<u>Research Proposal Presentation Transcript</u>

Subject: Research Methods and Professional Practice (April 2024)

Assignment: Research Proposal

Student Name: Michael Botha

Student Number: 12686220

Programme: MSc Computer Science

Slide 1:

Good day, my name is Michael Botha. I will be presenting a research project proposal related to a master's in computer science. The title of my project is supporting a PDH network by generating predictive maintenance prompts using Machine Learning.

Slide 2:

Considering the background of the situation at hand and beginning with the research problem. Currently, Powerlink Queensland uses an aged Plesiochronous Digital Hierarchy (PDH) network to connect high voltage systems together. Faults arise within such a domain due to various reasons. For instance, degradation over time where electronic components slowly wear until their functionality is no longer possible, at which point they break down. This affects resource utilisation as well as services in cases where there is no protection. Traditionally, fault response has been reactive which takes time and significant resources to restore services in the necessary timeframe. Real-time monitoring does not reveal any failure before occurring without human intervention, and only in situations when one might be prompted to analyse source measurements and alarms of varying importance. Furthermore, historically used data analytic tools and methods may not easily correlate related variables to create trends and associated predictions without substantial and lengthy analyses (University of Delaware, 2021).

In terms of the significance of the study. Resolving this issue will decrease service downtime, which means saving money and minimising risk carried by the business. Additionally, faults could be attended to from a much more prepared position assisting with business planning. Academically, the outcomes of this endeavour will prove the efficacy of popular machine learning models and open-source tools applied to everyday engineering problems (IEEE Communications Society, N.D.). Specifically, related to dated technologies still used by companies who now carry risk as they slowly roll out transition plans. Additionally, if enough support exists assets may be sweated with assurance of performance (Deloitte Analytics Institute, 2017).

Slide 3:

The research question that follows from the problem is: Can open-source Machine Learning models and tools be used effectively, to predict PDH equipment failures in real-time, and significantly decrease unplanned network disruptions?

Slide 4:

Assessing the aim inferred by the research question. At a high level the goal would be to decrease the number of service-affecting network disruptions. To achieve this, equipment failures need to be accurately predicted. As a hypothesis, this is achievable by training a Machine Learning model using data sets currently gathered for real-time equipment monitoring purposes.

Using the aim to dictate the objectives of the study. Firstly, historical data pertaining to the various metrics used for device monitoring need to be collated (Geron, 2022). Secondly, the collected data needs to be analysed to ascertain what characteristics are of importance to appropriately fit Machine Learning models (Geron, 2022).

Thereafter, the data needs to be cleaned, meaning the removal of outliers and instances where there is information missing (Geron, 2022). Fourthly, feature vectors and labels need to be provided to feed into supervised-learning algorithms (Burkov, 2019). Once the data has been completely organised various appropriate Machine Learning models can be selected and trained (Burkov, 2019). The sixth step will then be to test the trained models using portion of the data that was set aside for test cases (Burkov, 2019). Afterwards, the best performing model can be integrated into a portion of code that can facilitate real-time data throughput, allowing for the model to undergo production-level testing (Geron, 2022). The final step will be to analyse all the results that were gathered during the study and produce findings.

Slide 5:

After perusing the key literature the following was noted. Machine Learning has risen to prominence as a solution for complex algorithmic challenges within the last twenty years. Many organisations are trying to leverage the success of this technology to catapult themselves ahead of the competition (DeepLearningAI, 2023). A significant uptake of Machine Learning research has been witnessed as its bounds, applications, and nuances are sought. Moreover, many researchers are seeking to enhance the relevant techniques to achieve greater capabilities (DeepLearningAI, 2023). In essence, Machine Learning is the programming of an algorithm by using data (Burkov, 2019). Thereafter, if any data from the same statistical distribution as the original training data is fed into the trained model predicted results will be the output.

Communications is no different from any other industry as it too has sought the application of Machine Learning in various aspects from design to operations and elsewhere (IEEE Communications Society, N.D.). Boykin (2023) gives a comprehensive overview of where telecommunications service providers are trying to use Artificial Intelligence. Amongst many examples she lists network optimisation, customer service, fraud detection, predictive maintenance, marketing and sales, as well as network security. Ahmed et al. (2020) explain how critical the use of Machine Learning will be to meet the demands of future networks in their expansive survey of research pertaining to Machine learning in the industry. Furthermore, they cover various research for wireless technologies. Other areas covered includes Software Defined Networking, 5G, and security.

Although there has been a broad coverage of Machine Learning research in the communications space, the practical application of algorithms and open-source resources has not been significantly mentioned. Moreover, predictive maintenance has been mentioned but not expanded on in terms of focus areas. Additionally, most of the research focuses on using Machine Learning within the context of future communications technologies rather than how it could be used to leverage currently installed assets. Lastly, it will be beneficial to pursue research into how companies can use tools and best practices within current business structures and technology landscapes.

Slide 6:

Moving on to the Research Design. The theory development approach to be used is deduction, where the existing theories of data analytics, machine learning, communications, and other computer science knowledge form the foundational presuppositions for the study (Saunders et al., 2019). Formulating a hypothesis from the research question and known presuppositions yields the following statement. Open-source Machine Learning models and tools can be used effectively to predict PDH equipment failures within the Powerlink Queensland context, by being trained by equipment sensor data records, to significantly decrease unplanned network disruptions. This approach leads to the use of a quantitative methodological direction, as various numerical key performance areas and indicators will be used to assess the efficacy of the outcomes (Saunders et al., 2019). For instance, selected historical alarms, fault tickets, and sensor measurements stored in a database will be gathered as a sample. With regards to the research strategy, experiments will be planned and executed to assess the trained model's responsiveness to degradation introduced into the system being monitored (Saunders et al., 2019). For example, faults can be introduced into a PDH network to see the reaction of the trained model. Data collection will require both the collation of historical data as well as new data created through experimentation (Saunders et al., 2019). Both data sources are stored in databases on premises, therefore, require appropriate extraction techniques. Lastly, the analysis techniques to be used will be of a descriptive and inferential statistics nature, dissecting the various numerical data produced (Saunders et al., 2019). For instance, considering the frequency of faults before and after the application of a trained model and whether this is significantly different to the original distribution of values.

