IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693

Correlation analysis on Information retrieved from upstream segment production data in O&G Industry

A. K. Kavuru[1*], R. J. Ramasree[2], Md. Faisal[3]

[1]Department of Computer Science and Applications, SCSVMV University, Kanchipuram, India
[2] Department of Computer Science, RSVP, Tirupathi, India
[3] Department of Information Technology, Nizwa College of Technology, Nizwa, Sultanate of Oman

[*]Corresponding Author: arunudu@gmail.com, Tel.: +91-9494404362

Available online at: www.ijcseonline.org

Received: 21/07/2018, Published: 31/July/2018

***Abstract*—** Among the three segments in Oil and Gas industry, where huge data is produced every day, upstream segment and specifically production area is moving towards complete automation with the help of technologies such as IIOT and big data. The Oil and Gas Upstream production involves many sensors to collect the data that are attached to all the wells, and this data is processed through IOT Gateways into traditional servers. Such data can be analyzed using multiple analytics to predict and monitor so many components such as the temperature growth, water rate, gas rate, and oil rate etc., The current paper is to analyze the data produced from 5 different wells. Two months data is collected, summarized to find the correlation between the oil, water and gas produced from each well. The purpose is to find the relationship between each of the components. This helps in predictive maintenance and also gives information on how much oil and gas can be produced. When compared with the historical data, if the correlation coefficient is changing abnormally at any specific point the monitoring team must investigate and do proper maintenance.

***Keywords*—**IIOT- Industrial Internet of Things, IOT – Internet of Things, O&G - Oil & Gas, Corrleation

I. INTRODUCTION

The world is moving towards complete automation and high productivity. Industry 4.0 is the major technological industrial revolution and is the future of many industries such as oil and gas, manufacturing, medical monitoring, transport etc.,
Oil and gas is one of the major industries in the world. The major sectors of Oil and Gas industry can be summarized majorly as three streams such as Upstream, Midstream and downstream (Figure 1: Oil & Gas Industry Streams) [1].

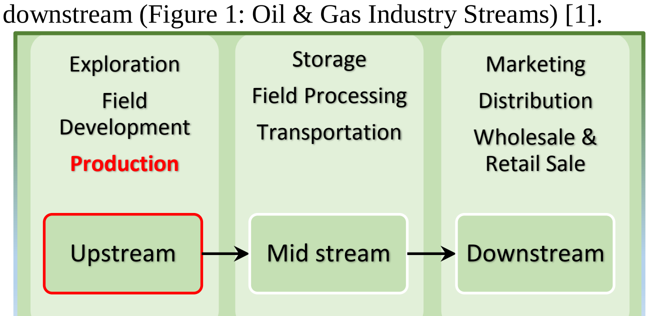

Figure 1: Oil & Gas Industry Streams

The upstream sector is the major sector, where the searching for oil and gas takes place. This process is called exploration. During exploration, the companies try to search for underground, or under water oil which is crude in nature and natural gas, by drilling wells. These wells are of two types. The first one is called exploratory well, used to explore the oil and gas. The second one is called operating wells, used to produce the crude oil and natural gas.
The midstream sector or middle stream sector is mainly used for transportation, field processing and storage. The downstream is the last stream where the distribution of oil & natural gas takes place, Involving marketing, wholesale and retail sales.
Implementing IIOT in Oil & Gas industry has been started long ago in all the streams & their functions. Still there is a lot of scope for research in terms of predictive analysis, diagnostic analysis, and descriptive analysis by applying big data technologies.
The current paper discusses on the different upstream associated components, especially in production area, that are used in oil and gas industry while transforming into Industry 4.0. It also focuses on one of the use cases where the well data is processed and analyzed.
With the advancement in technology, sensors were used with many of the machinery which has given required data for advanced analysis, (for example using Big Data analytics) from live data as well as historical.

The sensors used in upstream – production area not only just reads the relevant data but also helps in analyzing temperature, flow rate, pressure and solves many of the other problems. Sensors also help in infrastructure surveillance while producing the crude oil or natural gas.
Some of the sensors used in the upstream - production area is –

- Fiber optic Point sensors to check pressure and temperature
- Fiber optic distributed temperature sensors (DTS)
- Fiber optic sensors for infrastructure surveillance

The data taken from these sensors were generally collected through some devices that are famously known as IOT Gateways and then are loaded into traditional servers. The following architecture describes the total process.

