# Enhancing Distribution Grid Failure Data by Leveraging AI-driven text classification: A Danish Case Study

**Sundsgaard, Konrad; Yang, Guangya; Hansen, Jens Zoëga; Cafaro, Massimo**

*Published in:*
CIGRE Session 2024 Papers & Proceedings

*Publication date:*
2024

*Document Version*
Peer reviewed version

# Enhancing Distribution Grid Failure Data by Leveraging AI-driven text classification: A Danish Case Study

**Konrad SUNDSGAARD***
**Green Power Denmark**
**Denmark**
**ksu@greenpowerdenmark.dk**

**Guangya YANG**
**Technical University of Denmark**
**Denmark**
**gyyan@dtu.dk**

**Jens Zoëga HANSEN**
**Green Power Denmark**
**Denmark**
**jzh@greenpowerdenmark.dk**

**Massimo CAFARO**
**University of Salento**
**Italy**
**massimo.cafaro@unisalento.it**

**SUMMARY**

Text classification techniques based on supervised machine learning or natural language models are useful tools for automatically processing text and transforming unstructured text into structured data. However, to successfully apply them to specialised tasks within the power system and distribution grid domain, they need to be contextualized. This requires suitable training data. Fortunately, over the last fifty years, Danish distribution system operators have been saving text-based information in a database on power grid failures. This study leverages the high volume of the failure comments present in the database for the training of text classification models. To illustrate the potential benefits and capabilities of the models, a use case for estimating the root causes of medium-voltage cable failures is proposed. The results indicate that text classification models can positively contribute to data imputation and quality assurance. The study ends with emphasising on how the method could enable more robust models and expand to further applications or domains such as data convergence or named entity recognition.

**KEYWORDS**

Distribution Grid Failures, Data Quality, Text Classification, Natural Language Models, Asset Management

## INTRODUCTION

In the modern power system, enormous amounts of data is collected, processed, and utilised. A subset of the data is however not saved in a structured format but instead represented in human-readable text. Many power system stakeholders therefore face the challenge of efficiently processing the textual information. Consequently, machine learning (ML) approaches have been leveraged to facilitate the automation of text processing and are today increasingly applied in the overall information processing of power systems [1].

Traditionally, applications for text classification in the power system domain focus on using supervised classification techniques, such as support vector machines (SVM), random forest classifiers (RFC), or logistic regression (LR) [2]. Nowadays, the application of large language models (LLM) emerges [2] [3]. One example proposes to classify electric power audit texts with multi-grained, pre-trained LLMs [1]. The authors demonstrate that pre-trained and fine-tuned LLMs, such as bidirectional encoder representations from transformers (BERT), can outperform traditional supervised classification techniques. Similarly, the authors in [5] apply LLMs to the classification of text that describes a defect and point out that overfitting can be one common issue with text classification models. They recommend integrating data augmentation (EDA) techniques for development of more robust models [6].

High-quality data on component failures is crucial for enabling accurate, reliable, and informed decisions within the asset management of distribution grids. One reason is that failure data is often used as an input for data-driven reliability models, such as the prediction of failures on low [7] and medium voltage cables [8]. Distribution grid operators, however, often struggle with data scarcity or missing information regarding the past component failures. Text classification models offer an opportunity to improve the data extracted from failure reports and impute missing information.

To develop classification models for highly specialised tasks, such as classifying text on failure reasons, types, characteristics, special training data is required. Danish distribution system operators (DSOs) have been recording the information on power grid failures in the national fault and outage statistic (EL-FAS) since 1974 [9]. These "EL-FAS reports" are equipped with a "comment field" that is largely filled with text information, where maintenance personnel save information in free text. The high amount of comments on the failures of power grid components offers the opportunity to develop LLMs or apply other supervised ML techniques to perform text classification in the context of power grid failures. Eventually, the models can transform unstructured text on grid failures into structured information and enhance the overall failure data quality.

## MAIN CONTRIBUTION

This work trains classification models on texts saved within distribution grid failure reports in Denmark. The objective is to obtain improved failure data quality by using text classification models for missing information estimation and imputation. Thereby, the models are trained to provide probabilities for the classification prediction, which can serve as quality assurance indicators during the imputation process. The main contributions of this study can be stated as follows:

- To develop text classification models based on the comment fields included in the Danish national fault and outage statistic EL-FAS and evaluate the models' efficacy and limitations.
- To showcase the use of text classification models for the estimation and imputation of missing values.

