NLP-Based Fault Detection Method for Multifunction Logging-While-Drilling Services

Corina Maria Panait[1], Qian Su[2], Nahieli Vasquez[3], Ahmed Mosallam[3], Fares Ben Youssef[2], Olexiy Kyrgyzov[3], Hassan Mansoor[4], Anup Arun Yadav[5]

[1] *SLB, Str. Sergeant Constantin Ghercu, Bucharest, Romania*
CPanait5@slb.com
[2] *SLB, 135 Rousseau Road, 70592 Youngsville, Louisiana, USA*
QSu3@slb.com
FYoussef@slb.com
[3] *SLB, 1 rue Henri Becquerel, 92140 Clamart, France*
NVasquez23@slb.com
AMosallam@slb.com
OKyrgyzov@slb.com
[4] *SLB, Nowogrodzka Street 68, 02-014 Warsaw, Poland*
HMansoor2@slb.com
[5] *SLB, 8 Office 301 Commerzone IT park, Pune, India*
AYadav28@slb.com

Abstract

This paper presents a Natural Language Processing (NLP) method aimed at detecting faults within field failure reports of drilling tools. It builds on the definition of entities specifically matched to our unique requirements. These entities have been annotated within the dataset under the guidance of a Subject Matter Expert (SME), laying a foundation for our NLP method. By utilizing a model based on bidirectional encoder representations from transformers, the method achieves an F1-score of 88% in identifying entities and consequently detecting faults within field failure reports. This work is part of a long-term project aiming to construct a failure analysis and resolution system for drilling tools.

1. Introduction

The oil and gas industry relies heavily on logging tools that operate in extreme environmental conditions, including elevated temperatures, vibrations, and pressures. Such conditions can accelerate the degradation of tools, leading to potential failures. These failures not only compromise operations by providing inaccurate information but also result in delayed deliverables, tool repair, or even cancellation of the entire operation. Such setbacks translate into nonproductive time and substantial financial losses. Efficiency and speed in the maintenance process are critical when a logging tool fails. The maintenance team tackles the task of navigating extensive unstructured data to identify patterns of failures. Streamlining this maintenance workflow is essential to expedite the turnaround time of the tool, achieving its swift return to operational readiness and minimizing nonproductive intervals and associated financial losses. Manual analysis of these data proves extremely time-consuming in the context of an industry where avoiding downtime is a priority.

While Prognostics and Health Management (PHM) methods have been successfully utilized to enable predictive maintenance, this approach has its limitations (Mosallam, Laval, Youssef, Fulton, & Viassolo, 2018; Mosallam, Kang, Youssef, Laval, & Fulton, 2023; Mosallam, Youssef, et al., 2023; Kang et al., 2022). It primarily relies on equipment sensor data and does not take into account the rich source of information in maintenance logs, such as failure reports, asset performance, maintenance policies, and failure patterns collected during the life cycle of asset management. Through computerized analysis, i.e., NLP, it is possible to make the process considerably more efficient (Stenström, Al-Jumaili, & Parida, 2015). For instance, (Juan Pablo Usuga Cadavid & Fortin, 2020) utilized an NLP model named CamemBERT (Bidirectional Encoder Representations from Transformers) to predict the criticality and duration of maintenance issues based on free-form text comments from operators (Martin et al., 2019). Despite the unstructured and imbalanced nature of maintenance logs, the authors suggest that their approach can provide significant advantages in coupling production scheduling with maintenance logs, enabling the adaptation of planning to the shop floor. Despite the recent and rapid development of the NLP field, extracting meaningful insights from maintenance logs remains a challenging task. (Brundage, Sexton, Hodkiewicz, Dima, & Lukens, 2021) argue that current NLP tools are not suitable for engineering data and propose a domain-driven approach called Technical Language Processing (TLP). They suggest that key NLP tools need to be adapted to the maintenance domain based on available maintenance text-based data. For example, (Naqvi, Ghufran, et al., 2022) introduced a TLP approach utilizing a Case-Based Reasoning (CBR) framework paired with a domain-adapted BERT model to address maintenance issues through textual data from mining operations. This approach, particularly the use of a Transformer-based Sequential Denoising Autoencoder (TSDAE) for unsupervised fine-tuning and cosine similarity for case assessment, underscored the importance of domain-specific model training (Wang, Reimers, & Gurevych, 2021). (Lee & Marlot, 2023) proposed Oil & Gas domain-relevant entity and relationship extraction from drilling reports. The approach involves training a Named Entity Recognition (NER) model to identify key information or failure symptoms in the reports, such as equipment, operations, or events, which can be considered a fault detection task from maintenance logs. This is followed by a Relation Extraction model that identifies the relationships between the entities and recommends early mitigation using historical data. The authors also applied data augmentation techniques to increase the data samples and improve the model's performance in detecting rare entities. Finally, (Naqvi, Varnier, Nicod, Zerhouni, & Ghufran, 2022) propose an NLP method for fault diagnostics from maintenance logs. The study finds that fine-tuning CamemBERT outperforms classical NLP approaches and that data augmentation using deep contextualized embedding further improves performance.

