

# American Journal of Computer Science and Engineering Survey

Review Article

## OIL EXPLORATION USING SOFT COMPUTING

ASHITH KUMAR DATTA\*

*Associate Professor, Department of Computer Science, Shaqra Univesity, Shagra, Saudi Arabia*

\*Corresponding author e-mail: [drashitkumar@yahoo.com](mailto:drashitkumar@yahoo.com)

### ABSTRACT

Soft computing (SC) techniques provide wide variety of applications in data processing, analysis and interpretation. SC play a key role in the geo sciences due to the immense size and uncertainty associated with the data. The nature of SC assesses oil industry in oil exploration and optimization of oil wells. There is a significant change in oil industry due to the complex techniques and modern equipment. Intelligent systems like Neuro computing and Artificial Intelligence are available for the exploration of oil and popular evolutionary algorithms have effective methods for the optimization of oil wells but processing of vague data create problems for the existing techniques. Uncertainty data plays a crucial role to take vital decisions. The research uses seismic data in the process of oil exploration and compared the proposed method with the existing methods and results are favorable to the proposed research.

**Keywords:** Soft computing; Oil well; Oil field; Decision Tree; J48; Data mining

### INTRODUCTION

A process of drilling in the Earth brings petroleum oil hydrocarbons to the surface and the well is termed as oil well. Modern directional drilling tools allocate for powerfully deviated wells provide sufficient depth and with the proper equipment, actually become horizontal. This is of great value as the reservoir rocks which contain hydrocarbons are usually horizontal; a horizontal wellbore used in a production zone has more surface area than a vertical well, and increase the production rate. The use of deviated and horizontal drilling tools allow production team to reach reservoirs several distance away from the location for the production of hydrocarbons located below locations that are either difficult to rig depends on the environment. The introduction of data mining (DM) in the field of computer science in late 80's lead many researches in data analysis and discovered lot of statistical and data crunching tools. The massive growth of data in all kind of business urges to use data management and data warehousing applications. The growth of data mining in

the last three decades can be divided into three parts. In the first part, programmers developed machine learning algorithms to train machines to handle huge amount of data to generate pattern from it. In the second part, many business people realized the application of DM tools and started implementing in their business and took decision according to it. In the third part, web is a huge database with semi-structured data. DM tools were implemented to study the data in the web. Web mining is the concept derived from DM and successfully used for web management. The DM is a domain expertise tool and considered one of the pioneer applications of SC. Continuous research indicate that decision tree data mining algorithm produce best results. The further part of the paper will explain the literature work based on oil exploration and results and discussion of the methods employed in the research work.

### REVIEW OF LITERATURE

A review of recent applications of soft computing in oil exploration. Artificial neural network (ANN),

fuzzy logic, probability reasoning, and Bayesian belief network were methods highlighted in the study [1]. ANN has the ability to deal with linear and non – linear problems and ideal for the oil-exploration. Fuzzy logic has the ability to manipulate symbolic information in an effective way than other methods. The concept of fuzzy logic employed for seismic data interpretation and oil reservoir litho logy identification. Data in oil exploration are dynamic and uncertain and probabilistic reasoning used to handle uncertainty in decision making. BBN is suitable for casual rules and represents probabilistic relationship. The review explained the activities involved in the oil exploration like data acquisition and pattern recognition and prediction. A research on oil exploration using big data [2]. Seismic data management and analysis were the task optimized for the methods used in the research. Semma process used to disclose patterns hidden in the large volume of data. Oil exploration using data analysis is the paramount task. Insight, predict and optimize are the processes of big data used to explore oil from the huge amount of data. Big data analysis on safety mechanism and real analysis of the exploration of oil [3]. The research employed business intelligence tools, data warehouses and other transactional applications and generated better results comparing to existing methods. Hadoop system used to derive results from websites logs and complex databases. Real time analytics and recommendations were done by the system using the big data and Hadoop. A paper on overall maintenance of oil industry. The paper described the activities involved in the maintenance of equipment by collecting data from pumps and wells then adjust the repair schedule and prevent the failures [4-6]. Big data employed in the process of optimization of production volumes.