- To discuss and evaluate further use cases, such as quality assurance of failure reports, automatic report validation, and information extraction.

**EL-FAS: THE NATIONAL FAULT AND OUTAGE STATISTIC OF DENMARK**

The national distribution grid fault and outage statistic of Denmark, in Danish: “El selskabernes fejl- og afbrudsstatistik” (EL-FAS) is used in the study. The registration of faults in EL-FAS encompasses high-voltage, medium-voltage, and low-voltage grid components. Formally, it is devided after the voltage ranges :>100kV, 25-<100kV, 1-<25kV, and <1kV, which translates to the voltage levels : 400, 150 and 132kV, 30, 50 and 60kV, 10, 15, and 20kV, as well as 0,4kV. Most of the information within the EL-FAS database is saved in a structured format, which incorporates features such as failure type, reason, character, and date, as well as the number of interrupted customers or outage duration. However, besides the structured information, EL-FAS also provides a free-text field for additional comments or descriptions about the failures. Table I shows some examples of the failure report comments translated to English and linked to their respective features of asset class and component type, which are stored in a structured format. The table highlights that, even though relevant information is stored in the comment fields, the entries strongly differ in terms of information granularity, clarity, and structure. Nevertheless, this study builds on the assumption that ML models can leverage such information.

*Table I: Exemplary comments of Danish Failure reports translated to english linked to respective structured information of asset class and component type.*

| **Failure report comments (unstructured)** | **Asset Class** | **Component Type** |
|---|---|---|
| The transformer in XXXX short-circuited (possibly due to age) and tripped the high-voltage fuse. There was a bit of transformer oil on the floor afterwards. | Secondary Substation | Transformer |
| Manual deenergising XXXX. Emergency disconnection. Error in LV Busbar transformer | Secondary Substation | Low voltage switchboard |
| XXXX triggers as the termination in the same trfst XXXX in radial XXXX explodes. | Cable System | Cable termination |
| Drill tests were carried out on the XXXX site. The contractor drilled directly into the cable. The contractor has not contacted XXXX prior to the drilling, so it does not appear that cable information has been requested | Cable System | Underground cable (PEX) |

**TEXT CLASSIFICATION OF EL-FAS FAILURE REPORTS**

To showcase the potential of classification models of failure report comments while keeping computational efforts low, this study focuses on the application of traditional ML techniques on the last ten years of failure reports registered for Denmark. RFC classification models are trained for each feature, respectively. Thereby, each model fulfils the narrow task of classifying the comments according to a specific feature value and providing probabilities to indicate the classification certainty. Figure 1 exemplarily visualizes the set-up selected. Before training the models, the comments are vectorized. Simultaneously, the categorical features are transformed into numerical categories. To overcome the issue of imbalanced training data and to validate that the model identifies the classification capability from the vectorized text and not from the data distribution alone, the synthetic minority oversampling technique (SMOTE) is incorporated in the training pipeline. Furthermore, the RFC model is benchmarked against a dummy classification (DummyC) model, following the "stratified" strategy to ensure that the model provides meaningful improvements over random guessing.

