INTERDEPENDENCY OF GROUNDWATER DEPLETION USING BIG DATA ANALYTICS

[1]Mr. Sudhahar T N, [2]Ramagayathri B, [3]Rathika E, [4]Shreedevi S

[1]Assistant Professor, [2,3,4]UG Student
Department of CSE,
Agni College of Technology, Chennai

***Abstract*: Since the beginning, water has furnished us with life's essentials. Groundwater constitutes 30% of the global fresh water resource and it is a major water supply for over 2 billion people worldwide. Due to air pollution and rapid deforestation, the rate of groundwater depletion has increased. This results in water scarcity in human day to day life and agriculture, environment. Because of less number of trees in big cities, it can't bring rain which leads to lowering of water level in big cities. In the environment, we need to know how groundwater depletion is related with air pollution and deforestation. We proposed a system which predicts how groundwater level is dependent on deforestation and air pollution in the environment and the amount of danger the environment will be in the future.**

***Keywords*: deforestation, air pollution, groundwater depletion, big data analytics**

1. INTRODUCTION

For thousands of years, man has survived living in arid regions of the planet exclusively by skilfully managing that very important however scarce resource known as water. Groundwater may be a very important supply of water for domestic, agriculture and industrial use. Presently the contribution of groundwater is nearly thirty fourth of the overall annual water system. Groundwater consumption is increasing day by day thanks to continuous development of the economy, industrial enterprise, pollution, deforestation. These issues can have an effect on the approach to life of the human. So, it is necessary to associate the three staple items that any living being requires to live air, water and food. A good human life-style not solely depends on the economy, it additionally depends on their wellness. The levels of deforestation, pollution have to be controlled in order so that the living beings can still survive within the earth. It is vital to predict and maintain the ground water level. This prediction results in the quality of lifestyle of the individuals in the future. This paper focuses on the gathering of the large quantity of data, pre- processing it, building predictive models to predict the standing of what the amount of groundwater level can be within the close to future based on the amount of air pollution and deforestation.

2. RELATED WORKS
LITERATURE SURVEY

Meghana P. Patel has incontestable environmental condition and phylogeny impact on groundwater quality of agriculture dominated areas of southern and central Gujarat, Bharat: It describes changes in groundwater hydrochemistry within the south and central a part of Gujarat state in India to spot vulnerable hotspots beneath seasonal differences and phylogeny activities. Groundwater and soil samples were collected. The samples were analyzed for physico-chemical characteristics. Groundwater quality of Vadodara, Bharuch, and Tapi districts were plagued by geologic and phylogeny activities whereas spring water in Vansda and Ahwa were less affected because of restricted industrial enterprise similarly as geographical location. Even handed management of economical irrigation, industrial waste treatment, trendy agricultural practices, fresh water harvest home to conserve groundwater quality of study area units are important.

Jinghan Wang has uncontestable associate degree Air quality knowledge analysis and prognostication platform supported massive data: these days, with the continual development of massive knowledge technology, varied industries use massive knowledge technology to method and mine large knowledge, and notice the worth of information with efficiency. In terms of air quality processing, massive knowledge technology also can play a precise advantage. The platform is predicated on massive knowledge technology to style associate degree air quality knowledge analysis and prediction platforms together with knowledge layer, business layer, interaction layer and mental image platform. This style may be an important application for absolutely exploiting environmental knowledge info. it's powerful processing functions and measurability, that may be a reliable knowledge analysis and prediction platform.

Weitao Zou has an uncontestable survey of massive knowledge Analytics for sensible Forestry: correct and reliable biological science knowledge may be obtained by suggesting that of continuous watching of forests victimising advanced technologies, that provides a significant chance for the event of sensible biological science. However, with the advance of the exactitude and acquisition speed of information, the standard knowledge analysis and storage technology cannot meet the performance needs of current applications. During this paper, we have a tendency to summarize the analysis and work of the large knowledge in sensible biological science in recent years. Firstly, we have a tendency to review the history of the emergence and development of biological science and massive knowledge, then in short summarize the opportunities delivered to biological science by massive knowledge

technology. one in all the foremost necessary tasks of biological science, massive knowledge is to prepare the huge knowledge moderately and effectively and to calculate quickly. Therefore, we have a tendency to propose a five-layer design model of biological science massive knowledge, and summarize the connected work of information storage, query, analysis and application. Finally, the challenges of biological science, massive knowledge area units analyzed, and also the trend of future development is projected from 3 aspects.