## RESULTS AND DISCUSSION

The implementation of data analytics and prediction tool is a complex task due to scalability and time complexity. The proposed method and other methods used in the research were implemented in Java using i7 processor [7,8]. K-means, Naïve Bayes, K-NN and SVM are the methods compared with the proposed J48 methods. The algorithms were taken from Google algorithms and the dataset for the experiment were downloaded from international well data [9]. We have used two locations Saudi Arabia and Canada well data to show the ability of proposed method. Machine learning and automated tools need training to generate results, therefore during the training phase, selected data from the dataset given to the methods to learn the environment [10-12]. During the testing phase, the performance will be evaluated by calculating the time. The training phase data and (Figures 1 and 2) shows the relevant graph to the data generated during the training phase (Tables 1 and 2). J48 is the implementation of ID3 (Iterative

Dichotomiser 3) based on the classification algorithm. It has shown better training time than other methods. It has the ability to produce results in short duration with improved performance [13]. Saudi Arabia is the largest oil producer and Canada is the fifth in the world [14-16]. In the location of Saudi Arabia, the numbers of oil wells are more than the location of Canada. The training phase data shows that the employed methods had taken more time to learn the scenario and the attributes used for the training were seismic data, percentage of hydro carbon, distance from ground and latitude and longitude of the oil well. The trained attributes are vital to predict the location of oil well [17]. The testing time of the methods and the proposed method have better performance. The graph for the testing time of methods for the two locations (Tables 3 and 4). K-NN and SVM has nearest value to the proposed method but the accuracy of the methods is the better criteria to know the efficiency of the methods (Figures 3 and 4). The shows the percentage of accuracy produced by methods employed in the research [18-20]. The research work has shown overall 90% accuracy for all the methods and the proposed work has overall better efficiency than the other methods (Table 5 and Figure 5).

## CONCLUSION

SC is the combination of machine learning algorithms employed in the interest of development of application for real- world problem. The data mining algorithms were successfully implemented in all kind of business to provide decision in the complex situation. Oil exploration is the complex problem and data are vague and difficult to derive information and proposed method has achieved accuracy of an average of 92% for the dataset employed in the research. The ability of J48 to produce results achieved the better accuracy and shortest time than other methods employed in the research. The future work of the research is to expand the work for the other region in the world.

## REFERENCES

1. Chen MS, Han JW, Philip SY (1996) "Data Mining: An Overview from a Database Perspective". *IEEE Trans Knowledge Data Eng* 8: 866-883.
2. Beckman JR (1986) Model development to predict hydrocarbon emissions from crude oil storage and treatment tanks Report. Environ Prote Agen Air Resour Board.
3. Berkhin P (2002) Survey of clustering data mining techniques. Tech Rep Accrue Soft.
4. Biegert EK (2007) From black magic to swarms: hydrocarbon exploration using non-seismic technologies. EGM 2007 international workshop innovation in EM, grav and mag methods: a new perspective for exploration Capri Italy.

