Real-time Process Monitoring

Vinay KumarSomarapalli, Nageswara Rao Vesepogu

CGI Information Systems and Management Consultants Private Limited, Asia Pacific Solution Delivery Centre, Bengaluru, Karnataka, India, 560100

Abstract:

Process monitoring is the concept based onmanagingthe state of a business processesexecuted in anenterprise software. The named status and execution time are two important parameters to infer the state of a process in real-time. An execution of the process end to end is called as process instance. This paper covers monitoring process instances in real-time based on the data metric generated from event logs, where datametricsof aBusiness ProcessModels (BPM) – a weighted directed acyclic graph represents process model with its heuristics such as complete execution time of process, execution time between steps/activities and the status. Firstly, a set of process models are discovered from the events logs using unsupervised learning techniques andheuristics of different steps of a process are modeled using statisticalmethods.

Both process models and its heuristics are deployed to help process engineers to understand the path of executions of a business process, its state and performance in real-time.Path of executions involves most successful path, longest and shortest paths, error state if the process instances abruptly ends in between, state of failure etc., On the performance front the range of time taken to complete the process, for each step, most and least time consuming, identifying the steps taking more time than usual etc.,Process engineers/users are notified ifabnormalities are observed.

Keywords: Process Mining, Process Monitoring, BPM, PM4PY, Standard Deviation (SD)

1. Introduction

In an enterprise software, the business transactions are defined with a set of processes or sequence of activitieswhich normally follows a start to end. During the course of execution, every event of a process is recorded in log files and sourced to mining database to maintain detailed trails called event logs [1]. These event logs consistof traces and each trace with sequence of activities performed, execution time stamp, resource handling, time taken, cost incurred and other informative parameters. These recorded event logs are then extracted from the database and converted to format called XES [2] to explore the hidden insights through the process mining techniques i.e. discovery, conformance and extension [3]. With the help of explored insights, the end user or business analyst would be able to identify the bottlenecks causing the process to languid, deviations from the sequence of actions and other useful information and help to come up with appropriate measureslike areas required to be improved, ramp up or down the resources and improve the financial expenditure.

Process mining is most valuable approach to the business to improve the value delivered to the end customers.It helps the engineering teams to document, uncover best practices, notify if any deviations, an opportunity for improvements [4].

The process monitoring is distinguished into, a) *active monitoring* which is concerned with “real-time” propagation of relevant data concerning the enactment of business processes, such as the status or the execution time; and b) *passive monitoring* which delivers information about process instances upon request [4]. In either case, there should be some reference data, often called as data metric or simply metric to compare with the newly generated process instancefor status and execution time.

In this paper, we are more interested in active monitoring and the same will be discussed throughout. And, we define the term metric is considered to be agraph data of efficientBPM[1]and corresponding statistical data as heuristics, toexamine the statusand execution time of a process instance respectively.

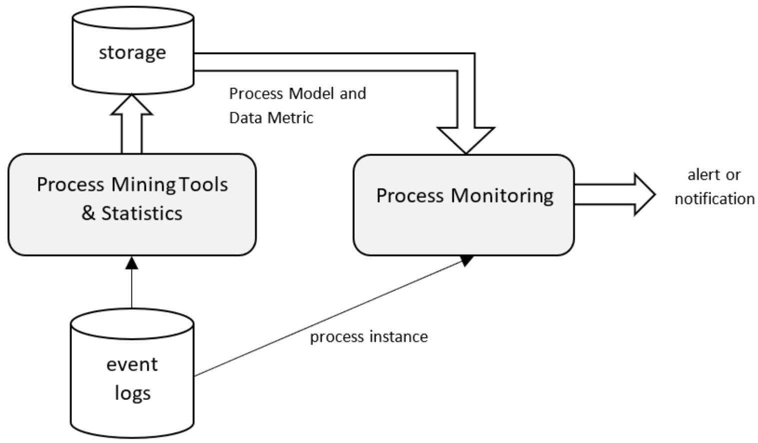

Figure 1. Overview of process monitoring implementation approach

From figure 1, the metric is generated from the historical event logs using process mining techniques and statistics. Where the metric is weighted directed acyclic graph, the directed acyclic graph represents the process modelandweights between nodes represent the execution time.