Yue Tai Yangtze Kiang has a giant knowledge platform for air quality analysis and prediction: He has planned a linguistics ETL (Extract-TransformLoad) framework on cloud platform for AQ analysis and prediction. He exploited metaphysics to concretize the link of PM a pair of.5 from varied knowledge sources and to merge those knowledge with constant conception however completely different naming into the unified info. The computing nodes area unit wont to execute data processing algorithms for predicting AQ, and storage modes area unit wont to store retrieved, preprocessed, and analyzed knowledge. He exploited browsers to point out the visualised result to demonstrate the estimation and prediction. It shows that the large knowledge access framework on the cloud platform will work well for air quality analysis.

3. ALGORITHM

In the existing methods, the prediction is done for ground water level based on the and that it needs to be handled both individually and collectively. Air pollution, deforestation and groundwater depletion are all an interconnected web of problems. When one is tackled the other problem may rise and vice-versa. So, there needs to be a system which shall collectively analyze each of these problems and then predict what the environment's status will be in the near future. Thus, we can take care of the situation before it goes out of our historical data of the ground water level and it is not associated with the levels of deforestation and air pollution. The environment and the people that live in it are affected by the amount of ground water level, deforestation and air pollution. So, these need to be controlled and tackled. There are existing solutions for all these problems separately. But, what we need to keep in mind is that each of the environment's problems are interrelated.

We proposed this system to overcome the demerits of the existing methods. We aim to create a system which can be accessed by both government and public. We not only predict data but also compare and interconnect the data about air pollution, groundwater depletion and deforestation to predict the future status of the environment using predictive models in Big Data Analytics. We have used the multiple linear regression model to predict the level of ground water level because Multiple Linear Regression (MLR) can be used to make a linear empirical relationship between air pollution, groundwater depletion and deforestation.
A forecast/prediction of air pollutants may be created through an equation during which unknown variables may be expressed as a performance of an exact variety of well-known variables. there's one variable quantity to be foretold in relevance the 2 or a lot of freelance variables. the overall kind of MLR may be expressed as
Y= b1+b2X2+…..+bk Xk e
where Y depends variable, X2, X3…....., Xk area unit freelance variables, b1, b2…....., atomic number 97 area unit rectilinear regression parameters ANd e is an calculable error term, that is obtained from freelance sampling from the traditional distribution with mean zero and constant variance. The aim of regression modeling is to estimate the b1, b2....., bk, which might be created as a victimization minimum sq. error technique. therefore we tend to conclude that this technique is ready to predict what the standing of the surroundings are in the longer term. This technique is simpler to use and safer.

