Title: Tracing Fault Causes from Fault Reports
Name: James Blackshaw
Supervisor: Dr Ramsay Taylor
Module Code: COM3610
Date: 04/10/2023

Managing and supervising the many buildings and spaces within university is a massive task. The Estates and Facilities Management team are responsible for repairing and maintaining the infrastructure belonging to the university. This encompasses groundskeeping, safety and security, equipment maintenance and more. Therefore, this team have a vital role in making sure our university runs smoothly, safely and efficiently. Without them our university would be hazardous and ill-equipped for daily operation. So how do we make their job easier?

Fault reporting by humans can often be vague and not very helpful. Sometimes there's a lack of knowledge by the reporter on the fault subject, leaving a vague or ambiguous report. Other times there could be many possible underlying issues, any one of which could be the primary culprit. One example to demonstrate is "My office is full of water!". Some possible underlying issues could be the roof is leaking, there's a burst pipe somewhere in the building, a faulty appliance or just simple human error (they left the window open). Predicting these fault reports manually and trying to find the root cause may turn out to be a labyrinthine task.

Machine learning allows us to use data and build models with the aim of identifying patterns to make predictions. This could be combined with the power of natural language processing which can extract the most important information from fault reports, leading to more accurate predictions.

By collecting prior data from Estates to be used as training data then utilising natural language processing and machine learning, we can aim to more accurately and effectively predict fault causes. This would in theory reduce the amount of time and resources used to solve the issue. With the University of Sheffield pledging to reduce carbon emissions to net-zero (The University of Sheffield, 2020) any step in reducing resource wastage and carbon footprint will help towards this goal.

Our first problem to be solved is data acquisition. This will hopefully be addressed with Estates providing various prior fault reports with solutions to which we can then use as training data for our models. After receiving this data, we would need to do some pre-processing to handle it effectively, and to make sure report formats are consistent. Various models and techniques could be applied to our problem to predict faults and propose solutions. I will need to analyse, research and implement various machine learning and natural language processing methods to compare and conclude if any of the models we create make accurate and effective predictions.

Python is the most obvious language of choice for carrying out the programming of our task. With libraries such as PyTorch, Tensorflow, SpaCy and more which could be used to aid in our development of an effective fault prediction model, python is very much suited to our task. A possible simple web application could also be developed using the Django web

framework to provide an interface for technicians to use our tool and possibly help in evaluation of our project. Or instead, I could apply unseen data to our models to analyse the accuracy of our models. These are some ideas which are definitely subject to change, but the overarching aim of our project is to develop a model which can predict causes from our fault reports, then evaluate our results.

In terms of going forward and a general plan. My first port of call would be to research any previous literature around fault reporting and predictions. By analysing this I can get accustomed to certain methods and techniques that could be required during my project and get a clearer idea of direction. I can use the conclusions found within the literature to help me and more efficiently tackle problems. This would last 1-2 weeks.

As machine learning is something I have only limited experience with, I will need to research deeper into techniques and methodologies that would be useful to implement. Any machine learning techniques such as Supervised (Neural Networks, Naive Bayes etc) or Unsupervised (K-means Clustering, Principal Component Analysis etc) learning would have to be more thoroughly researched. So, taking the time to research these methods and any literature to do with these would be quite valuable to me. So ideally, I would put 2-4 weeks of research into this.

I would then hopefully be confident enough to try and implement some ideas researched and possibly start programming some initial methods. Starting with handling the training data. Are there any text pre-processing techniques which I could implement on our data? Is there any missing data? I would spend maybe 1-3 weeks establishing a dataset ready for the more complex machine learning methods I want to implement. While also doing up any initial setup like installing software and libraries plus getting familiar with any new tools.

The bulk of the work would be on algorithm & model selection, implementation and analysis. This could take a month or more. Then a further few weeks evaluating and refining these models. I could then perhaps think of a developing a possible user interface which could be developed to use in conjunction with the tool, considering cloud deployment etc.

Then finally testing and validating on unseen data, possibly making refinements again. I should analyse my finding in detail and make conclusions on any findings. This would be reaching the final months of the project so finalising the report writing and documentation would also need to be done. Finishing off with an insightful and detailed presentation.

References

The University of Sheffield. (2020). *University of Sheffield pledges carbon neutral campus by 2030*. [online] Available at: https://www.sheffield.ac.uk/news/university-sheffield-pledges-carbon-neutral-campus-2030 [Accessed 3 Oct. 2023].