The generation of metric plays a vital role in process monitoring.The efficientprocess modelis definedin 2 ways. First, manually by the analyst with good business knowledge dealing with, or, Second, through automated process discovery [5]. In the latter case, the F1 score or F-score defined in [3, 5], which is the harmonic mean of the two measurements i.e. fitness and precision as a single metric to determine the efficientprocess model, $F1\ score = 2 * \frac{(fitness*precision)}{(fitness+precision)}$.Then, the statistical analysis performed on event logs to generate, 1) average throughput time and its standard deviation, 2) average and its standard deviation of time taken to complete the specific activity, which will be discussed in the coming sections.

The generated metric is then stored into the storage device in a convenient format and used to compare during the process monitoring against the process instance received at real-time a) for any deviation in the activity, and b) for the breach in time beyond 3 * Standard Deviation (3SD). Thereafter, the alert or notification is generated to end users for quick action on the process instance, on which abnormalities observed.

PM4PY framework [6] and supporting librariesare used predominantly for modeling.

[1] The process model defines the best sequence of activities is or to be followed to make the business process compliant and free from bottle necks, economical losses or from other forms of problems. i.e. generating the directed acyclic and weighted graph.

The rest of the paper is organized is as follows. Section-2 briefs on dataset we worked on, Section-3 specifies approach followed to generate metrics, Section-4 details on the evaluation performed with the metric, monitoring result inSection-5 and finally we conclude in Section-6.

2. Setup and dataset

The whole analysis and generation of metric required for monitoring is developed using python and PM4PYframework [6] for modelling and mining and several other compatible libraries.

As part of model building, the event logs of business processes of an enterprise applications are extracted and converted into Extended Event Stream (XES) format, which is the standard format to source events to PM4PY Library for the cross verification. The dataset published in [7] is used to explain the model parameters and verification.

3. Metric Generation

In active process monitoring, it is very important to know the status and execution time of aprocess instance, this would help to check for deviation and breach in time. By doing so, business will get ample time to act and adopt quick measures and thus save from any economic losses.

To perform process monitoring in real-time, we need a metric to compare with new process instance(s) generated by the system. To generate a data metric, we take a simple and efficient approach - 1) process discovery to come up with efficientprocess models, and 2) statistical methods to get acceptedexecution time for an activity or whole process.

Prior to the generation of process model, the event logs are preprocessed to prune loops and concurrent activities [8, 9].

A. Business Process Model (BPM)

Discovering the process from the event logs, we might end up with spaghetti-like model. But for originations to be very beneficial the lasagna-like model is preferred and said to beefficientBPM. This is achieved in the following ways,

1. **Manual:** The analyst exploits the event logs and comes up with several factors like different ways the process isperformed (called variants), their frequency, bottlenecks causing the process to slowdown leading to delay in completion and so on. Based on the derived factors, the analyst will create the BPM either a) through varying the noise threshold [6] or, b) by selecting top performing variants and performing the graph merging [6, 9], and sometimes, the analyst is provided with graph editing options – add, delete or bypass activity(ies).
2. **Automatic:**Automated discovery process operations are applied on the event logs sourced from an enterprise application and resulting the business process model as output [5]. During the operation the F1-Score or F-Score [3, 5], is used to generate BPM based on thebest score obtained,

$$F1\ score = 2 * \frac{(fitness * precision)}{(fitness + precision)}$$

To begin with, thegreedy approach is followed to capture the efficientBPM as given below,

1. Initialize the noise threshold,
2. Generate the process model based on the noise threshold from any of the discovery technique,
3. Evaluate and note the F1 score for the generated model,
4. Vary the noise threshold (varied from 0 to 1 with step size of 0.5) and repeat step 2 and 3, and
5. Finally, the BPM selected with the best F1 score.

   From either of the approach, the directed acyclic graph is generated for monitoring the status of instances.

B. Statistical Data

After generation ofBPM from the above steps, the corresponding execution time – the overall time taken to complete the process and the time tookto complete each activity is evaluated.

In general, the average is considered and estimated for future process instances. But, there maybebe a cases or instances in the past, which might have taken more than the defined time aka. anomalies, these instances will affect the calculation of average and SD leading to extra execution time. So, these exceptional cases are handled before creating the statistical data for monitoring reference.