4.SYSTEM ARCHITECTURE

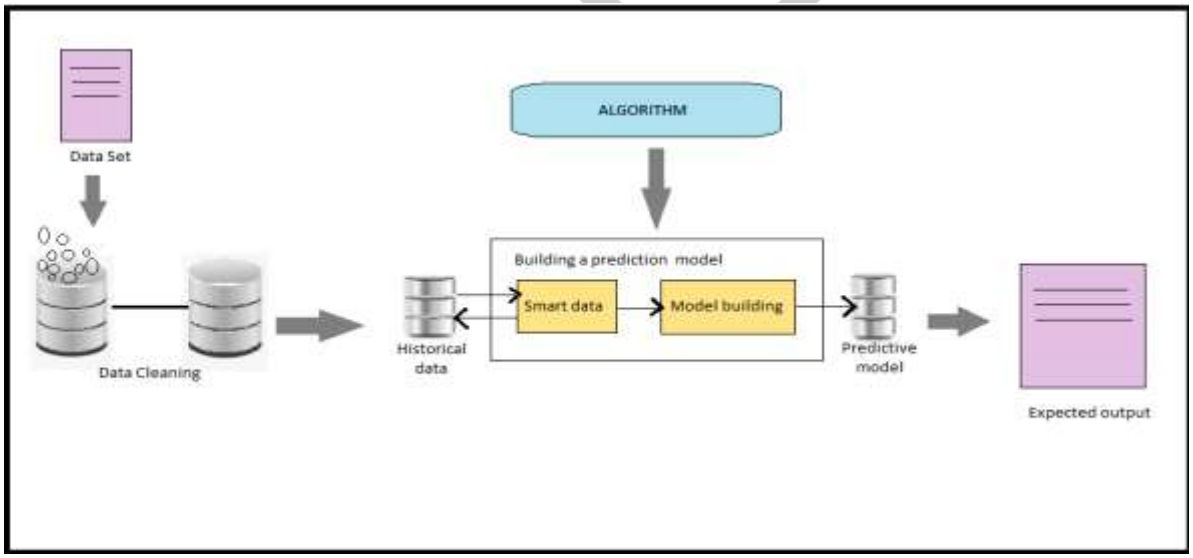

Fig .1 SYSTEM ARCHITECTURE

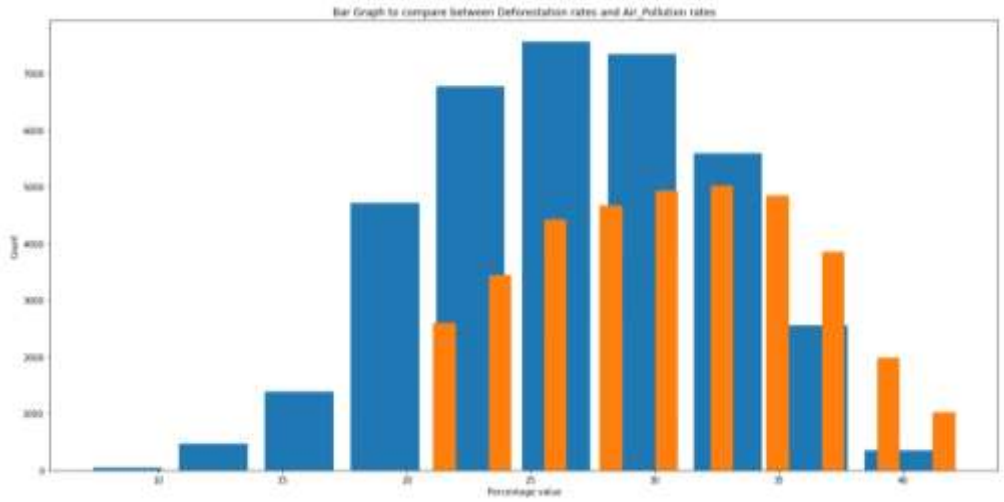

Fig.2 COMPARISON BETWEEN DEFORESTATION AND AIR POLLUTION

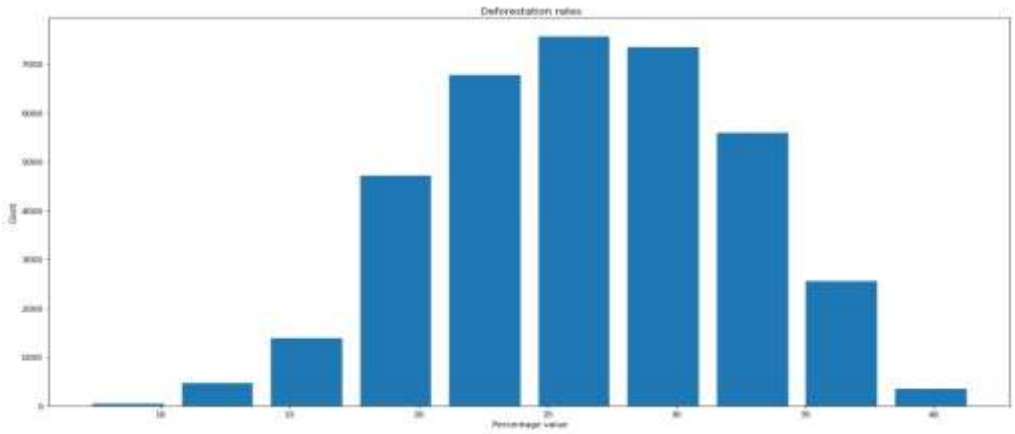

Fig.3 DEFORESTATION RATE OVER THE YEARS

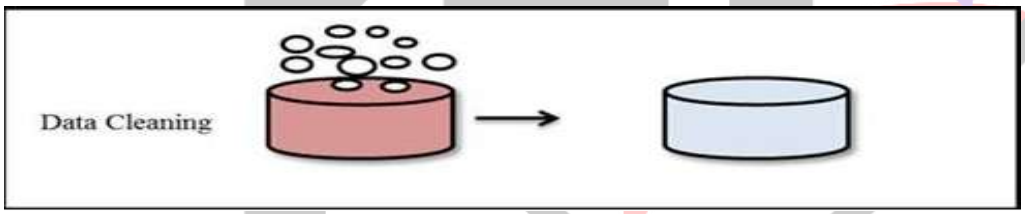

Fig.4 DATA CLEANING PROCESS

5. SYSTEM MODULES

A.AIR POLLUTION MODULE

In this module, we collect a large amount of data of air pollution in the environment. There is a chance to find some data are missing in the dataset or some unwanted data are in the dataset. We need a dataset which does not include any outliers and others. So, we use the data cleaning method to remove the outliers. Then this module preprocesses the data and displays the predicted data about air pollution of the environment.

B.DEFORESTATION MODULE

In this module, we collect a large amount of data of deforestation in the environment. There is a chance to find some data are missing in the dataset or some unwanted data are in the dataset. We need a dataset which does not include any outliers and others. So, we use the data cleaning method to remove the outliers. Then this module preprocesses the data and displays the predicted data about air pollution of the environment.

