IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 7, July 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907035 |

Asymmetric Allocation for Breakdown Prediction Using Ensemble Learning With Oversampling Algorithm

Aarthi R, Arthi S, Celshia Helan A, Mrs. K. Suja Rajeswari

Student, Department of CSE, Panimalar Institute of Technology, Chennai, India

Student, Department of CSE, Panimalar Institute of Technology, Chennai, India

Student, Department of CSE, Panimalar Institute of Technology, Chennai, India

Assistant Professor, Department of CSE, Panimalar Institute of Technology, Chennai, India

ABSTRACT: An imbalanced classification is an example of a classification problem where the distribution of examples across the known classes is biased or skewed. The distribution can vary from a slight bias to a severe imbalance where there is one example in the minority class for hundreds, thousands, or millions of examples in the majority class or classes. Imbalanced classifications pose a challenge for predictive modelling as most of the machine learning algorithms used for classification were designed around the assumption of an equal number of examples for each class. This results in models that have poor predictive performance, specifically for the minority class. This is a problem because typically, the minority class is more important and therefore the problem is more sensitive to classification errors for the minority class than the majority class. We proposed a model which handles the imbalanced data and to predict the fault occurrences and their solution to rectify the fault as well as updating the new pipeline failure data in the model. In the industry fault rectification requires much amount of time and effort, finding the error is also a tedious process. This will reduce the cost and time efficiency in the oil and gas production industries.

KEYWORDS: imbalanced classification, oversampling, breakdown prediction

I.INTRODUCTION

Imbalanced classification refers to a classification predictive modelling problem where the number of examples in the training dataset for each class label is not balanced. Machine learning in oil & gas can be used to enhance the capabilities of this increasingly competitive sector. Not only can it help to streamline the workforce. The technology can also be used to optimise extraction and deliver accurate models. The future scope includes analysing the sensor data in the oil and gas industry to increase monitoring each and every part of the machineries and machine learning approaches used in the predictive maintenance to enhance the life of the pipelines and machineries. The implementation of this concept requires additional model training to solve complex problems.Imbalanced data sets in real-world applications have a majority class with normal instances and a minority class with abnormal or important instances. The Synthetic minority over-sampling technique is specifically designed for learning from imbalanced data sets which gives better machine.

II.EXISTING SYSTEM

In existing system, if any failure occurs nay experienced professional will be incharge to resolve the issue. If any very rare issue arises in the pipeline, finding the cause and deriving the steps and man power is a time consuming and tedious process. Usage of machine learning for predicting the cause and failure improves the efficient problemsolving method. As well as while training the model, there arises a problem i.e. imbalanced dataset. Imbalanced classification is the problem of classification when there is an unequal distribution of classes in the training dataset. The imbalance in the class distribution may vary, but a severe imbalance is more challenging to model and may require specialized techniques. It is possible that the imbalance in the examples across the classes was caused by the way the examples were collected or sampled from the problem domain. This might involve biases introduced during data collection, and errors made during data collection.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907035 |

III.PROBLEMS IN EXISTING SYSTEM

In existing system, if any failure occurs nay experienced professional will be incharge to resolve the issue. If any very rare issue arises in the pipeline, finding the cause and deriving the steps and man power is a time consuming and tedious process. Usage of machine learning for predicting the cause and failure improves the efficient problem solving method. As well as while training the model, there arises a problem i.e. imbalanced dataset. Imbalanced classification is the problem of classification when there is an unequal distribution of classes in the training dataset. The imbalance in the class distribution may vary, but a severe imbalance is more challenging to model and may require specialized techniques. It is possible that the imbalance in the examples across the classes was caused by the way the examples were collected or sampled from the problem domain. This might involve biases introduced during data collection, and errors made during data collection.