5. Bishop CM (1999) Neural networks for pattern recognition. 164-193.
6. Biswas G, Weinberg JB, Fisher DH (1998) ITERATE: A conceptual clustering algorithm for data mining. IEEE Trans Syst Man Cybern Part C Appl Rev 28: 219-230.
7. Bodine JH (1984) Waveform analysis with seismic attributes. Oil Gas J 84: 59-63.
8. Bott RD (2004) Evolution of Canada's oil and gas industry. Canadian Center for Energy Information.
9. Han J, Kamber M (2006) Data Mining Concepts and Techniques 4.
10. Yao YY, Zhong N, Zhao Y (2008) A Conceptual Framework of Data Mining. Studies in Comp Intel 118: 1-515.
11. Yao YY, Zhong N, Zhao Y (2001) A Three layered Conceptual Framework of Data Mining. 215-221.
12. Yao YY, Dasarathy BV (2003) "A Step towards the Foundations of Data Mining" in Data Mining and Knowledge Discovery. Theory Tools Technology 254-263.
13. Qu H, Zhao WZ, Hu SY (2006) Oil & Gas Resources Status and the Exploration Fields in China Petroleum Exploration. 4: 1-5.
14. Pan JP, Jin ZJ (2004) Potentials of petroleum resources and exploration strategy in China Acta Petrolei Sinica 25: 1-6.
15. Stundner M, Al Thuwaini JS, (2001) How Data Driven Modeling Methods like Neural Networks can help to Integrate Different Types of Data into Reservoir Management.
16. Cai YD, Gong JW, Gan IR, Yao LS (1993) Hydrocarbon reservoir prediction using artificial nerve network method. Oil Geophys Prospect 28: 634-638.
17. Camps-Valls G, Gomez-Chova L, Calpe-Maravilla J, Soria-Olivas E et al (2003) Support vector machines for crop classification using hyper spectral data. 2652: 134-141.
18. Chakarbatti D, Faloutsos C (2006) Graph mining: laws, generators and algorithms. ACM Comput Surv 38Article 2.
19. Chandra M, Srivastava AK, Singh V, Tiwari DN, Painuly PK (2003) Lithostratigraphic interpretation of seismic data for reservoir characterization. In: AAPG international conference Barcelona.
20. Chapelle O, Vapnik V, Bouquet O, Mukherjee S (2002) Choosing multiple parameters for support vector machines. Mach Learn 46: 131-159.

**Table 1.** Training Time (in seconds) for the location in Saudi Arabia.

Methods	Seismic Data	Percentage of Hydrocarbon	Distance from Ground	Latitude & Longitude
J48	0.178	0.261	0.272	0.314
K-Means	0.214	0.291	0.283	0.364
Naïve Bayes	0.192	0.242	0.286	0.412
K-NN	0.191	0.260	0.281	0.292
SVM	0.184	0.312	0.317	0.319

**Table 2.** Training Time (in seconds) for the location in Canada.

Methods	Seismic Data	Percentage of Hydrocarbon	Distance from Ground	Latitude & Longitude
J48	0.084	0.112	0.078	0.056
K-Means	0.121	0.125	0.134	0.101
Naïve Bayes	0.097	0.159	0.145	0.114
K-NN	0.187	0.145	0.089	0.077
SVM	0.107	0.127	0.095	0.081

**Table 3.** Testing Time (in seconds) for the location in Saudi Arabia.

Methods	Seismic Data	Percentage of Hydrocarbon	Distance from Ground	Latitude & Longitude
J48	0.098	0.101	0.085	0.145
K-Means	0.125	0.154	0.114	0.189
Naïve Bayes	0.137	0.138	0.142	0.189
K-NN	0.115	0.119	0.121	0.162
SVM	0.099	0.128	0.126	0.149

**Table 4.** Testing Time (in seconds) for the location in Canada.

Methods	Seismic Data	Percentage of Hydrocarbon	Distance from Ground	Latitude & Longitude
J48	0.091	0.101	0.160	0.147
K-Means	0.112	0.132	0.192	0.241
Naïve Bayes	0.101	0.212	0.242	0.312
K-NN	0.098	0.174	0.174	0.180
SVM	0.094	0.118	0.181	0.174

**Table 5.** Accuracy of results (in percentage).

Methods	Saudi Arabia	Canada
J48	95	97
K-Means	91	92
Naïve Bayes	90	90
K-NN	93	92
SVM	93	95

