Proposal for Semester Thesis: Analyzing Heat Pump System Failure Patterns using Natural Language Processing

Introduction

EKZ Contracting finances and operates heat pump systems in multi-family homes and building complexes. Monitoring and analyzing the performance of these systems is crucial to ensure their efficient operation and minimize downtime due to failures. Field service logs and maintenance logs provide valuable information about the operational issues and maintenance activities performed on these systems. However, manually extracting meaningful insights from these logs can be time-consuming and error-prone. This proposal aims to utilize natural language processing (NLP) techniques to automate the analysis of field service logs and maintenance logs of heat pump systems. The primary objective is to identify clusters or patterns of different failures of the systems, which will serve as labels for a larger data science project focused on failure cause identification.

Objectives

The main objectives of this semester thesis are as follows:

1. Literature review: Conduct a literature review to gain a comprehensive understanding of relevant NLP techniques. Identify existing studies or frameworks that have utilized NLP for similar purposes.
2. Preprocess and clean the data: Apply necessary preprocessing techniques to clean and standardize the textual data. This may include removing irrelevant information, normalizing text, handling missing data, and addressing inconsistencies in the logs.
3. Implement natural language processing techniques: Utilize NLP algorithms and techniques to extract relevant features from the logs.
4. Identify failure clusters or patterns: Apply clustering algorithms or pattern recognition techniques to identify distinct clusters or patterns of failures in the heat pump system logs. These clusters will serve as labels for failure identification in the subsequent data science project.
5. Evaluation and Comparison: Evaluate the performance of the selected models using appropriate metrics. Compare different NLP techniques and algorithms to identify the most effective approach for this specific task.

Data

The data provided by EKZ consists of about 3'500 field service logs regarding about 2'000 service visits at about 600 heat pump installations. The text is in german language and is saved in an excel table, including service log id, visit id, installation id, source of alarm, alarm information, status, date of log and visit and the actual text and comments that the service personnel in each log.

Expected Outcomes

Upon completion of this semester thesis, the following outcomes are anticipated:

1. An implementation of NLP techniques for feature extraction from the textual logs.
2. Identified clusters or patterns of failures in heat pump systems, which will serve as labels for failure identification in the subsequent data science project.
3. Evaluation results and performance metrics of the clustering or pattern recognition algorithms applied to the logs.

Deliverables

1. A well-documented codebase implementing the clustering and labeling models, along with preprocessing techniques and analysis scripts.
2. A report summarizing the methodology, results, and insights obtained from the analysis of field service logs.
3. Visualizations showcasing the clustered logs and their associated insights.
4. Presentation slides summarizing the project goals, methodology, results, and future directions.
Timeplan

1. Familiarization with the project (~1 week)
   a. Understand the data & the scope
   b. Read similar theses
2. Literature review and method selection (~1-2 weeks)
3. Modelling and method selection (~5 weeks)
4. Evaluation of results (~2 weeks)
5. Write report and finalize presentation (~2 weeks)