IV.PROPOSED SYSTEM

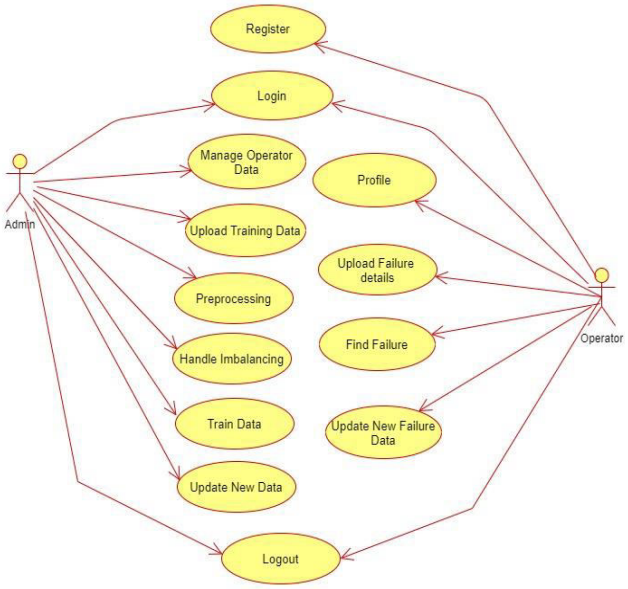
One of the most noticeable impacts of machine learning in oil & gas focused industries is how it transforms discovery processes. Applications employing machine learning in oil & gas enable computers to quickly and accurately analyze huge amounts of data. After this information has been gathered and analyzed modern software applications can construct accurate geological models. By using machine learning in oil & gas in this manner has allowed engineers to auto-track the working of all the process. By using machine learning in oil & gas in this manner has allowed engineers to identify the type of damage occurs and rectify them more efficiently. In particular, machine learning algorithms can be used for case-based reasoning (CBR). This means that the algorithms can be used to quickly sort through massive databases of recorded problems. The algorithms are then able to identify similar cases. Once a similar case, or cases, is identified the software identifies how the issue has previously been solved. So that at the time of experienced operators not available or in a hard situations this algorithm helps other professionals to resolve the issue with a note and solving videos.

V.ADVANTAGE OF PROPOSED SYSTEM

- ✓ This allows operatives to predict, accurately, what is actual reason for the failure.
- ✓ This can speed up problem resolution times, saving money and improving efficiency.
- ✓ Data training using the imbalanced data will be solved using oversampling approach.
- ✓ The production efficiency will be improved with less amount time taken to resolve.
- ✓ New type of failures will be trained with professional knowledge.
- ✓ Past similar failure data will be used to predict and rectify the problem.
- ✓ The cause and rectification steps of the project will be provided to the operator at the resolving time.

