# Deep Monument Asymentric Allocation for Breakdown Prediction

**Ameer Basha .H[1] Dhivakar .R[2] Rohan Robello Paul .A[3] Dr. K. Pandikumar[4]**
[1,2,3]Research Students [4]Assistant Professor
[1,2,3,4]Department of Computer Science and Engineering
[1,2,3,4]Dhanalakshmi College of Engineering Chennai, Tamil Nadu, India

*Abstract*— An imbalanced classification problem 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:*** Over Sampling Algorithm, Breakdown Prediction, imbalanced classification

## I. INTRODUCTION

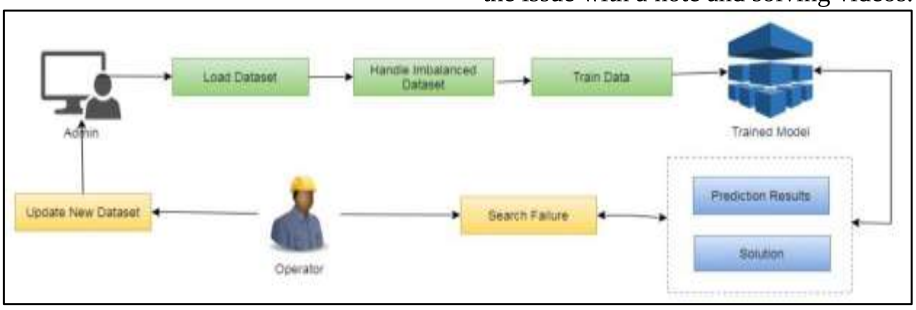
Imbalanced classification refers to a classification predictive modeling problem where the number of examples in the training dataset for each class label is not balanced. Machine learning in oil & gas can be used to enhance the capabilities of this increasingly competitive sector. Not only can it help 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. DOMAIN: MACHINE LEARNING

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop a conventional algorithm for effectively performing the task.Machine learning is closely related to computational statistics, which focuses on making predictions using computers.The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine learning, and focuses on exploratory data analysis through learning. In its application across business problems, machine learning is also referred to as predictive analytics.

## III. PROPOSED METHODOLOGY

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 analyse huge amounts of data. After this information has been gathered and analysed 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.



## IV. Scope of the Project

Imbalanced classification refers to a classification predictive modeling 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

### A. Advantages

- 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.
- 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.

### B. Disadvantages

- Finding the failures is done by using human knowledge.
- Finding cause and rectification is quite hard.
- Even experienced professionals were struggling to find the cause of failure.
- Resolving issue will consume much time and cost.
- Production will be paused at the time of rectification.

### C. Block diagram

Modules

- Load Dataset
- Oversampling
- Model Training
- Failure Prediction
- Training new data

### D. Load Dataset

In oil and gas industry there are various types of pipes which is connecting to different places inside the industry to transfer the extracted oil and gas to different locations inside the industry. There are faults occurring on the pipelines when they face certain degradation. In this module we load the failure dataset to our model to provide training on the particular dataset. Loading up of the dataset requires some time due to large amount, after this the training algorithm will work.

### E. Oversampling

The loaded datasets were preprocessed and extracted for the number of different types of failures. Then the total numbers of occurrences of the failures were calculated and grouped by the types. There are lots of imbalanced calculations in the dataset, So that the oversampling algorithm is applied here. It will balance the imbalanced dataset type to a proper count after that it will give for training the model.

### F. Model Training

In the module our learning algorithm will get trained on the training dataset. The loaded dataset will be preprocessed initially to ensure the integrity of the data. Then the oversampling is applied. Finally the dataset will be trained to predict the failure.

### G. Failure Prediction

In this module the algorithm will predict the type of failure which is occurred in the pipelines of the oil and gas industry. At the time of failure occurrences the operator of the particular section will search for the solution to resolve the issue. The operator can interact with the model to fetch the error. And the solution will be provided.

### H. Training new data

In this module, if at any rare case the occurred failure is a new one, then the new type of failure with their attributes and solutions will be analyzed, added and trained to the model. Through this module the training model will adapt new type of scenarios also.

## V. Bottomline and Future Enhancement

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 standardised 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.

### A. *Hardware and Software Requirements*

1) *Hardware Requirements:*

Processor : Intel (R) Pentium (R)
Speed : 1.6 GHz and Above.
RAM : 4 GB and Above.
Hard Disk : 120 GB.
Monitor : 15'' LED SVGA
Input Devices : Keyboard, Mouse.