We present an NLP-based fault detection approach leveraging a NER model based on the BERT framework for Logging-While-Drilling Service (LWD) (Hansen & White, 1991) (Figure 1). This approach addresses equipment failures within unstructured data, offering a robust solution for the oil and gas industry. The BERT model, chosen for its advanced contextual understanding of words in text, excels at identifying and classifying key entities related to equipment failure and operational procedures within textual data. Custom and actionable entities were defined through extensive data labeling to enhance the model's proficiency in recognizing relevant information.

The rest of this paper is structured into four sections. Section 2 presents a description of the failure investigation process. The method and the results are presented in section 3 and 4, respectively. Finally, section 5 concludes the paper.

Figure 1. Multifunction LWD Service.

2. Failure Investigation Process

Field incidents and maintenance data are utilized to construct the dataset required for this work. These data are stored across various business systems in diverse formats. Field incidents are reported in the Failure Investigation Business System known (FIBS). Field Crew generates a field failure report which includes the following details:

1. Basic event information, including event date, location and the suspected failed technology.
2. The primary content is the "Field Failure Description," a free-text input where the Field Crew documents all their observations concerning the event sequence and the failure symptoms of suspected technology. This part is the cornerstone of our analysis, as identifying historical events with similar failure symptoms is crucial.
3. The "Remedial Actions Attempted," also a free-text input, where the Field Crew records all the actions that they attempted to restore the tool to normal operation.

Maintenance Data is primarily stored in a maintenance business system. A field failure Work Order (WO) is generated for each technology implicated in field incidents and flagged by the field crew when creating the FIBS report. The main section of each Work Order for technicians involved in failure investigations is the failure description section, where they provide free-text input on analysis, testing, and findings. This segment serves as a valuable source of insights and knowledge, helping others who may encounter similar failures with the same technology. After completing the investigation and if any failed components are identified, this failed part is then recorded in a dedicated business system. The failed components are confirmed immediate causes in all historic incidents, which can be utilized to establish a failure Pareto chart, illustrating the probability of specific causes for distinct failure symptoms. The failure investigation process is summarized in Figure 2.

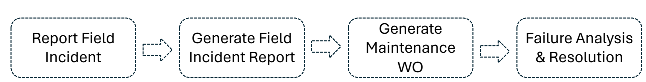


Figure 2. Failure Investigation Process.

A critical part of this process is analyzing failure reports of LWD service to determine whether the downhole conditions have had any detrimental effects on the operability. Due to the complexity of this service, analysis of this vast amount of data is very time-consuming and prone to error if performed manually. Therefore, fault detection from maintenance logs is of utmost importance for operation. An automated tool, which can determine different failure symptoms from maintenance logs with minimal user input, removes variability, eliminates human error and provides an efficient decision on the required maintenance in a fraction of the time. The reliability benefits are clear and provide significant cost savings both for the client in terms of reduced Non-Productive Time at the rigsite and for the Original Equipment Manufacturer in terms of reduced Materials and Supplies (M&S) during maintenance and troubleshooting.

3. Proposed Method

Currently, this work focuses solely on one technology, a multi-function Logging-While-Drilling tool designed for oil and gas well drilling applications. This technology integrates a comprehensive suite of formation-evaluation measurements (resistivity, porosity, density, natural gamma-ray, etc.) and drilling parameters (temperature, pressure, shock, vibration, etc.) into a single housing. Given the domain-specific nature of the data, it was necessary to create bespoke entities tailored to our unique requirements. This crucial task was carried out by an SME. The SME annotated key entities within a vetted dataset, laying the groundwork for a robust NLP model. This model is adept at identifying and categorizing phrases that fall into predefined entity groups, each critical for deciphering the complex narratives within the data:

1. Failure Symptom: This category captures the explicit details of failures, such as 'png stopped firing' or 'no porosity data.' Identifying these allows for a precise understanding of the failure characteristics.