VI.ARCHITECTURE & UML DIAGRAMS

A. USE CASE DIAGRAM

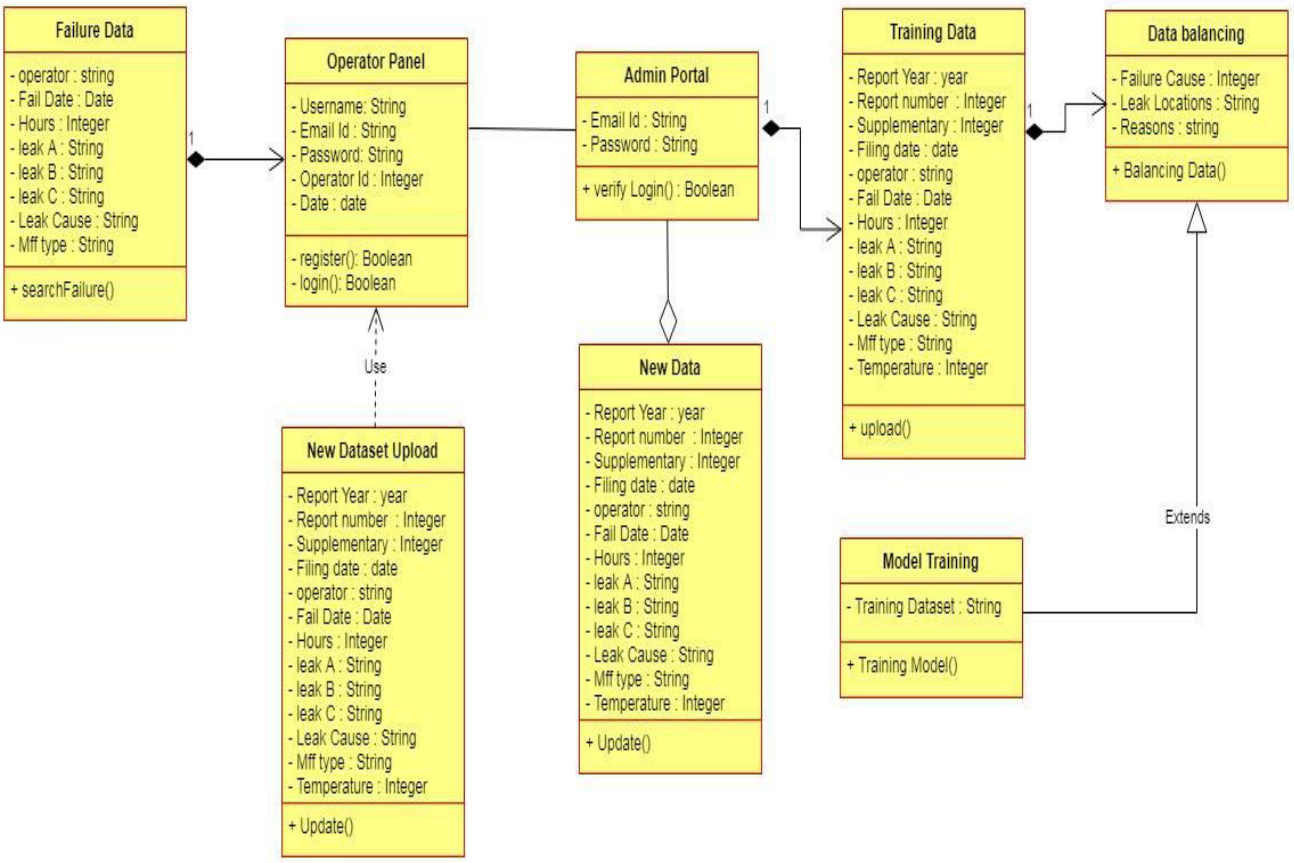


| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

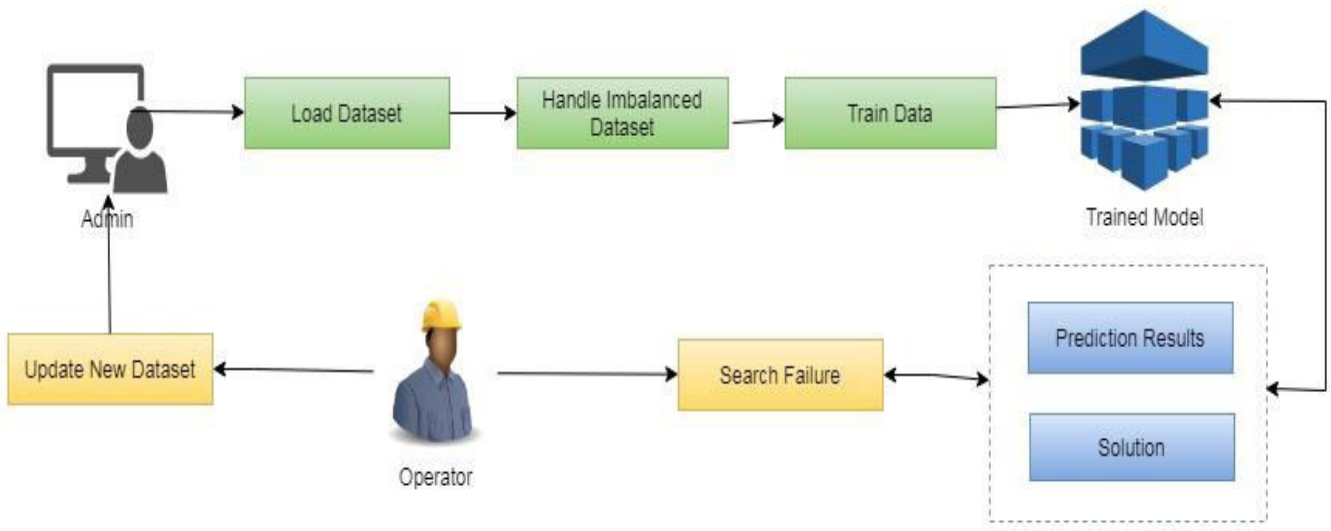
|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907035 |