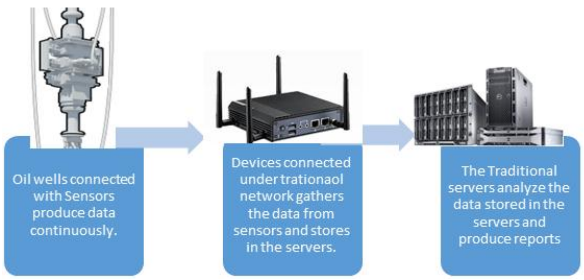

Figure 2: Architecture of current Oil and Gas environment in upstream production area

There is a huge scope of research with the data available in oil and gas and can be used for analysis and predictions. In the current use case, the data collected from the sensors is taken for correlation analysis.
The next part of this paper is organized as follows. Section II contains the related work done in oil and gas industry, based on sensors applied on the wells and their data collection and storage methods, implementing IOT and Big data and also on different types of analytics used on the data to produce relevant output. Section III discusses the theoretical aspects and practical implementation. Section IV provides the methodology followed. Section V gives an overview of the results produced with proper discussion. Section VI concludes work done in this paper along with the future scope also mentioned.

II. Related Work

1. Robert Karlsson et al., 1996:- presented experimental design for kinetic analysis of protein. The use of a reference surface made it possible to identify conditions where a response related to changes in matrix conformation could be ignored. When data was inconsistent with a one to one reaction, it was possible to obtain good fits to an entire data set assuming several other reaction schemes including parallel, competitive and two-state reactions. This analysis helped to understand biometric sensors.
2. Abdulelah Bin Mahfoodh et al., 2017: have introduced a system for monitoring and maintaining well data quality and integrity using different types of sensors and big data systems. The paper also uses the big technologies and techniques such as Hadoop, Map reduce etc.,
3. Mohammed Y Aalsalem et al., 2018: - have presented needs and requirements of Wireless Sensor networks in Oil and Gas industry. WSN applications are necessary for Oil & Gas industry to perform operations on two major segments such as "Process Control" and "Factory Automation". In upstream sector, the WSN is used to monitor the operations happening in exploration and production sectors.
4. Qian Zhu et al., 2010: - have proposed IOT Gateway system based on GPRS protocols, to have data transmission between Wireless sensors networks and mobile communication networks. The applications of IoT cover logistics, intelligent building systems etc., IOT gateway works as a bridge to connect sensor networks with traditional communication networks.
5. Chien-Chi Kao et al., 2017: have proposed comprehensive study on sensors that work under water. Internet of Underwater things is a new concept for underwater wireless sensor networks and their applications. These applications are helpful in operations such as environmental monitoring, underwater exploration, disaster prevention etc.,
6. Nader Mohamed et al., 2008: have proposed fault tolerant sensor network architecture for monitoring pipeline infrastructures. This architecture helps in efficient power management, and efficient routing for wireless sensor nodes to extend the life of network.
7. Pan Yi et al., 2010: have proposed a monitoring system for production well based on wireless sensor networks, where the data can be browsed on internet or by mobile phone and receive alerts whenever there is some fault such as overflow, high temperature etc.,
8. Tarek R Sheltami et al., 2016: have provided the popular software methods from the survey conducted for pipeline leak detection, pressure point analysis etc., Some of the major methods suggested were negative pressure wave method, Mass balance method, pressure point analysis, real time transient modelling and statistical methods for detecting leaks.
9. Wazir Zada Khan et al., 2017: have proposed a novel IOT based architecture for data collection from connected objects in all three categories of operations in O&G, i.e., in Upstream, midstream and downstream. This architecture has got four layers where in the first layer the data from O & G equipment such as pipelines, pumps, well heads, storage tanks etc., were shown. The second layer is shown with smart objects such as applications, long and short-range communications and

Sensors.
The third layer is the gateway with applications, long and short-range communications and Sensors. The last layer is shown as Server / control center with Management applications database servers and middle / long range communications.