2. Data Channels: These entities encompass technical log parameter values like 'state changing to 2304.' Their recognition is vital for correlating technical readings with failure events.

3. Operational Actions: This group includes actions taken in response to issues, such as 'downlink to shutdown png.' Understanding these actions aids in assessing the effectiveness of operational responses.

4. Drilling Conditions: These entities report on the physical conditions during drilling, like 'top cement was tagged.' Recognizing these conditions is essential for contextualizing failures within their operational environment.

The proposed method consists of three steps: data collection, data preparation, and modelling, as illustrated in Figure 3.

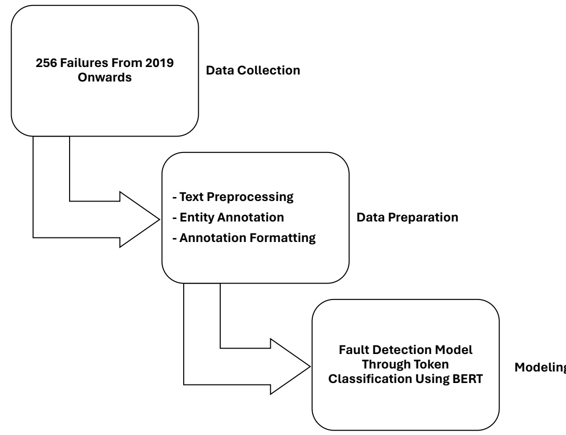


Figure 3. Proposed Method.

3.1. Data Collection

The dataset comprises 256 failure descriptions, extracted from internal database tables, focusing on data from the year 2019 onwards. This time restriction was applied because the majority of relevant and validated data, as identified by the SME, begins from this period.

3.2. Data Preparation

The data processing involved three main steps: text preprocessing, entity annotation, and adaptation to the dataset format of the (*Hugging Face*, Accessed: 2024-05-27) library, a widely-used platform that provides pre-trained models for natural language processing tasks:

1. Text Preprocessing: In this step we concentrated on removing non-alphanumeric characters and converting all text to lowercase to ensure uniformity and reduce complexity in the dataset. Additionally, we removed sections of text that originated from application-dependent formatting, such as incident dates and job numbers, as these did not contribute valuable information for fault detection.
2. Entity Annotation: We utilized syntactic strategies to ensure qualitative consistency throughout the entire dataset. A significant decision in this process was to include verbs and adverbs in all entity categories. This approach was adopted to maintain grammatical consistency across entities, which is crucial for reducing the risk of misclassification. The annotation was executed by the SME using the Doccano application, resulting in a JSON (JavaScript Object Notation) file, a format used for storing structured data. This file contains failure descriptions and lists of three chained elements, detailing the starting character, ending character, and associated category for each entity. While only 256 reports were labeled, the annotated data itself is extensive, with each failure description often containing multiple annotated phrases and entities.

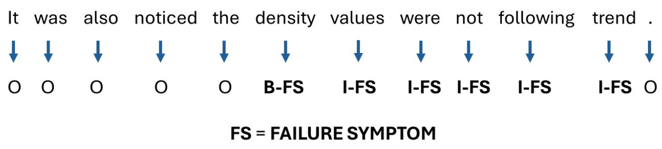


Figure 4. Hugging Face IOB Formatting.

3. Annotation Formatting: We converted generated annotations into a format compatible with the Hugging Face token classification paradigm. After validating the model, a corresponding tokenizer was utilized to divide the text into subwords. Subsequently, the character spans from Doccano were converted into Inside-Outside-Beginning tagging (IOB) format, a requirement for Hugging Face. An example of such formatting can be observed in Figure 4, where some tokens in a sentence are assigned the corresponding IOB formatted *FAILURE SYMTPOM* labels. The total amount of BI tags is $11,333$ as illustrated in Figure 5. The tokens ids and attention masks were also retained during the final conversion to the Dataset object. This object adheres to best practices for data splitting in cross-validation, featuring a distribution of 204 instances for training, 26 for validation, and 26 for testing.