B. CLASS DIAGRAM



C. SYSTEM ARCHITECTED

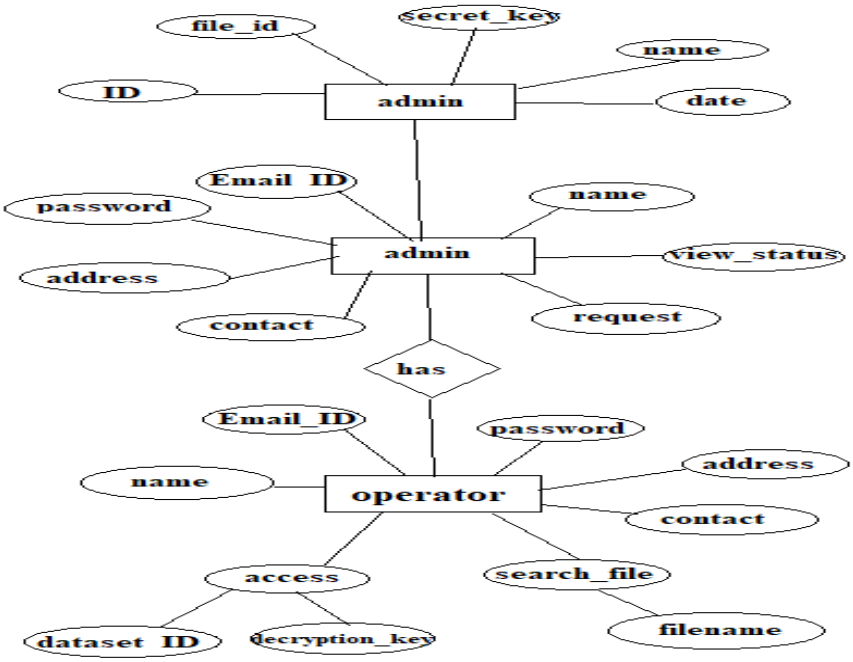


| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

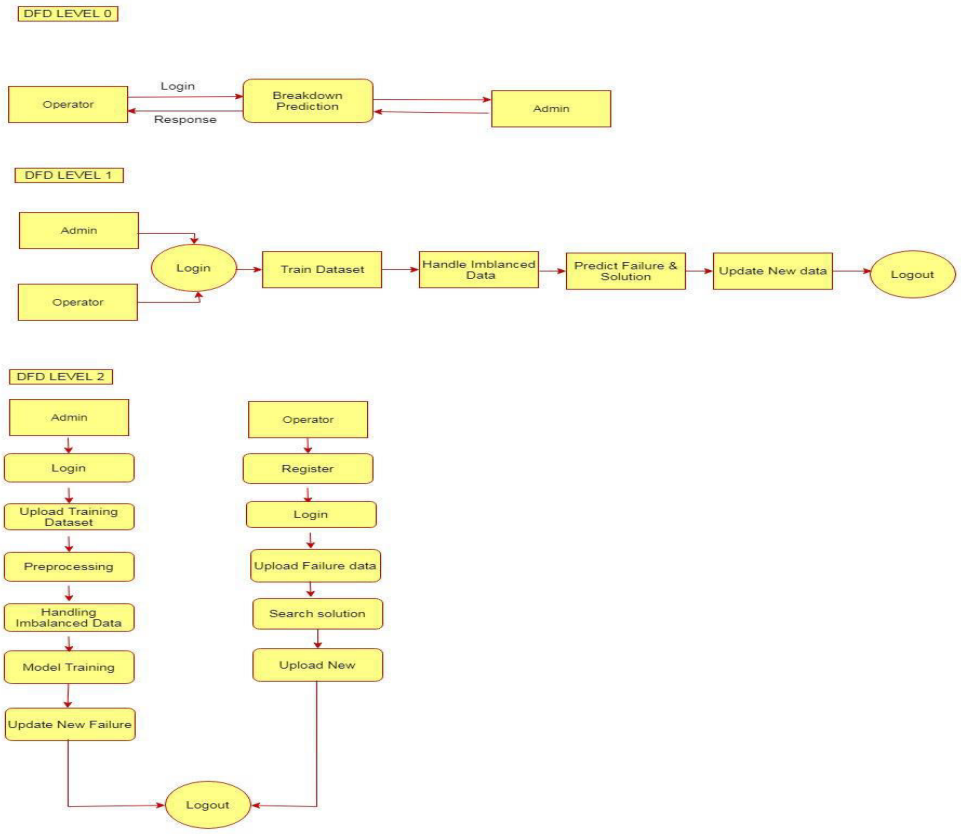
|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907035 |