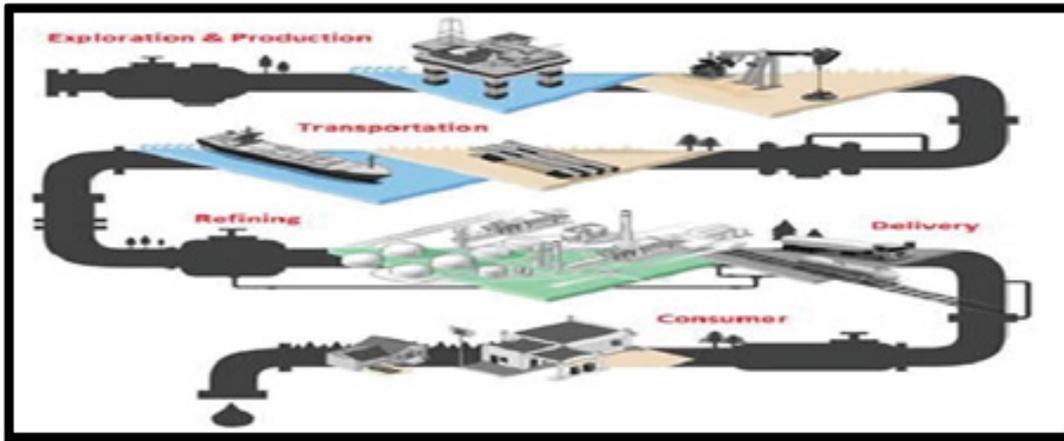


Figure 1. Training Time (in seconds) for the location in Saudi Arabia.

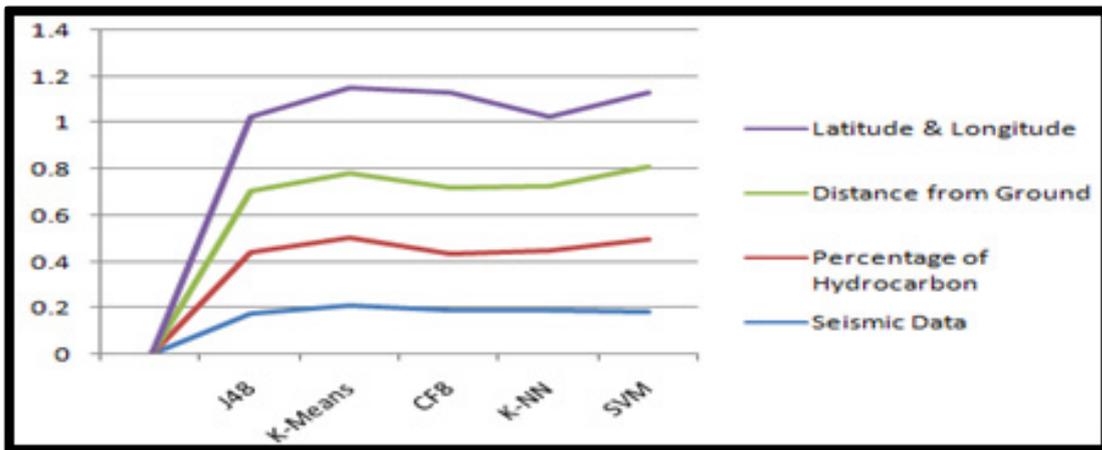


Figure 2. Training Time (in seconds) for the location in Canada.

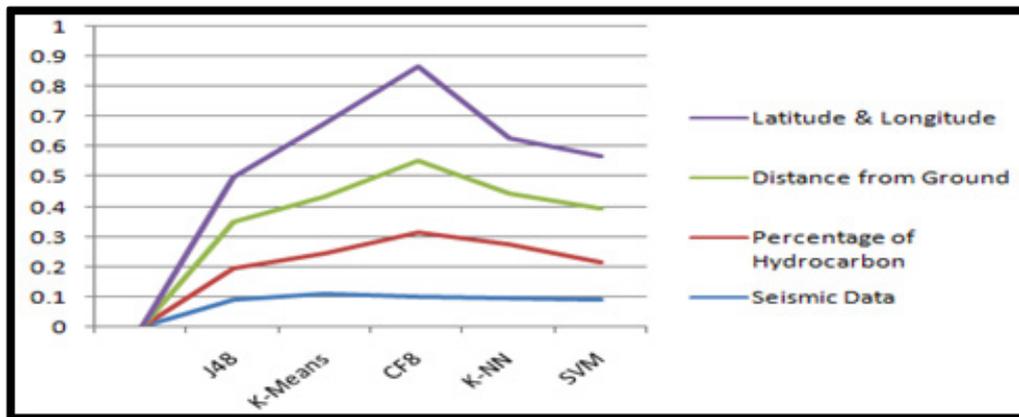


Figure 3. Testing Time (in seconds) for the location in Saudi Arabia.