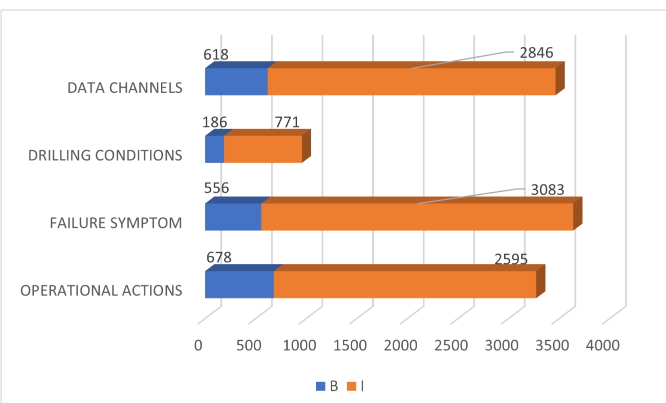


Figure 5. Bar Distribution Graph of BI-tags.

3.3. Modeling

Formally, we can define every failure description report, X, as a sequence of tokens:

$$X = \{x_0, x_1, ..., x_n\} \tag{1}$$

where x_i represents the i_{th} token in the report.

Our objective is to obtain a sequence of predicted labels, Y,

$$Y = \{y_0, y_1, ..., y_n\} \tag{2}$$

where y_i is the label of the x_i token.

For this purpose, we began by examining various unsupervised learning models to extract insights from text. However, we found that at the time of solution development, none were trained on an industry-specific corpus. Most available NER models were designed to identify general entities like date, location, person, and company, which did not align with our business-specific needs.

After completing the annotation of entities, we undertook a comparative analysis of several NLP models. This included, but was not limited to, Spacy NER, ”distilbert-base-uncased”, and ”bert-base-cased”. Our objective was to identify the most suitable model for our fault detection use case. Among these, ”bert-base-cased” demonstrated initially promising results. Consequently, we focused on fine-tuning its hyperparameters to optimize performance.

The BERT model is pretrained on two tasks: masked language modeling and next sentence prediction. Each token in BERT is represented by a combination of its token embedding, segment embedding, and position embedding. During our fine-tuning process, we utilized BERT's default activation function, GELU, along with a final classification layer that employs a softmax function to determine class probabilities. For token classification, we extract the hidden layer representation of each token and apply the softmax function to compute the probabilities for each class.

The formula for the softmax function is detailed below, where K is the number of classes and z_i is the output of the classification layer:

$$Z = [z_0, z_1, ..., z_K] \tag{3}$$

$$s(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{4}$$

4. RESULTS

In line with the time constraints of our business objectives, we explored the hyperparameter space for the ”bert-base-cased” model exhaustively. To evaluate our model's performance, we primarily focused on the F1-score, a metric derived from precision and recall. The formulas for precision, recall, and F1-score are provided below:

$$precision = \frac{TP}{TP + FP} \tag{5}$$

$$recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 - score = 2(\frac{precision * recall}{precision + recall}) \tag{7}$$

Entity	Precision	Recall	F1-score	No. of Instances
B-DATA CHANNELS	0.92	0.94	0.93	3816
I-DATA CHANNELS	0.81	0.74	0.78	2378
B-DRILLING CONDITIONS	0.50	0.18	0.27	227
I-DRILLING CONDITIONS	0.59	0.27	0.37	1073
B-FAILURE SYMPTOM	0.63	0.80	0.71	450
I-FAILURE SYMPTOM	0.72	0.79	0.75	2711
B-OPERATIONAL ACTIONS	0.50	0.78	0.61	460
I-OPERATIONAL ACTIONS	0.59	0.49	0.54	1708
macro avg	0.69	0.64	0.68	4762
weighted avg	0.88	0.88	0.88	4762

Table 1. Classification Report of BERT Checkpoint.

This effort led to a satisfactory model checkpoint, achieving an F1-score of 88%. This score reflects the model's precision and reliability in recognizing entities (See Table 1).

The most effective entities identified by the model were Data Channels and Failure Symptom, with weighted averages of 81% and 74% respectively. These averages were calculated by multiplying the F1-score by the corresponding support value and dividing by the sum of support for each BI-tags pair. Operational Actions showed average performance with a 55% score, while Drilling Conditions lagged at 35%. The lower performance in these categories correlates with a reduced token support count, indicating the model's sensitivity to class imbalances.

Our analysis of the text data revealed that the weakly predicted categories had less variance in our validated data sample, underscoring the need for more representative data for these categories. Other insights pertain to the nature of technology failures. We observed that there are a limited number of ways in which failures manifest, and this limitation aided the model's performance in identifying well-represented entities. Specifically, certain phrases indicating a specific techlog parameter value or failure symptom recur more frequently in the failure descriptions.