10. Mohammed Y Aalsalem et al., 2017: have proposed an intelligent IoT based monitoring system which involves smart objects for reliable and efficient monitoring of O&G wells. According to the proposed architecture the O & G wells are connected with smart objects that can take decisions on their own without human intervention. A Smart Gateway is used as a bridge between the control center / Server and the smart objects. These gateways are also equipped with different types of sensors to measure temperature, flow and pressure sensors. This system can trigger alarms on detection of anomalies.
11. Gianmarco De Francisci Morales et al: have introduced mining with IOT big data stream learners for classification, regression, clustering and pattern mining. The different scalability issues in IOT applications and data streams on distributed engines such as spark, amza etc., were also discussed.
12. Hossein Hassani et al: have conducted research on the industry's engagement and adoption of big data in all streams operations. The research in particular to upstream sector is actively engaging with big data to achieve efficiency gains and finds solutions to its aching financial and productivity issues.

III. Tehoretical Aspects and Practical Implementation

Correlation Analysis:
Correlation analysis is a statistical method that helps the organization to study the strength of relationship between two or more attributes. Redundancy may be caused due to inconsistency in attributes and such redundancies are identified with the help of correlation analysis.
In general correlation coefficients are shown in a matrix whose value lies between -1 to 1. If the correlation coefficient is between 0 & 1 it is known as positive correlation or between 0 & -1 known as negative correlation.
If the correlation coefficient is 1 then it is considered to be as perfect positive correlation, or if it is -1 then it is considered to be as perfect negative correlation. If the correlation coefficient is between -1 to -0.75 or 1 to 0.75, it is considered as higher negative correlation or higher positive correlation respectively. If the correlation coefficient is between -0.75 to -0.5 or 0.5 to 0.75, it is considered as moderate negative correlation or moderate positive correlation respectively. If the correlation coefficient is between -0.5 to -0.25 or 0.25 to 0.5, it is considered as lower negative correlation or lower positive correlation respectively. If the correlation coefficient is 0 then it is considered to be as no correlation.
Let us consider X as the Average Oil Rate and Y as the Average Gas Rate. Let us consider the sample data as shown below.

X	Y
188.1256	11019.52
197.9145	11604.88
206.4205	12113.51
211.2673	12403.34
215.9054	12680.69
218.9557	12863.08

To find the correlation between the two variables, the famous statistical correlation formula considering the method actual mean method, is as shown below.

$$r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}}$$

Where X = $(x - \bar{x})$ and Y=$(y - \bar{y})$ i.e,

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$$

where r is the correlation coefficient, x is (X-$\bar{x}$) where $\bar{x}$ is the mean of X and y is (Y-$\bar{y}$) where $\bar{y}$ is the mean of Y.
After completing the calculations, r=0.898864171. So as mentioned above the correlation is higher positive correlation.
Rapid Miner Studio 8.0 has been used for practical implementation, application having the features to use the correlation matrix operator to perform on huge data set.
In this research article data has been collected from 5 wells from different geographical locations to find out the correlation among the attributes.
The following methodology has been followed and implemented based on the data collected from anonymous oil and Gas Companies from multiple wells distributed in different geographical locations.

III. 1 Business understanding

Acquire the data from the sensors of the wells.
The data collected from the sensors from different wells are taken into a data source. In the current use case it is planned to find correlation between the average oil, average gas and average water produced from a particular well. To perform this task approximately 60 days data is used.

III. 2 Data Understanding

DayTime: The date and time of the average data collected.
AvgOilRate: The average oil produced on a particular day from the well.
AvgGasRate: The average Gas produced from the same well on a particular day.

AvgWaterRate: The average water came out from the same well on a particular day.

III. 3 Data Preparation

Understand the data and prepare the final data set.
The data coming from sensors is huge and not processed. As an initial step the data is cleansed and then compiled into average rates.