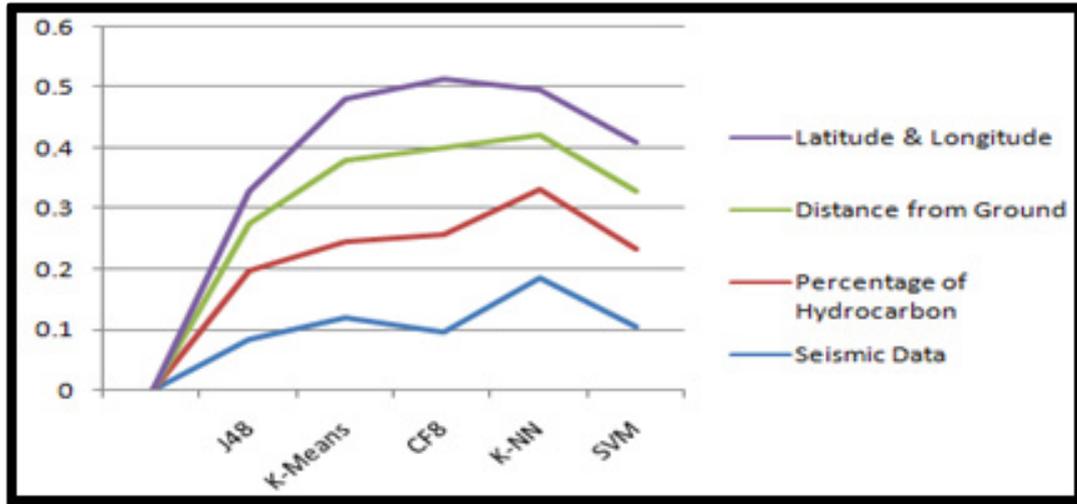


Figure 4. Testing Time (in seconds) for the location in Canada.

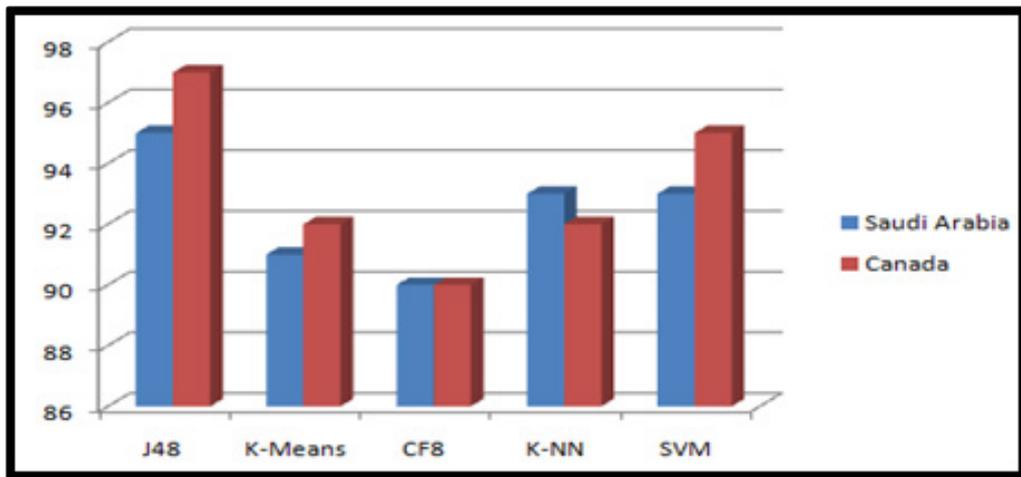


Figure 5. Accuracy of results (in percentage).