5. Conclusions

In this paper, we presented a fault detection NLP-based method from maintenance logs. The methods builds on identifying four technical-defined entities, essential to the failure investigation process. This approach entailed fine-tuning "bert-base-cased" model which achieved an F-1 score of 88%, underscoring the model's precision and reliability in recognizing critical entities.

The practical implications of our work are significant, with the potential to improve operational decision-making through enhanced pattern recognition in historical failure data. The impact of our findings is geared towards improving operational efficiency, reducing downtime, and cutting costs.

In future work, we aim to expand our research from identifying failures to comprehensively diagnosing them. This will involve a more detailed examination of failure events to extract insights into their causes and impacts. By advancing from simple detection to in-depth diagnostics, we will offer not just identification but also solutions.

Additionally, we intend to develop a Case-Based Reasoning system that will complement our NLP framework. This system will feature a similarity model to gather and compare similar cases, bringing forward solutions that have been effective in the past. This enhancement is expected to not only pinpoint failures but also recommend validated resolutions, thereby streamlining the path from problem recognition to problem-solving.

The integration of the CBR system is expected to leverage historical insights and expert knowledge, evolving into a dynamic model that improves with each new dataset. This step will mark a significant advance in intelligent fault detection systems, pushing the boundaries of what is currently possible in operational efficiency and safety.

References

Brundage, M. P., Sexton, T., Hodkiewicz, M., Dima, A., & Lukens, S. (2021). Technical language processing: Unlocking maintenance knowledge. *Manufacturing Letters*, *27*, 42-46. doi: https://doi.org/10.1016/j.mfglet.2020.11.001

Hansen, R., & White, J. (1991). Features of logging-while-drilling (lwd) in horizontal wells. In *Spe/iadc drilling conferencespe/iadc drilling conference.* doi: 10.2118/21989-MS

Hugging Face. (Accessed: 2024-05-27). `https://huggingface.co/`.

Juan Pablo Usuga Cadavid, S. L. R. P., Bernard Grabot, & Fortin, A. (2020). Valuing free-form text data from maintenance logs through transfer learning with camembert. *Enterprise Information Systems*, *16*(6), 1790043. doi: 10.1080/17517575.2020.1790043

Kang, J., Varnier, C., Mosallam, A., Zerhouni, N., Youssef, F. B., & Shen, N. (2022). Risk level estimation for electronics boards in drilling and measurement tools based on the hidden markov model. In *2022 prognostics and health management conference (phm-2022 london)* (p. 495-500). doi: 10.1109/PHM2022-London52454.2022.00093

Lee, M., & Marlot, M. (2023). Information Retrieval from Oil and Gas Unstructured Data with Contextualized Framework. , *2023*(1), 1-5. doi: https://doi.org/10.3997/2214-4609.202332039

Martin, L., Muller, B., Suárez, P. J. O., Dupont, Y., Romary, L., de la Clergerie, É. V., ... Sagot, B. (2019). Camembert: a tasty french language model. *CoRR*, *abs/1911.03894*. Retrieved from `http://arxiv.org/abs/1911.03894`

Mosallam, A., Kang, J., Youssef, F. B., Laval, L., & Fulton, J. (2023). Data-driven fault diagnostics for neutron generator systems in multifunction logging-while-drilling service. In *2023 prognostics and health management conference.*

Mosallam, A., Laval, L., Youssef, F. B., Fulton, J., & Viassolo, D. (2018). Data-driven fault detection for neutron generator subsystem in multifunction logging-while-drilling service. In *PHM society european conference.*

Mosallam, A., Youssef, F. B., Sobczak-Oramus, K., Kang, J., Gupta, V., Shen, N., & Laval, L. (2023). Data-driven degradation modeling approach for neutron generators in multifunction logging-while-drilling service. In *2023 prognostics and health management conference.*

Naqvi, S. M. R., Ghufran, M., Meraghni, S., Varnier, C., Nicod, J.-M., & Zerhouni, N. (2022). Human knowledge centered maintenance decision support in digital twin environment. *Journal of Manufacturing Systems*, *65*, 528-537. doi: https://doi.org/10.1016/j.jmsy.2022.10.003

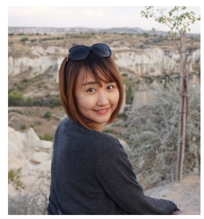
Naqvi, S. M. R., Varnier, C., Nicod, J.-M., Zerhouni, N., & Ghufran, M. (2022). "Leveraging Free-Form Text in Maintenance Logs Through BERT Transfer Learning". In L. Troiano, A. Vaccaro, N. Kesswani, I. Díaz Rodriguez, & I. Brigui (Eds.), *Progresses in artificial intelligence & robotics: Algorithms & applications* (pp. 63–75). Cham: Springer International Publishing.