2) *Software Requirements:*

Operating system : Windows 7 / 8 / 8.1 / 10.
Coding Language : JAVA / J2EE.
Java Version : jdk 8.
IDE : Eclipse Oxygen.
Database : MySQL v5.1.
Database Tool : HeidiSql v11.0.
Application Server : Apache Tomcat 8.X / 9.X.

## VI. OUTOUT SCREENSHOT











## 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 standardised 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] F. Bao, Y. Deng, Y. Zhao, J. Suo, and Q. Dai, "Bosco: Boosting corrections for genome-wide association studies with imbalanced samples," IEEE Trans. Nanobiosci., vol. 16, no. 1, pp. 69–77, Jan. 2017.
[2] F. Bao et al., "Probabilistic natural mapping of gene-level tests for genome-wide association studies," Briefifings Bioinf., vol. 19, no. 4,pp. 545–553, 2017.
[3] Y. Deng et al., "Information transduction capacity reduces the uncertainties in annotation-free isoform discovery and quantifification," Nucleic Acids Res., vol. 45, no. 15, p. e143, 2017.
[4] H. He and E. A. Garcia, "Learning from imbalanced data," IEEE Trans. Knowl. Data Eng., vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
[5] Z. Akata, F. Perronnin, Z. Harchaoui, and C. Schmid, "Good practice in large-scale learning for image classifification," IEEE Trans. Pattern Anal.Mach. Intell., vol. 36, no. 3, pp. 507–520, Mar. 2014.
[6] S. Wang, L. L. Minku, D. Ghezzi, D. Caltabiano, P. Tino, and X. Yao, "Concept drift detection for online class imbalance learning," in Proc.Int. Joint Conf. Neural Netw., Aug. 2013, pp. 1–10.
[7] X.-Y. Liu, J. Wu, and Z.-H. Zhou, "Exploratory undersampling for class-imbalance learning," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 39, no. 2, pp. 539–550, Apr. 2009.
[8] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell.Res., vol. 16, no. 1, pp. 321–357, 2002.
[9] Y. Sun, M. S. Kamel, A. K. C. Wong, and Y. Wang, "Cost-sensitive boosting for classifification of imbalanced data," Pattern Recognit.,vol. 40, no. 12, pp. 3358–3378, 2007.
[10] B. Zadrozny, J. Langford, and N. Abe, "Cost-sensitive learning by cost-proportionate example weighting," in Proc. IEEE Int. Conf. Data Mining, 2003, pp. 435–442.
[11] Y. Deng, F. Bao, Y. Kong, Z. Ren, and Q. Dai, "Deep direct reinforcement learning for fifinancial signal representation and trading," IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 3, pp. 653–664, Mar. 2017.
[12] Y. Deng, Z. Ren, Y. Kong, F. Bao, and Q. Dai, "A hierarchical fused fuzzy deep neural network for data classification," IEEE Trans. Fuzzy Syst., vol. 25, no. 4, pp. 1006–1012, Aug. 2017.
[13] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," in Proc. Int. Conf. Mach. Learn., 2016, pp. 478–487.
[14] M. T. Law, R. Urtasun, and R. S. Zemel, "Deep spectral clustering learning," in Proc. Int. Conf. Mach. Learn., 2017, pp. 1985–1994.
[15] K. Wang, Z. Ming, and T.-S. Chua, "A syntactic tree matching approach to finding similar questions in community-based qa services," in Proc. ACM 32nd Int. SIGIR Conf. Res. Develop. Inf. Retr., 2009, pp. 187–194.
[16] R. Wald, T. M. Khoshgoftaar, A. Fazelpour, and D. J. Dittman, "Hidden dependencies between class imbalance and diffificulty of learning for bioinformatics datasets," in Proc. IEEE 14th Int. Conf. Inf. Reuse Integr. (IRI), Aug. 2013, pp. 232–238.
[17] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-,boosting-, and hybrid-based approaches," IEEE Trans. Syst., Man, C,Appl. Rev., vol. 42, no. 4, pp. 463–484, Jul. 2012.
[18] R. A. Sowah, M. A. Agebure, G. A. Mills, K. M. Koumadi, and S. Y. Fiawoo, "New cluster undersampling technique for class imbalance learning," Int. J. Mach. Learn. Comput., vol. 6, no. 3, p. 205, 2016.
[19] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-SMOTE: A new oversampling method in imbalanced data sets learning," in Proc. Int. Conf. Intell. Comput., 2005, pp. 878–887.

[20] H. He, Y. Bai, E. A. Garcia, and S. Li, “ADASYN: Adaptive synthetic sampling approach for imbalanced learning,” in Proc. IEEE Int. Joint Conf. Neural Netw., Jun. 2008, pp. 1322–1328.