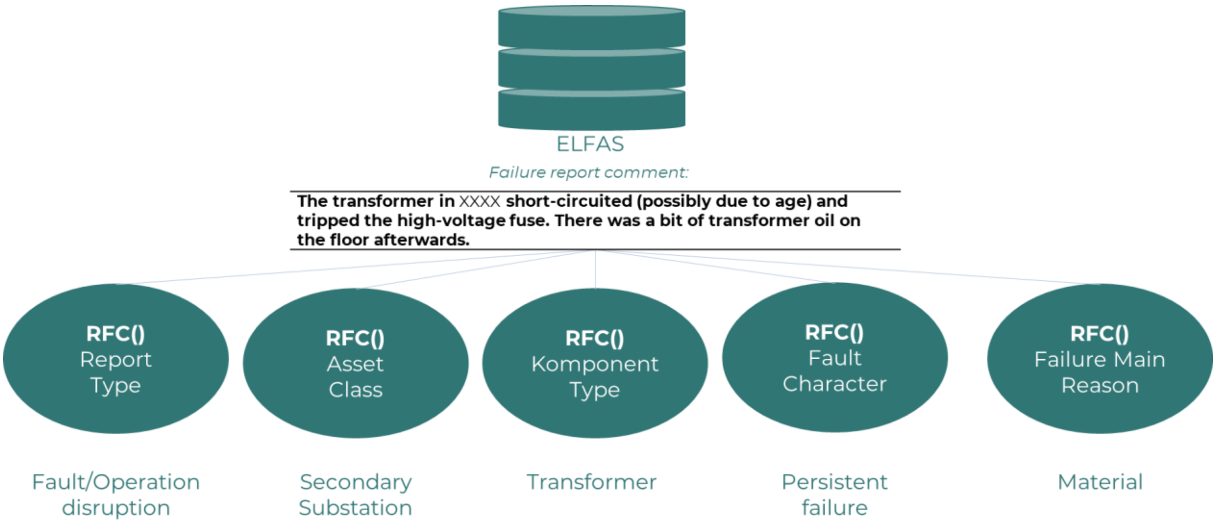


*Figure 1 : Visualisation of the set-up for ELFAS comment classification based on individual multiclass RFC() trained after probabilities.*

Table II lists the resulting performance metrics : accuracy and F1 score for the trained models. The accuracy thereby measures the proportion of correct predictions among all evaluated cases. Consequently, it captures the overall correctness of a model [10].

$$Accuracy = \frac{Number\ of\ correct\ Predictions}{Total\ Number\ of\ Predictions}$$

The accuracy assumes that all types of classification errors carry the same cost, which is generally not the case. Therfore the metrices precision and recall are added. Precision measures the model's ability to identify instances of a particular class correctly, whereas recall measures the model's ability to identify all instances of a particular class [10].

$$Precision_i = \frac{True\ positives\ for\ Class\ i}{True\ positives\ for\ Class\ i\ +False\ positives\ for\ Class\ i}$$

$$Recall_i = \frac{True\ positives\ for\ Class\ i}{True\ positives\ for\ Class\ i\ +False\ negatives\ for\ Class\ i}$$

The F1 score represents the harmonic mean of precision and recall and thus provides a balanced measure of a model's performance in a single value. In this study it is obtained by using a weighted averaging of the precision and recall values for all classes.

$$F1\ Score = 2\ x\ \frac{Weighted\ Average\ Precision\ x\ Weighted\ Average\ Recall}{Weighted\ Average\ Precision + \ Weighted\ Average\ Recall}$$

The calculated metrices, as shown in Table II, indicate that the models obtain a basic capability of classifying the comments after features, such as the report type, affected asset class, component class, and failure main or sub reason. The comparison with the results of a stratified DummyC indicates that the model learns meaningful patterns in the provided comments.
However, the overall accuracy does not reveal how well the model performs on each individual class. This can be obtained through the constructon of confusion matrix, which lists the true negatives, false positives, and false negatives per class. The inspection of the normalised confusion matrixes reveals, as shown in Figure 2, that some of the underlying values are harder to predict than others.

*Table II: Performance metrics of traditional multiclass RFC() of the failure report comments, structured after investigated features and benchmarked against a dummy classifier*

| | **RFC()** | | **DummyC(), "stratified"** | |
|---|---|---|---|---|
| **Feature** | **Accuracy** | **F1 Score** | **Accuracy** | **F1 Score** |
| Report Type | **95.93%** | **95.52%** | 32.54% | 44.32% |
| Asset Class | **93.44%** | **93.15%** | 15.67% | 21.88% |
| Failure Main reason | **86.14%** | **85.24%** | 2.80% | 4.23% |
| Component class | **80.30%** | **79.03%** | 2.92% | 4.29% |