C.GROUNDWATER DEPLETION MODULE

In this module, we collect a large amount of data of groundwater depletion in the environment. There is a chance to find some data are missing in the dataset or some unwanted data are in the dataset. We need a dataset which does not include any outliers and others. So, we use the data cleaning method to remove the outliers. Then this module preprocesses the data and displays the predicted data about air pollution of the environment.

D.FINAL MODULE

In this module, the predicted output is displayed to describe the status of the ground water level which is predicted by associating with deforestation and air pollution . It is predicted using Multiple linear regression predictive models of Big Data Analytics. This module helps to predict the future status of the environment and how to tackle it efficiently.

6. METRICES

Even if cash was out there, most of the technologies are still in their pre-commercial stage and there are several considerations concerning information privacy, security and accuracy. ancient storage will value loads of cash to store huge information. several huge information is unstructured.This methodology overcomes all the demerits of existing ways. This methodology can provide a result or foreseen information with ninety fifth accuracy.

7. CONCLUSION AND FUTURE WORKS

Prevention is better than cure. Instead of worrying about a Disastrous event that might occur in the near future it's better to take proper measures to prevent it from occurring. Our system provides a way to do this, with the help of the predictive models, we're able to predict what the status of the environment will be in the future. Better ways to enhance the environment is possible, as the problem is tackled before it gets too large to handle.The proposed system is used to inter-relate groundwater level with air pollution and deforestation. It is used to predict the ground water level of the future based on historical data and dependencies with air pollution and deforestation. We collect a large amount of data and clean that data. Then the cleaned data pre-processed by predictive modelling. It provides an insight of what problems that may occur if the current situation continues. In future we will conduct more experiments using big data analytics to give the best and more accurate data.

REFERENCE

[1] Wu J, Guo S, Li J, et al. Big data meet green challenges: Big data toward green applications[J]. IEEE Systems Journal, 2016, 10(3): 888900.
[2] Ayyalasomayajula H, Gabriel E, Lindner P, et al. Air quality simulations using big data programming models[C]//2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService). IEEE, 2016: 182-184.
[3] Prediction of Ground Water Level Based on DE-BP Neural Network by lishan Ma, Dekui Yuan, Jianhua Tao lishan Ma, Dekui Yuan, Jianhua Tao
[4] Climatic and anthropogenic impact on groundwater quality of 1 agriculture dominated areas of southern and central Gujarat, India by Meghana P. Patel, Bharat Gami, Akash Patel, Pankaj Patel
[5] K. Wang, H. Li, Y. Feng, and G. Tian, "Big Data Analytics for System Stability Evaluation Strategy in the Energy Internet", IEEE T. Ind. Inform., vol. 13, no. 4, pp. 1969-1978, Apr. 2017, DOI: 10.1109/TII.2017.2692775
[6] B. Shangguan, P. Yue, Z. Wu, and L. Jiang, "Big spatial data processing with Apache Spark", presented at Int. Conf. AgroGeoinformatics, Fairfax, VA, USA, Aug. 7-10, 2017
[7] Mohammad Amin Bhat, M. G. (2016). Assessment of Groundwater Quality for Irrigation Purposes using Chemical Indices. Indian Journal of Ecology, 43(2), 574-579.
[8] C.D. Modhera, G. R. (2017). Groundwater Quality Assessment for Irrigation Water Use in Mahi Right Bank Command Area, Gujarat, India. In ICRISET (Ed.), Kalpa Publications in Civil Engineering, Vol.1 (pp.343-351). vidyanagar, Gujarat: InternationalConferenceonResearch and Innovations in Science, Engineering &Technology
[9] Zhang, D., Pan, S. L., Yu, J., & Liu, W. (2019). Orchestrating Big Data Analytics Capability for Sustainability: A Study of Air Pollution Management in China. Information & Management, 103231. doi:10.1016/j.im.2019.103231
[10] Miah, S.J., Vu, H.Q., Gammack, J., and McGrath, M., A big data analytics method for tourist behaviour analysis. Information & Management, 2017. 54(6): p. 771-785
[11] Optimised Big Data analytics for health and safety hazards prediction in power infrastructure operations Anuoluwapo Ajayi, Lukumon Oyedele , Olugbenga Akinade, Muhammad Bilal, Hakeem Owolabi, Lukman Akanbi, Juan Manuel Davila Delgado Big Data, Enterprise and Artificial Intelligence Laboratory (Big-DEAL), University of the West of England, United Kingdom,https://doi.org/10.1016/j.ssci.2020.104656