Stenström, C., Al-Jumaili, M., & Parida, A. (2015). Natural language processing of maintenance records data. In *International journal of comadem* (Vol. 18).

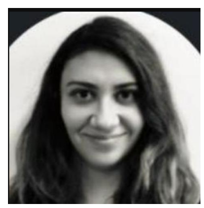
Wang, K., Reimers, N., & Gurevych, I. (2021). TS-DAE: using transformer-based sequential denoising auto-encoder for unsupervised sentence embedding learning. *CoRR*, *abs/2104.06979*. Retrieved from `https://arxiv.org/abs/2104.06979`

BIOGRAPHIES

Corina Maria Panait is a Data Scientist at SLB Romania, within the Data Science & AI Hubs. Her 1.5 years of experience in the company concentrate on NLP solutions, designed to obtain quick insights from unstructured text data in the following business areas: PHM, Health & Safety and Generative AI. Corina holds a Master's in Data Science from the University of Bucharest and a Bachelor's in Economics Cybernetics from the Bucharest Academy of Economic Studies.



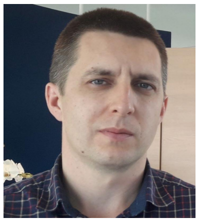
Qian Su is currently a Technical Engineer for the Offshore Atlantic Basin Well Construction. In her role, she provides technical support with a focus on electronics for failure investigations, conducts reliability improvement initiatives, and digital applications. Qian holds a Master's degree in Electrical Engineering from the Institut National des Sciences Appliquées de Lyon.



Nahieli Vasquez is a Data Scientist with SLB France for around a year and a half. She has several years of experience working in NLP applications. Prior to this position, she has worked in applying NLP methods to curriculum vitae scoring and social network moderation. Her areas of interest include NLP, Gen AI, Cloud services and End-to-end product development. She holds a Master Degree in Quantitative Economics and Econometrics from El Colegio de Mexico.

Ahmed Mosallam is the Data Science & AI European Hub Manager at SLB technology center in Clamart, France. He has his Ph.D. degree in automatic control in the field of PHM from University of Franche-Comte in Besanc¸on, France. His main research interests are signal processing, data mining, machine learning and PHM.

Fares Ben Youssef Fares Ben Youssef is the Reliability and COSD Manager, where he leads a team of Engineers and Technicians focused on several aspects of Well Construction for Offshore Atlantic basin. His responsibilities include delivering Fit for Basin projects, which involve digital, material, mechanical, and electrical design changes tailored to the Offshore Atlantic basin requirements. In 2011, Fares earned a Master's degree in Electronics and Telecommunication Engineering from the University of Paris-Saclay, France.



Olexiy Kyrgyzov graduated with B.S. and M.S. in CS in 2001 and 2002, respectively, from Dnipro National University (Dnipro, Ukraine). He received his Ph.D. in EE from the Northeastern University (Boston, USA) in 2010, where he worked on research projects as an assistant of Prof. Deniz Erdogmus. The topic of his dissertation was 'Non-redundant tensor decomposition'. After graduation he moved to France and worked on a wide range of projects in state research centers. He is currently a senior data scientist in DS&AI Hub Eur at SLB (Clamart, France). His research focuses on machine learning and artificial intelligence with applications to signal processing, natural language processing, and image analysis. He works to solve real-world Oil & Gas data analysis problems with delivering final compact and efficient products.

Hassan Mansoor is Full Stack Developer at SLB Poland. He holds a Master's Degree in Software Engineering from Poznan University of Technology, Poland. He has over 7 years of experience as a software developer, where he has worked with several cloud and web technologies.

Anup Arun Yadav is an AI Application Manager at SLB, PITC. He holds a Master's Degree in Computer Science from Pune University, India. He has over 13 years of experience as an application developer, where he has worked with several cloud and web technologies.