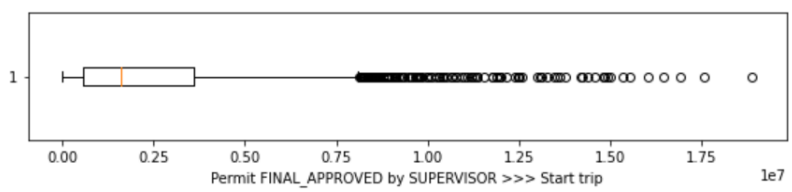

Figure 2. Box plot for execution time (seconds) from “Permit FINAL_APPROVED by SUPERVISOR” and “Start trip” activity.

From figure 2, when the execution time between 2 activities plotted on to the box plot, it is observed that certain processes with exceptional or delayed falling beyond the max whisker which are marked and anomalies. These anomalies aretreated as outliers and ignored from calculation. We select the time falling between max and min percentile, say 80 and 20 respectively [10]. The percentile is decide depending on the severity of business for which the metric is generated.

The mean (Average) and standard deviation (SD) arecalculated from the values falling between max and min percentiles and these values acts as weights for BPM generated earlier,

$$mean_{time} = \frac{1}{n}\sum time_i \ in \ seconds)$$

$$sd_{time} = \sqrt{\frac{\sum(time_i - mean_{time})^2}{n}}$$

4. Monitoring

In the previous section, we have discussed about the generation of data metric. And now, we willdiscuss on the approach applying metric to monitor the process instances generated in real-time.

From figure 1, the event logsare stored and sourced from event database. Start and End event and its detailed trial are sourced to modeling module to build process models. The process models and along with its heuristics is deployed to monitor therunning instance of a process in real-time. The process monitoring module will register and alert and trigger a notification to service engineers if any abnormalities is observed against the data metric generated. This would help the business to revert with action to quickly adopt special measures that can mitigate from eventual consequences.

a. Status

It is good know the status of every instance tocheck for deviation is observed and correct it immediately or analyze the reason. During the evaluation process, the following steps are performed based on graph traversal with breadth first concept,

1. Read the process instance data and data metric,
2. Start with first activity,
3. Get the list of next activity(ies) of the current activity from the BPM defined in data metric,
4. Get the next activity from the process instance,
5. Compare the activities extracted from step 3 and 4,
6. Compare,If the next activity from step 4 available in the activities in step 3?
    a. Yes,
        i. go to step 3
    b. No,
        i. notify the end or business user

b. Execution time

Similar to monitoring for deviation in status, every activity will have to be completed in the defined, so it also important to monitor *overall* execution time and time of *previous activities* took to attain completion status.

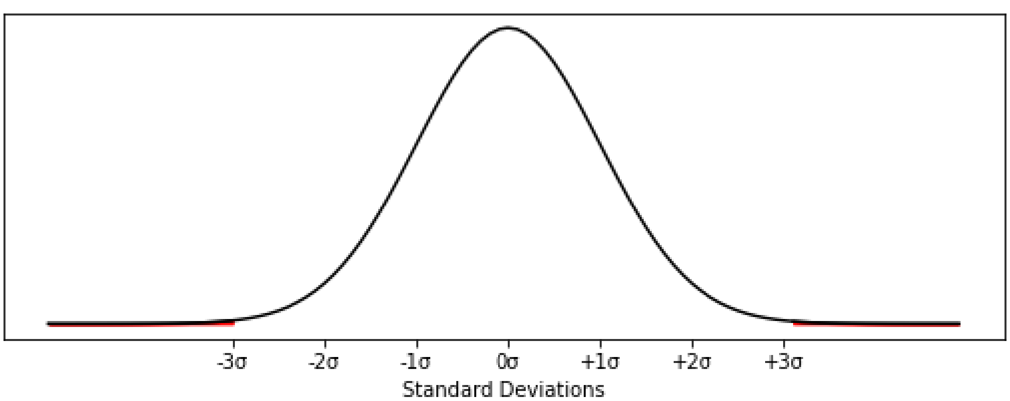

Figure 3. Plot of standard deviations

In our paper, we define ±3σ (3SD)as threshold for testing. The realistic threshold is defined based on the type of business. Suppose, if the financial institutions need a strict compliance the threshold can be reduced to ±2σ or ±1σ. If the activity execution time falls beyond the threshold then the end users or business are notified.