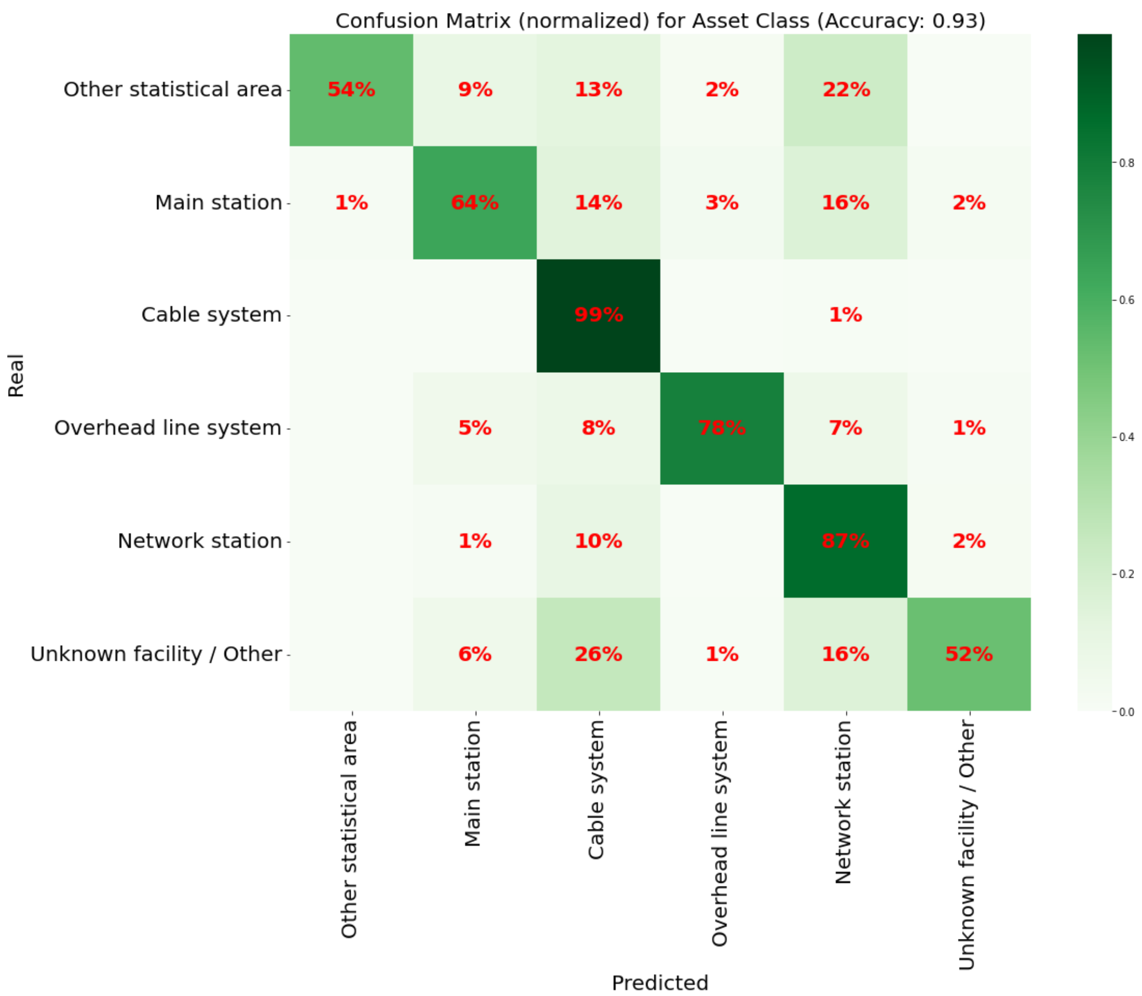


*Figure 2: Normalized confusion matrix for estimation the asset class based on failure report comment fields.*

The ELFAS database not only stores the comments but also contains tabular information. This gives the opportunity to transform the tabular information into a string format and add it to the comment fields during training. Table III illustrates, based on the asset class classification model, how model performance can be improved by combining structured data and comments. This is reasonable as many of the stored features highly correlate with each other, e.g., the asset class and the component type. For cases with high data availability, the integration of tabular data can thus strongly improve the classification performance. On the other hand, an integration of the tabular data reduces the applicability of use cases where only text strings and no further structured information is available.

*Table III : Showcase of model improvements by combining structured data and comments during the training of the asset class classification model.*

| | **RFC() – Accuracy** | |
|---|---|---|
| **Feature** | Only Comments | Structured Data + Comments |
| Asset Class | 93.44% | 99.05% |

**LEVERAGING TEXT CLASSIFICATION FOR MISSING VALUE IMPUTATION**

As outlined in [10], obtains EL-FAS data quality challenges for certain features and components, such as material properties, or failure reasons. To showcase the benefit of applying text classification to failure reports, this study illustrates the estimation of failure root causes of medium voltage cables as one practical use case to use such models to mitigate data quality issues. EL-FAS distinguishes between the "initialising cause" - what triggers the failure- and the root cause - the actual reason why the component failed due to the initialising cause.

