposed model.

### *Conflict of interest*

The authors have declared no conflict of interest.

## References

- Bruntland, G.H., 1987. Our common future. World commission on environment and development, University Press Oxford.
- Graphet Inc, 2000. Energy data mining and analysis toolset. [http:// www.graphet.com/index.php?id=24](http://www.graphet.com/index.php?id=24)
- Shah, S. and M.K.L. Bhatti. 2009. Crisis of electrical energy in Pakistan and future guidelines for policy makers. *Int. J. Basic Appl. Sci.*, 9(9): 1-17.
- Shaheen, M. and Z. Khan. 2015. A method of data mining for selection of site for wind turbines. *Renewable and Sustainable Energy Reviews*, Accepted in Volume 48. <https://doi.org/10.1016/j.rser.2015.04.015>
- Shaheen, M., M. Shahbaz and A. Guergachi. 2013. Context based positive and negative spatio temporal association rule mining. *Knowledge-based Syst.*, 37: 261-273. <https://doi.org/10.1016/j.knosys.2012.08.010>
- Shaheen, M., M. Shahbaz, A. Guergachi and Z. Rehman. 2010. Data mining applications in hydrocarbon exploration. *Artif. Intell. Rev.*, 35(1): 1-18. <https://doi.org/10.1007/s10462-010-9180-z>
- Shaheen, M., M. Shahbaz, A. Guergachi and Z. Rehman. 2011. Mining sustainability indicators to classify hydrocarbon development. *Knowledge-based Syst.*, 24: 1159-1168. <https://doi.org/10.1016/j.knosys.2011.04.016>
- Shaheen, M., S. Iqbal and F. Basit. 2013. Labeled clustering: A method to label unsupervised classes. *Proc. 8<sup>th</sup> International Conference on Internet Technology and Secured Transactions*. pp. 210-214. <https://doi.org/10.1109/ICITST.2013.6750193>
- Sharma, L.K., O.P. Vyas, U.S. Tiwary and R. Vyas. 2005. A novel approach of multilevel positive and negative association rule mining for spatial databases. Springer Verlag Berlin, pp. 620-629. [https://doi.org/10.1007/11510888\\_61](https://doi.org/10.1007/11510888_61)
- Smith, J.I., 2002. Tutorial on principal component analysis. University of Otago New Zealand, [www.cs.otago.ac.nz/cosc453/student\\_tutorials/principal\\_components.pdf](http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf).
- United Nations Department of Economic and Social Affairs (UNDESA), 2001. *Indicators of Sustainable Development: Guidelines and Methodologies*. 2<sup>nd</sup> ed.
- Vera, I., L. Langlois and H. Rogner. 2005. Energy indicators for sustainable development. International Atomic Energy Agency IAEA, Austria, pp. 6-20.
- Wu, X., C. Zhang and S. Zhang. 2004. Efficient mining of both positive and negative association rules. *ACM Trans. Inf. Syst.*, 22(3): 381-405. <https://doi.org/10.1145/1010614.1010616>