5. Results

To simulate the process monitoring, we have created the synthetic data with data in [7] as referenceand processed it as real-time instance against the data metric generated from the steps mentioned in section 3. For better understanding of process monitoring, the status and execution time of BPM graph is better visualized as,

- ***EfficientBPM***: black color with dotted lines,
- ***Compliant processinstance***: Green with solid lines, and
- ***Non-compliant process***: Red color with solid lines

From the figure 4 and 5, the execution abnormalities observed – deviationin execution sequence of activities and both execution timeand delay in completion of activity.

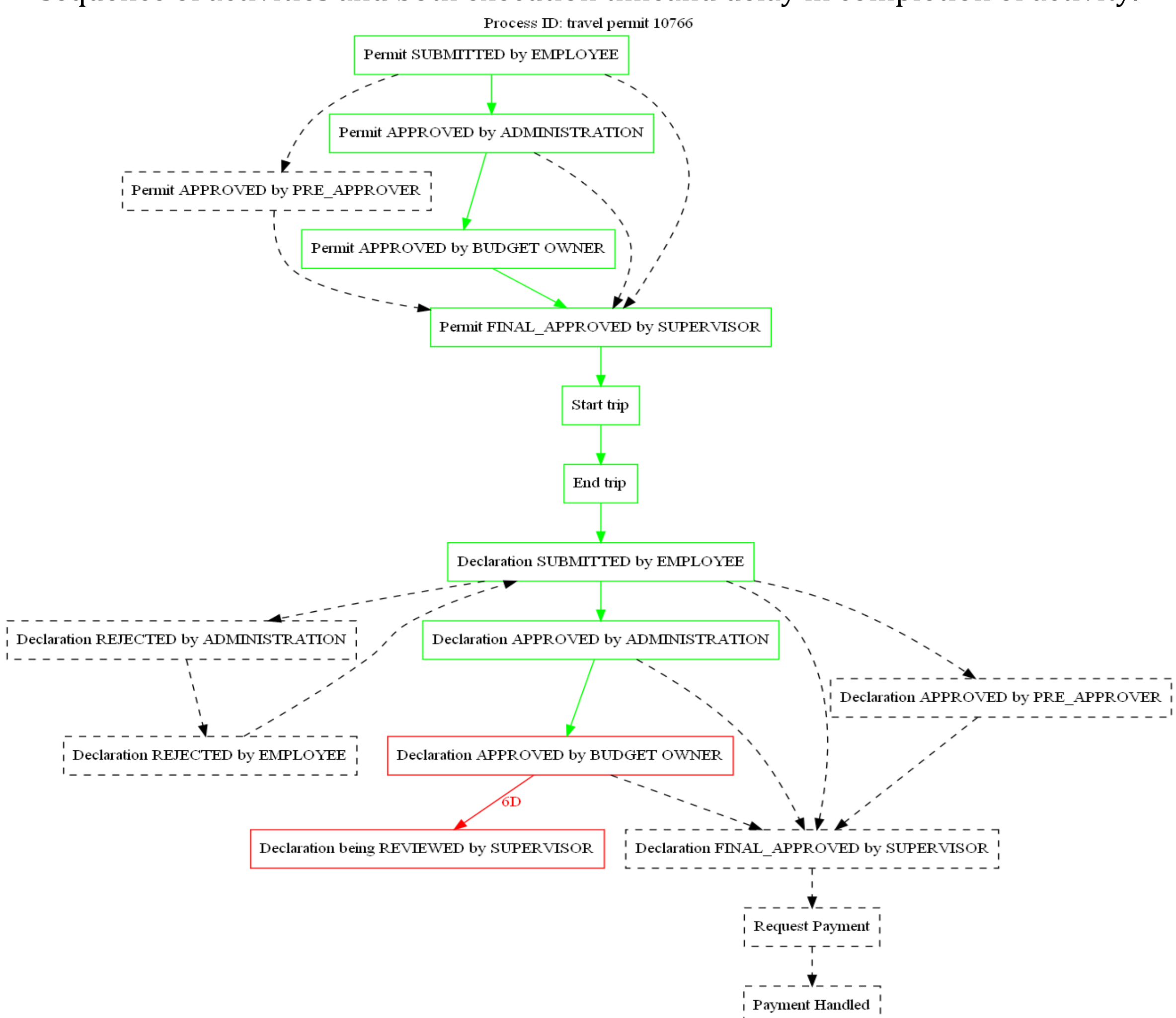

Figure 4. Deviation observed the sequence of activity performed

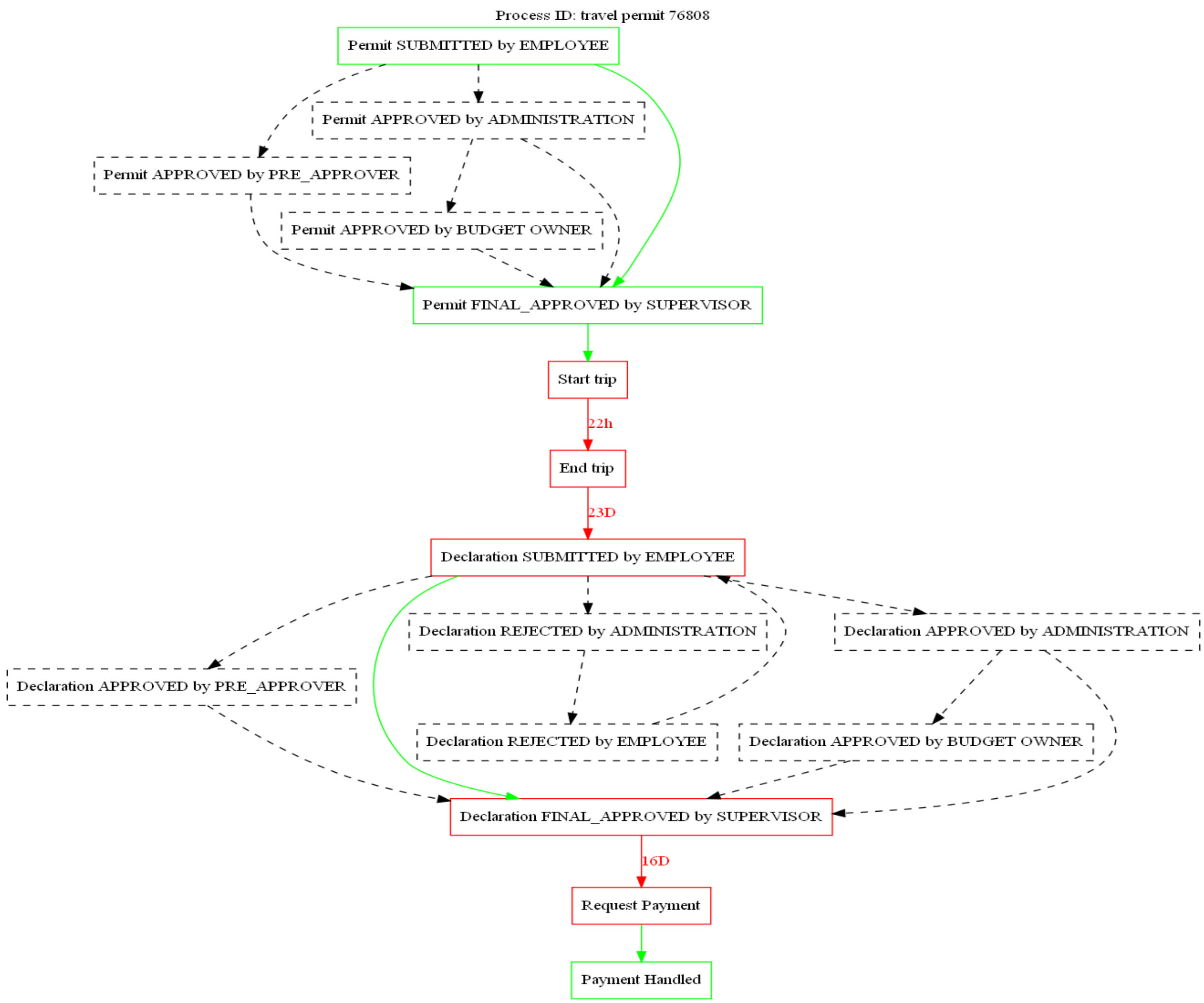

Figure 5. Observed deviation and defiance in time

6. Conclusion

The statistical approach for process monitoring is simple and efficient, helps the business and organizations to define and monitor the process on real-time basis. Defining BPM is crucial, but simplified with automated BPM discovery approach. The severity level of monitoring i.e. from low to critical for the business can be decided based on the percentile during data metric generation and level of standard deviation during monitoring the process instances. The notification from the abnormalities, will help them to revert into action immediately and implement quick measures to mitigate the eventual consequences. The proposed approach is easy to deploy and fast in monitoring even when many process instances generated by the application.