In this context, one concern is that the failure reasons, particularly the root causes, are not always properly tracked. The root cause is a mandatory field. However, for a high amount of failure reports, either "No underlying root cause", in Danish: “Ingen dybereliggende”, or "Other (reason to be stated in report comment)”, in Danish: “Andet (årsag anføres i bemærkning til rapport)”, is selected.

This indicates that the root causes are not always stored in a tabular format but might be provided in the comment field. Therefore, to propose an estimate of the root cause, a text classification model based on the comments is proposed. Table IV presents the resulting performance metrics of the trained model and indicates that, already without considering any further structured information, the RFC strongly outperforms the dummy model. Figure 3 visualises the confusion matrix of the four most frequent root causes of medium voltage cables, “Corrosion or age”, “Lack of request for cable information during digging”, “Material, construction, or dimensioning” and “Vibration or pressure impact”. Again, it can be observed that certain root causes can be better predicted from the model than others.

*Table IV: Performance metrics of multiclass RFC() to estimate root causes of medium voltage cables benchmarked to a dummy classifier with stratified strategy.*

| | **RFC()** | | **DummyC(), “stratified”** | |
|---|---|---|---|---|
| **Predicted Feature** | **Accuracy** | **F1-score:** | **Accuracy** | **F1-score:** |
| Failure sub reason | 86.40% | 85.24% | 7.23% | 10.73% |

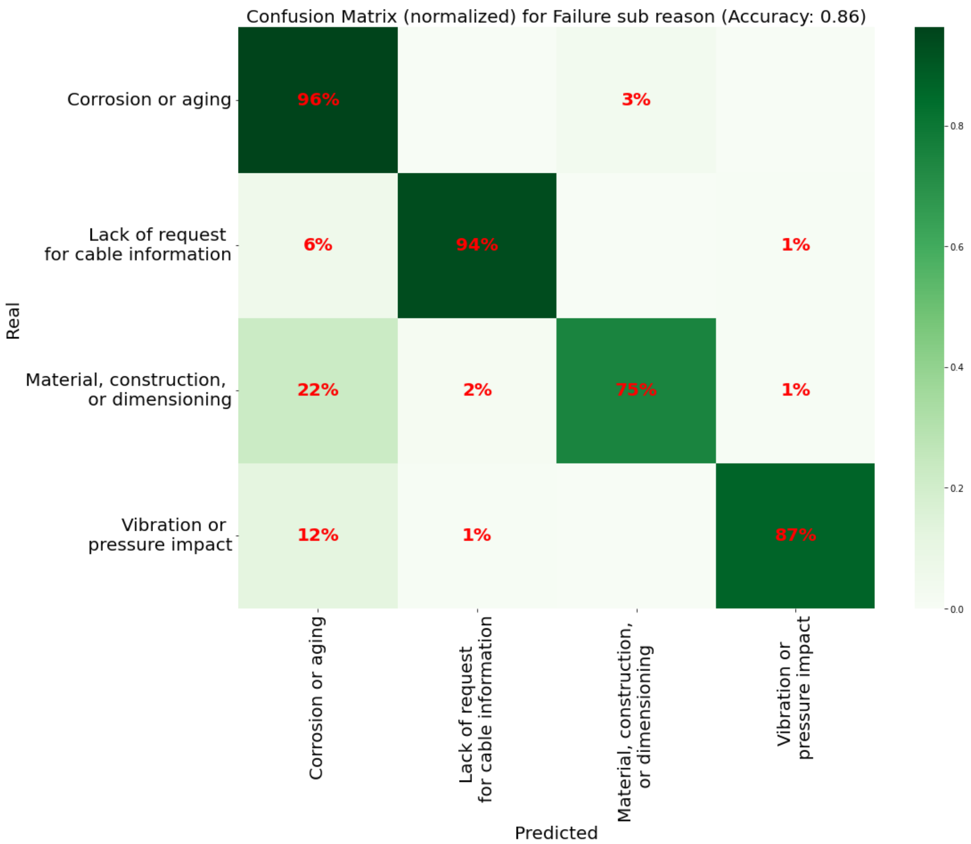


*Figure 3: Confusion matrix for RFC model to classify text on medium voltage cable failures after the root cause. For a better readability, the confusion matrix is limited to the four most frequent root causes.*

Eventually, the trained model can be used for estimating missing values in the structured data. In EL-FAS there are over 10.000 cable failures registered, for which the root cause of the failure is not further specified but where a comment field exist. By estimating the root cause of the failures through classifying the provided comment field, data quality might be improved. Together with the root cause estimate, the RFC model also provides a probability, indicating the certainty of the prediction made. This emphasizes that the prediction certainty differs from comment to comment. A high classification probability over all unseen comment fields thus indicates robust model performance and meaningful comments.
Figure 4 visualises the distribution of the predicted probabilities for the root cause estimations provided by the presented text classification model. It can be observed that the majority of the model predictions remains between 40-70% certainty, which again highlights improvement potential for future studies. Lastly, Table V presents two exemplary instances with high model certainty.