ExampleSet (61 examples, 0 special attributes, 5 regular attributes)

Row No.	Well Name	Daytime	Avg Oil Rate	Avg Gas Rate	Avg Water Rate
15	D	Oct 15, 2016 12:00:00 AM GST	188.126	11019.525	488.786
16	D	Oct 16, 2016 12:00:00 AM GST	197.915	11604.878	514.593
17	D	Oct 17, 2016 12:00:00 AM GST	206.420	12113.511	537.018
18	D	Oct 18, 2016 12:00:00 AM GST	211.267	12403.338	549.796
19	D	Oct 19, 2016 12:00:00 AM GST	215.905	12680.688	562.024
20	D	Oct 20, 2016 12:00:00 AM GST	218.956	12863.084	570.065
21	D	Oct 21, 2016 12:00:00 AM GST	197.052	11553.309	512.319
22	D	Oct 22, 2016 12:00:00 AM GST	215.584	12661.472	561.176
23	D	Oct 23, 2016 12:00:00 AM GST	190.280	11148.370	494.466
24	D	Oct 24, 2016 12:00:00 AM GST	191.632	11229.225	498.031
25	D	Oct 25, 2016 12:00:00 AM GST	191.501	11221.364	497.685
26	D	Oct 26, 2016 12:00:00 AM GST	88.720	5075.295	226.715
27	D	Oct 27, 2016 12:00:00 AM GST	0	0	0
28	D	Oct 28, 2016 12:00:00 AM GST	0	0	0
29	D	Oct 29, 2016 12:00:00 AM GST	167.778	9802.818	435.143
30	D	Oct 30, 2016 12:00:00 AM GST	0	0	0
31	D	Oct 31, 2016 12:00:00 AM GST	0	0	0

Figure 3: Data got imported into the Rapidminer project

As a lateral step, the following fields are considered to use in the model, to find the correlation.
Well name: Each well is given a name to identify while doing the analysis.

III. 4 Model

To identify the correlation among the fields selected, one of the famous data analysis tool, Rapidminer is used with the CRISP statistical model.
Initially the data is imported into the project. Data filtration can be done by selecting the required attributes and excluding the unwanted data and managing the missing values. Then the correlation matrix can be created by including the correlation functionality in the tool and then deploy the project to get the required result. The complete process is shown in the picture given below.

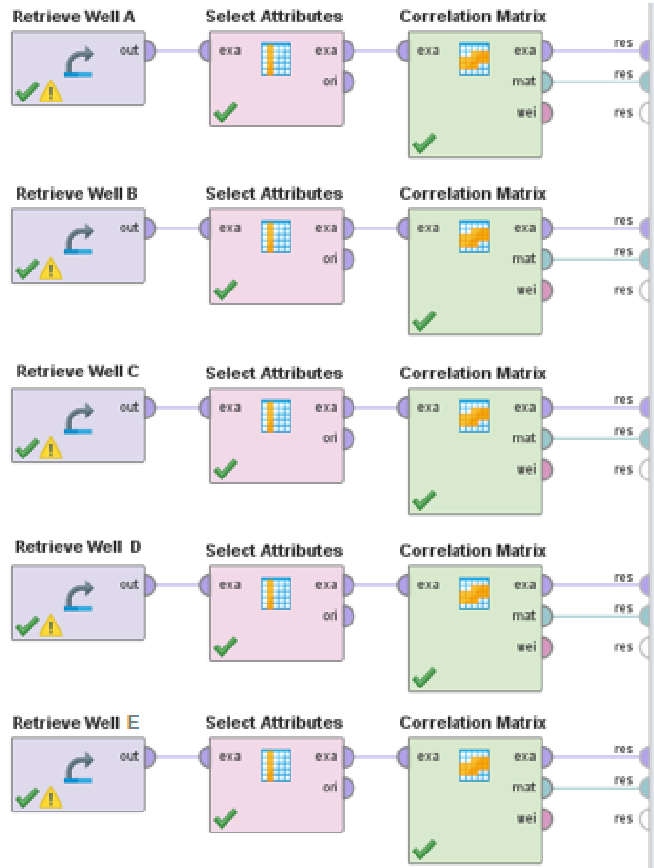

Figure 4: Rapidminer model to create Correlation Matrix from multiple wells.