Slide 7:

Examining the important aspects of ethical considerations and risk management. Beginning with ethics; at a high level the overarching professional and ethical standards presented by bodies such as the British Computer Society need to be maintained (BCS, N.D.). For instance, the unbiased reporting of findings and fair representation of data. Additionally, standard academic research practices pertaining to ethics must be adhered to such as laid out by Dawson (2015) in his explanation of project requirements. In terms of the particular features of this study that need special consideration, dataset anonymity will be important. That is, certain information such as site names will need to be replaced with names like “site 1”, to keep details anonymous. Another important consideration is data confidentiality, where data which is not open to the public domain and could put the company at risk in any way should not be shared. Lastly, all stakeholders should be made aware of the project goals, implications, and expectations.

Pertaining to risk management, equipment failures will need to be invoked during experimentation. Therefore, it will be important to ensure that this is done in a test environment to prevent any affect on live services. Secondly, the criticality of the infrastructure means that any impeding of operations must be prevented. This can be controlled by following standard business protocol where applicable. Thirdly, with the access of data archives it will be necessary to follow data processing best practices to prevent any loss of data. Lastly, it will be important to focus on practical and academic outcomes, to ensure that company resources are not wasted.

Slide 8:

In terms of the artefacts that will be produced. A trained Machine Learning model will be the core outcome, however, other supporting software will need to be produced. For instance, backend code which will digest real-time data through various application programming interfaces. Additionally, some form of user interface will be required to present infographics and highlight assets of concern presented as output of the trained model.

Slide 9:

Considering the project timeline and assuming the allowable total time to be roughly eight months or thirty-two weeks; gathering of data should take three weeks. Data analysis can be run consecutively with data cleaning and organising which will take a total of three weeks. Six weeks has been assigned to train the models. After a two-week offset from training, testing will begin and also run for six weeks. Testing the trained model in a real-time environment has a four-week allocation. Lastly, analysis of results and drawing up the final report are planned to run consecutively over an eight-week period. One will note that the total allocated time is not thirty-two weeks. This has been planned to allow for any deviations or challenges that might affect the project's timeframe.

References:

Ahmad, I., Shahabuddin, S., Malik, H., Harjula, E., Leppänen, T., Lovén, L., Anttonen, A., Sodhro, A., Alam, M., Junttiantti, M., ylä-jääski, A., sauter, T., Gurtov, A., Ylianttila, M., Riekki, J. (2020) Machine Learning Meets Communication Networks: Current Trends and Future Challenges. *IEEE Access* vol. 8: 223418-223460. DOI: 10.1109/ACCESS.2020.3041765.

BCS. (N.D.) BCS Code of Conduct. Available from: https://www.bcs.org/membership-and-registrations/become-a-member/bcs-code-of-conduct/ [Accessed 06 July 2024].

Boykin, F. (2023) ‘AI Use Cases in Telecom’ [Slideshow Presentation]. Available from: https://www.youtube.com/watch?v=UDIRRGRLGNM [Accessed 07 July 2024].

Burkov, A. (2019) *The Hundred-Page Machine Learning Book*. Andriy Burkov. Available from: https://leanpub.com/theMLbook [Accessed 18 June 2024].

Dawson, C. (2015) *Projects in Computing and Information Systems: A Student's Guide*. Harlow: Pearson.

DeepLearningAI. (2023) Machine Learning Specialization. Available from: https://www.youtube.com/watch?v=wiNXzydta4c&list=PLkDaE6sCZn6FNC6YRfRQc_FbeQrF8BwGI&index=2 [Accessed 07 July 2024].

Deloitte Analytics Institute. (2017) Predictive Maintenance: Taking pro-active measures based on advanced data analytics to predict and avoid machine failure. Available from: https://www2.deloitte.com/content/dam/Deloitte/de/Documents/deloitte-analytics/Deloitte_Predictive-Maintenance_PositionPaper.pdf [Accessed 06 July 2024].

Geron, A. (2022) *Hands-On Machine Learning with Scikit-Learn, Keras & Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 3rd ed. Sebastopol: O'Reilly Media.

IEEE Communications Society. (N.D.) Machine Learning in Communications and Networks: Call for Papers (Final Issue). Available from: https://www.comsoc.org/publications/journals/ieee-jsac/cfp/machine-learning-communications-and-networks#:~:text=Modern%20machine%20learning%20techniques%20provide,routing%2C%20transport%20protocol%20design%2C%20andc [Accessed 06 July 2024].

Kepner, J., Gadepally, V. (2020) Artificial Intelligence and Machine Learning [YouTube Video]. RES.LL-005 Mathematics of Big Data and Machine Learning. Massachusetts Institute of Technology: MIT OpenCourseWare

Saunders, M., Lewis, P., Thornhill, A. (2019) *Research Methods for Business Students*. Eighth Edition. United Kingdom: Pearson Education Limited.

University of Delaware. (2021) Machine Learning vs. Statistics. Available from: https://onlinestats.canr.udel.edu/machine-learning-vs-statistics/ [Accessed 06 July 2024].