*Table V : Two exemplary selected and imputed root cause instances, where the text classification model indicates a high probability.*

| **Failure report comments (unstructured)** | **Root Cause [Imputed]** | **Probability [%]** |
|---|---|---|
| XXXX, digging damaged cable XXXX | Lack of request for cable information | 92% |
| 10 kV cable demolished by digging machine. | Lack of request for cable information | 93% |

However, such results emphasise that there remains another risk associated with the imputation through comment classification. Even though the logical link between digging related comments and a lack of requests for cable information is obvious, there still exist digging related failures, where cable data has been properly requested. If this information is not provided in the comment, the model will have difficulties to properly identify such cases.

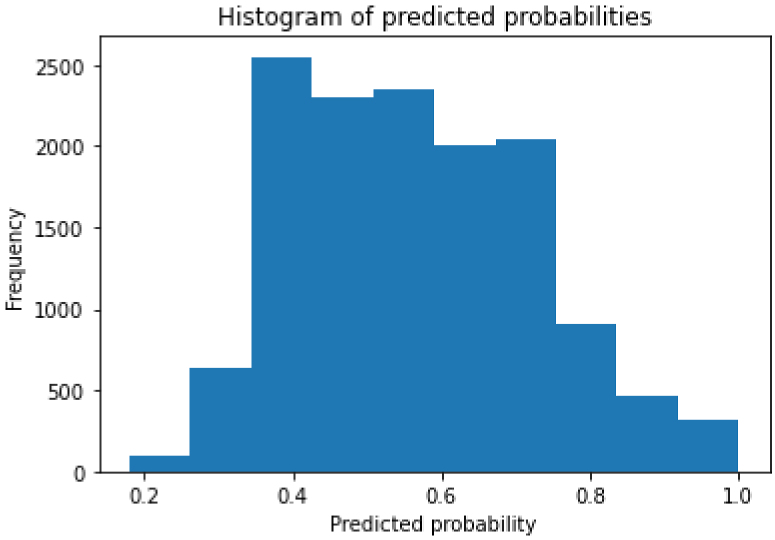


*Figure 4: Histogram of certainties for the imputed root causes, provided by the text classification model*

**DISCUSSION AND FUTURE WORK**

This work presents a first attempt to leverage text classification in the context of distribution grid failures. However, the presented applications obtain several flaws that should be tackled by future improvements. One flaw of the presented models is for example the current inability to inform if no useful prediction can be made from the given text. Ideally, the model should also raise flags for the instances where no useful information is provided.

It is known that data augmentation techniques, such as EDA, can mitigate the issue of overfitting. Consequently, future studies should integrate EDA techniques during model training to provide more robust models. In addition, naturally the available features, e.g. asset class and component type, depend on each other, where individual training of one model per feature is not practical. Therefore, future work should adapt the current model set-up so that it can estimate all feature values based on the given ELFAS comment. For this also other model designs, e.g. LLM should be considered and benchmarked against the current results.

Many of the features within the EL-FAS failure reports cannot be estimated using text classification, as they do not follow defined categories. The location, installation date or specific component-type information of the failed components are good examples of this. As a result, other natural language processing techniques, such as Named Entity Recognition (NER) might come at hand to extract the desired information from the given text. Future work should thus develop NER models for identifying non standardised information in the failure report comments. Some information, such as dates or location can be identified by pretrained NER language models right away, however, the identification of more specific information, e.g. specific cable material types, might require further fine-tuning efforts, including the construction of suitable training data. Even though complete historical failure reports can be used to automatically identify information in a subset of comments, manual efforts of naming the entities might still be necessary for many instances. Again, the integration of LLM might be beneficial, as previous research has shown that LLMs are capable of generalising for multiple downstream tasks, such as classification as well as named entity recognition. Eventually, this might enable models that not only generalise for different feature classification tasks, but also different text processing tasks in the context of distribution grid failure reports.