D.ER DIAGRAM



E. DATA FLOW DIAGRAM

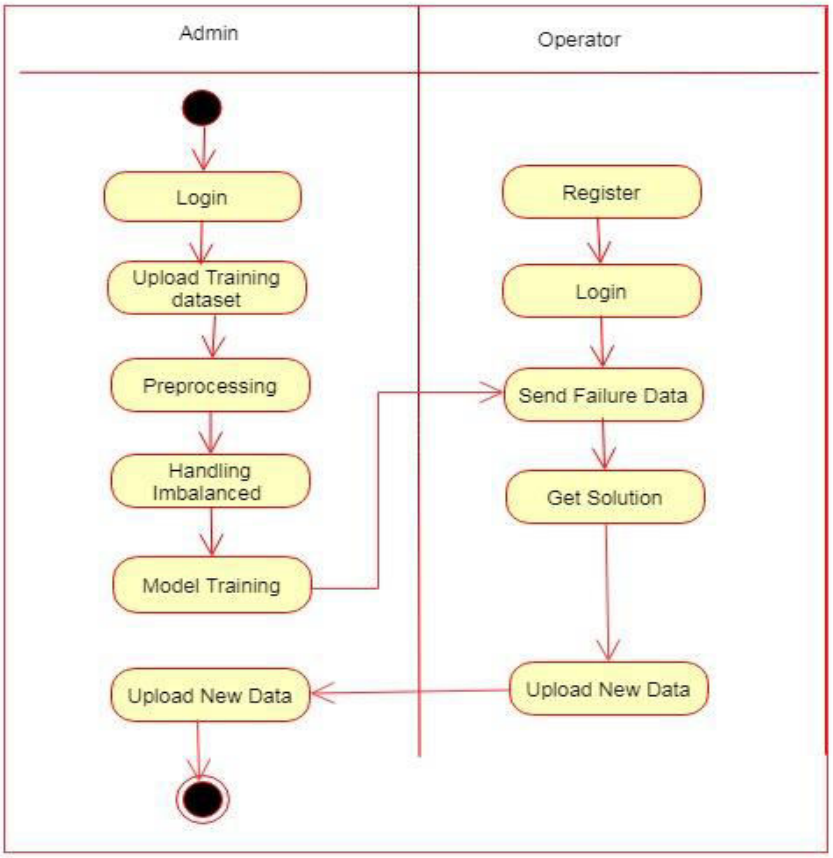


| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907035 |

F. ACTIVITY DIAGRAM



VII.CONCLUSION

Machine learning in oil & gas will not replace manual operatives entirely. While it will account for some streamlining, human operatives will still be required. Using machine learning in oil & gas industries will allow skilled workers to become more efficient. It can also save them from conducting needless tasks. Finally adopting more applications centred on machine learning in oil & gas can make the industry safer. However, it is necessary and workers will soon become adept at these skills. Workers who can properly deliver all these skills will become commonplace. They will also be best served by machine learning algorithms that are fed by standardised, quality data. This will yield the best possible results. Tasks such as collecting and maintaining data will fall increasingly on AI and machine learning, in oil & gas as well as other industries. While this will allow for a standardized information base to be created it won't completely negate the need for human workers. Machine learning in oil & gas will not only improve the customer experience but can also help to keep costs low across the process. The possible advantages that can be brought by machine learning in oil & gas to this competitive sector are massive.

REFERENCES

1. "An improved ensemble learning for imbalanced data classification" by Zhengwu Yuan, Pu Zaho.
2. "Auto-ensemble: A adaptive learning rate scheduling based deep learning model ensembling" by JUN Yang , FEI WANG
3. "Demonstration if imbalanced data learning based failure prediction in self-optimizing optical networks with large scale field topology" by Boyuan Yan, Yongli Zhao, YajieLi ,Xiaosong Yu, Jie Zhang ,Yilin, Haomian Zheng
4. "Oversampling algorithm based on spatial distribution of data sets for imbalance learning" by Yiran Liu, Wanjiang Han, Xiaoxiangwang, Qi li
5. "An improving majority weighted minority oversampling technique for imbalanced classification problem" by Chao-Ran wang, Xin-Hui Shao

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details