IV. METHODOLOGY

The proposed research article is based on the qualitative research where researcher has to emphasize on a problem unlike the quantitative methodology. CRISP methodology has been chosen to elaborate the context-based problem into the distinct phases towards to achieve the objective of the project. There are six phases involved into the CRISP methodology as follows:

a. Business understanding
b. Data understanding
c. Data preparation
d. Modeling
e. Evaluation
f. Deployment

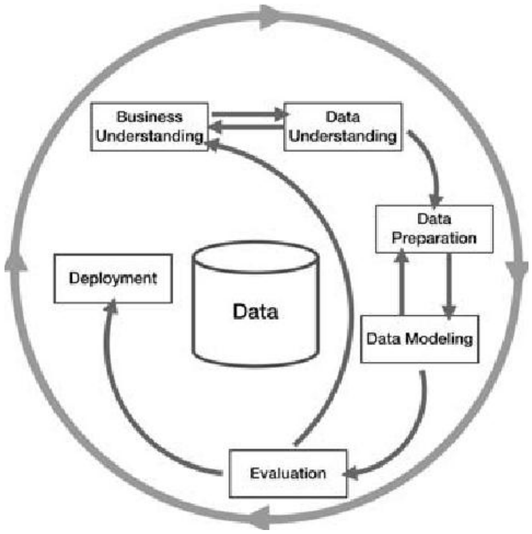

Figure 5: CRISP Model followed in finding the correlation matrix.

In the current use case, the data from the wells which are geographically located in multiple locations were collected, to find the correlation among the different coefficients such as average oil rate, average water rate and average gas rate.

V. RESULTS AND DISCUSSION

V. Evaluate the data and process the results:
Figure 6, shows the correlational matrices produced from each of the well taken as sample. These matrices are produced with the help of Rapidminer software. As explained in Section III, the correlation matrix shows the following are the key correlations observed in each of the wells.
The following table shows the correlation based on Average Oil Rate, with Water and Gas.

Well	Type of correlation	With Water	With Gas
A	Perfect Correlation	1	1
B	Perfect Correlation	1	1
C	Positive Moderate Correlation	0.462	0.462
D	Perfect Correlation	1	1
E	Perfect Correlation	1	1

Table 2: Results of the correlation based on Oil with Water and Gas

Well A

Attributes	Avg Oil Rate	Avg Gas Rate	Avg Water Rate
Avg Oil Rate	1	1.000	1.000
Avg Gas Rate	1.000	1	1.000
Avg Water Rate	1.000	1.000	1

Well B

Attributes	Avg Oil Rate	Avg Gas Rate	Avg Water Rate
Avg Oil Rate	1	1.000	1.000
Avg Gas Rate	1.000	1	1.000
Avg Water Rate	1.000	1.000	1

Well C

Attributes	Avg Oil Rate	Avg Gas Rate	Avg Water Rate
Avg Oil Rate	1	1.000	0.462
Avg Gas Rate	1.000	1	0.462
Avg Water Rate	0.462	0.462	1

Well D

Attributes	Avg Oil Rate	Avg Gas Rate	Avg Water Rate
Avg Oil Rate	1	1.000	1.000
Avg Gas Rate	1.000	1	1.000
Avg Water Rate	1.000	1.000	1

Well E

Attributes	Avg Oil Rate	Avg Gas Rate	Avg Water Rate
Avg Oil Rate	1	1.000	1.000
Avg Gas Rate	1.000	1	1.000
Avg Water Rate	1.000	1.000	1

Figure 6: Correlational Matrices

Based on the observation, currently the correlation among the attributes is either Perfect or positive. In any case if there is any change of correlation, this may be either because of sensor malfunctioning or the actual change occurred while producing the oil, which needs to be verified and investigated by concern teams.

VI. CONCLUSION and Future Scope

According to multiple online sources, only three percent of data in the world is used in a meaningful way. Further research in this area will allow the world to use the data even more usefully.
Currently the data is getting stored in anonymous servers and it is planned to further extend the study by introducing the big data architecture replacing the current architecture.
This improves the possibility of applying more and more analytics that helps in predictive analysis and maintenance.
The analysis can be improved with the use of Big Data analytics such as FLUME / KAFKA to fetch the data directly from the sensors through IOT Gateway and can store in Hadoop distributed file system HDFS. Apache Spark Streaming can be used to fetch the live data and finally machine learning algorithms are required, that can be facilitated by using Spark MlLib.