References

1. W. van der Aalst, T. Weijters and L. Maruster, "Workflow mining: discovering process models from event logs," in IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 9, pp. 1128-1142, Sept. 2004.

2. W.M.P. van der Aalst, B.F. van Dongen, J. Herbst, L. Maruster, G. Schimm, and A.J.M.M.Weijters. Workflow Mining: A Survey of Issues and Approaches. Dataand Knowledge Engineering, 47(2):237–267, 2003.
3. W. van der Aalst, Process Mining - Data Science in Action. Springer,2016
4. de Medeiros A.K.A. et al. (2007) An Outlook on Semantic Business Process Mining and Monitoring. In: Meersman R., Tari Z., Herrero P. (eds) On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops. OTM 2007. Lecture Notes in Computer Science, vol 4806. Springer, Berlin, Heidelberg.
5. A. Augusto et al., "Automated Discovery of Process Models from Event Logs: Review and Benchmark," in IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 4, pp. 686-705, 1 April 2019.
6. PM4PY, https://pm4py.fit.fraunhofer.de/
7. BPI Challenge 2020, https://data.4tu.nl/collections/_/5065541/1
8. Martin N., Martinez-Millana A., Valdivieso B., Fernández-Llatas C. (2019) Interactive Data Cleaning for Process Mining: A Case Study of an Outpatient Clinic's Appointment System. In: Di Francescomarino C., Dijkman R., Zdun U. (eds) Business Process Management Workshops. BPM 2019. Lecture Notes in Business Information Processing, vol 362. Springer, Cham.
9. A. Augusto, R. Conforti, M. Dumas and M. La Rosa, "Split Miner: Discovering Accurate and Simple Business Process Models from Event Logs," 2017 IEEE International Conference on Data Mining (ICDM), 2017
10. Devore, Jay L. 2016. Probability and statistics for engineering and the sciences.
11. Dumas, M., La Rosa, M., Mendling, J., &Reijers, H. A. (2018). Fundamentals of business process management (2nd ed.). Rajendiran, P., et al. "Customer Relationship Management in the Manufacturing Industry, Using Data Mining Techniques." *International Journal of Educational Science and Research (IJESR)* 7.6 (2017): 63-70.
12. Rajendiran, P., et al. "Customer Relationship Management in the Manufacturing Industry, Using Data Mining Techniques." *International Journal of Educational Science and Research (IJESR)* 7.6 (2017): 63-70.
13. Deepk, D., and Beedu Rajendra. "Investigation of the effect of process parameters on material removal rate in machining of is 5986 fe 410 steel by wire-edm: the Optimization and regression modeling." *International Journal of Mechanical and Production Engineering Research and Development* 9.6 (2019): 411-420.
14. Mohan N. A. N. J. A. N. G. U. D., Sathyashankara Sharma, and Ritesh Bhat. "A comprehensive study of glass fibre reinforced Polymer (gfrp) drilling." *International Journal of Mechanical and Production Engineering Research and Development* 9.1 (2019): 1-10.
15. Nyssabayeva, G. R., et al. "Synthesis of modified nanocarbon materials and determination of their adsorption capacity." *International Journal of Mechanical and Production Engineering Research and Development* 10.1 (2020): 305-314.
16. Onoduo, Vincent A., I. K. P. E. IKPE MARIUS, and Oluchukwu F. Anowor. "Non-oil export and

economic growth in Nigeria: A time series econometric model." *International Journal of Business Management & Research (IJBMR)* 3.2 (2013).

17. Rath, B., Biswajit Das, and J. Panigrahi. "Sustainability impact assessment of developmental programmes undertaken by mining industries as part of the corporate social responsibility for value creation." *International Journal of Mechanical and Producition Engineering Research and Development* 8.3 (2018): 23-30.