This study selected the estimation of root causes of medium voltage cable failures based on failure reports comments as one practical application case. However, more application areas are possible. Another case could be given in relation to medium voltage switchgear failures. For certain analysis, knowledge of the type of switchgears that failed is essential. However, such information is not always tracked properly in the failure reports. This again offers the possibility to train either a NER or classification model to estimate the correct switchgear type. The text processing of failure reports can also be applied to further data processing domains. One example is the quality assurance of failure reports, which can be a resource intensive process. ML text classification models could be leveraged to identifying inconsistencies in the reports and thus increase the efficiency of the validation process. Similar to the practical applications for missing value imputation, one can also develop use cases for report validation. One example is the incorrect registration of joint failures as cable failures.

There exist also other data enrichment techniques, that are not based on language models or text classification. Prominent examples are state of the art imputation techniques, such as miss forest (MF), k-nearest neighbours (KNN), synthetic data imputation based on variational auto encoders (VAE) or other deep learning models. It would be interesting to compare the results of the models and to investigate how such models could be applied together to obtain ideal performance.

## CONCLUSION

This study successfully applied ML text classification techniques to comment fields of Danish distribution grid failure reports. It is shown that the developed models obtain a general capability of classifying the comments after categorical values for features such as asset class, component class, failure main or root cause, etc. Furthermore, a basic approach for utilising text classification for imputing missing values within the failure report was proposed. A practical use case of estimating the root cause of medium voltage cable failures was elaborated. The discussion on future research activities points to the application of state-of-the-art LLMs and EDA techniques to increase the robustness of the current models, to the exploration of further use cases, e.g., the estimation of medium voltage switchgear types, as well as to an integration of text based and statistical based imputation techniques in a common imputation workflow.

## DISCLAIMER

This manuscript has been developed within the framework of the InnoCyPES project, financially supported by the European Union's Horizon 2020 research and innovation program, through the Marie Sklodowska Curie grant agreement No 956433.

## BIBLIOGRAPHY

[1] Q. Meng *et al.*, “Electric Power Audit Text Classification With Multi-Grained Pre-Trained Language Model”, doi: 10.1109/ACCESS.2023.3240162.

[2] J. Hartmann, J. Huppertz, C. Schamp, and M. Heitmann, “Comparing automated text classification methods,” 2018, doi: 10.1016/j.ijresmar.2018.09.009.

[3] “Getting started with the built-in BERT algorithm | AI Platform Training | Google Cloud.” Accessed: Jul. 31, 2023. [Online]. Available: https://cloud.google.com/ai-platform/training/docs/algorithms/bert-start

[4] X. Tian, C. Li, and B. Zhao, “A Novel Classification Model SA-MPCNN for Power Equipment Defect Text,” *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 20, no. 6, Aug. 2021, doi: 10.1145/3464380.

[5] J. Wei and K. Zou, “EDA: Easy data augmentation techniques for boosting performance on text classification tasks,” in *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 2019. doi: 10.18653/v1/d19-1670.

[6] M. H. P. Klerx, J. Morren, and H. Slootweg, “Analyzing Parameters That Affect the Reliability of Low-Voltage Cable Grids and Their Applicability in Asset Management,” *IEEE TRANSACTIONS ON POWER DELIVERY*, vol. 34, no. 4, 2019, doi: 10.1109/TPWRD.2019.2903928.

[7] O. Faivre, S. Folleville, and K. Guiga, “Optimizing Network Replacement with AI”, pp. 3–6, in International Conference on Electricity Distribution (CIRED), Madrid, Spain, 2019.

[8] J. Z. Hansen, “Results from Danish Failure Statistics for Medium Voltage XLPE Cables”, 22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013).

[9] K. Sundsgaard, J. Zoëga Hansen, G. Yang, M. Cafaro, and P. Kjær Hansen, “Exploring Data Collection and Fusion of Medium Voltage Cable Failures - A Danish Case Study,” in *IEEE PES Innovative Smart Grid Technologies Conference EUROPE 2023 (ISGT)*, Grenoble, France, 2023.

[10] Evidently AI. (No publication date). Accuracy, precision, and recall in multi-class classification. Evidently AI. Retrieved February 4, 2024, from https://www.evidentlyai.com/classification-metrics/multi-class-metrics.