VII. REFERENCES

[1] August 2006 - Global Industry classification Standard (GISC®) – by Stanford and Poor
[2] Abdulelah Bin Mahfoodh, Mohamad Ibrahim, Maan Hawi, Khalid Hakami (2017). "Introducing a Big Data System for Maintaining Well Data Quality and Integriyt in a World of Heterogeneous Environment". Society of Petroleum Engineers, SPE – 188082 – MS
[3] Mohammed Y Aalsalem, Wazir Zada Khan, Wajeb Gharibi, Muhammad Khurram Khan, Quratulain Arshad (2018), "Wireless Sensor Networks in Oil and Gas Industry: Recent Advances, Taxonomy, Requirements and Open Challenges". Journal of Network and computer applications.
[4] Qian Zhu, Ruicong Wang, Qi Chen, Yan Liu and Weijun Qin (2010), "IOT Gateway: BridgingWireless Sensor Networks into Internet of Things". IEEE/IFIP International Conference on Embedded and Ubiquitous Computing.
[5] Chien-Chi Kao, Yi-Shan Lin, Geng-De Wu and Chun-Ju Huang (2017), "A Comprehensive Study on the Internet of Underwater Things: Applications, Challenges, and Channel Models". MDPI;
[6] Nader Mohamed and Imad Jawhar (2008), " A Fault Tolerant Wired/Wireless Sensor Network Architecture for Monitoring Pipeline Infrastructures ", The Second International Conference on Sensor Technologies and Applications
[7] Pan Yi, Xiao Lizhi, Zhang Yuanzhong (2010), "Remote real-time Monitoring System for Oil and Gas Well Based on Wireless Sensor Networks", IEEE.
[8] Tarek R Sheltami, Abubakar Bala, Elhadi M Shakshuki (2016), "Wireless Sensor Networks for leak detection in pipelines-Survey", Springer – Verlag Berlin Heidelberg
[9] Wazir Zada Khan, Mohammed Y Aalsalem, Muhammad Khurram Khan, Md. Shohrab Hossain and Mohammed Atiquzzaman (2017) , "A reliable Internet of Things based on Architecture for Oil and Gas Industry", ICACT
[10] Mohammed Y Aalsalem, Wazir Zada Khan, Wajeb Gharibi and Nasrullah Armi (2017), "An intelligent oil and gas well monitoring system based on internet of things" IEEE
[11] Gianmarco De Francisci Morales, Albert Bifet, Latifur Khan, Joao Gama and Wei Fan (2016) "IOT Big Data Stream Mining", KDD, ACM.
[12] Hossein Hassani and Emmanuel Sirimal Silva (2018), "Big data: a big opportunity for the petroleum and petrochemical industry" OPEC.

Authors Profile

Mr. A K Kavuru pursued Bachelors of Computer Science from Nagarjuna University in the year 1997, Masters in Computer Applications in AY 2000 from Madras University and M.Phil in computer science from Periyar univeristy in the year 2013. He is currently pursuing Ph.D. from SCSVMV university, Kanchipuram, India and working as Lecturer in IT under Department information technology, Nizwa college of technology, Nizwa, Sultanate of Oman. His main research work focuses on Big Data Analytics, Data Mining, IoT and Information Retrieval. He has 16 years of teaching experience and 6 years of Research Experience.

Dr. R. J. Ramasree is professor in Department of Computer Science, Rashtriya Sanskrit Vidya Peetha, Tirupathi, Andhra Pradesh, India. She has pursued her Ph.D in computer science in the area of Natural Language processing. She worked as systems engineer in Sri Padmavathi Womens University, Tirupathi, Andhra Pradesh, India. She acted as coordinator for innovative academic projects of natural languate processing and shabdhabodha as interdisciplinary courses sanctioned by UGC. She also coordinated machine translation project sanctioned by MHRD, India. She is acting as technical coordinator for many projects in RSVP, Tirupathi, funded by UGC. She wrote many books, course materials and several papers in the area of programming for Sanskrit students and NLP. She has presented many research papers at national, international level and published in reputed journals.

Dr. Md. Faisal, pursued Bachelor of Science in Computer Application (Hon's) from Ranchi University, Master of Computer Applications from Visvesvarayya Technologial University, Belgaum in year 2005 and Ph.D in Computer Science – Data Science from SJJT university, Rajasthan and currently working as Lecturer in Department of Informaton Technolgoy , Nizwa College of Technology, Oman since 2010. He is a member of IEEE & IEEE computer society since 2015, a reviewer of the SPRINGER journal of Big Data since 2017. His main research work focuses on Big Data Analytics, Data Mining, IoT and Software Engineering. He has 12 years of teaching experience and 8 years of Research Experience.

