Swinburne University of Technology

Start Date: September 16th 2024

Due Date: November 1st 2024

COS40007: Design Project

Group 2 - Theme 3

Submitted By:

Arnob Gosh

103494114

Nguyen Tran

103844463

Alexandre Cluzet

105318766

Jacob F. Larsen

105323472

Nick Watson

103600076

Contents

1 Introduction

1.1 Background and motivation

In the future, several members of our group aspire to become industrial engineers with a focus on optimizing production chains. They are particularly interested in how systems can be streamlined to improve efficiency and reduce waste. The project on Vegemite emerged as a perfect example of such optimization in practice. After some discussion, we unanimously agreed to pursue this project, partly because one group member is an avid fan of Vegemite, which added an extra layer of enthusiasm. This AI model is intended for companies looking to optimize their production chain, detect anomalies. By integrating this model into their systems, company can optimize their production chains but also detect potential disruptions before they escalate, leading to more consistent product quality and reduced downtime. Overall, the project aims to provide practical tools for companies seeking to improve performance and ensure long-term sustainability.

1.2 Project objectives

- Determine the recommended values of machine settings during production process for different class of product quality.
- Determine what anomalies can occur in production process for which a production run can fail.

We aim to develop an AI that can determine the optimal machine settings to achieve the desired product quality and analyze anomalies, identifying where errors in the machine settings occur.
For users, the AI model will provide actionable insights on machine settings adjustments in real time and alert operators to potential failures. This would improve efficiency, reduce waste, and increase the reliability of product quality during the manufacturing process.
Through this project, We would like to learn how to develop predictive machine learning model that can recommend optimal operational settings and also detect anomalies.

1.3 Project schedule

The Project Schedule will briefly be expected like below. As the project progress, more information and tasks might be added for better result:

Week	Weekly Achievement	Arnob	Alexandre	Jacob	Nick	Nguyen
7	Group Formation	Attend Tutorial, Forming Team and Divide Task				
8	Project Brief	Data	Introduction	Data	Requirements	Introduction
9	Project Training and Development	Data Pre-processing	Training and Validating Data		Data Labeling	
10	Project Finalise / Presentation Preparation	AI Demonstration of Final Selected Model			Powerpoint Slide Preparation	
11	Presentation	Powerpoint Presentation / Report Preparation				
12	Project Demo / Report Preparation	Demonstration of Final Selected Model / Receiving Feedback and Prepare Final Report				

Figure 1: Project Schedule

2 Data

2.1 Data Source

The data for this project has been derived from a combination of machine sensor readings and machine settings during the processing and production of Vegemite. This data has been differentiated into multiple .csv files based on the output consistency quality of the product, labelled "good", "low bad" and "high bad". Within these files are logs of all available datapoints, separated by the time of settings configuration. These files will be used to provide an insight into which values are correlated with better or worse quality outcomes. The available data columns consist of batch IDs, raw materials used, set points for controllable components and observed values from sensors, with each series possessing a prefix (TFE, FFTE or Extract Tank) to identify which system of the production line the data has been taken from. Along with the output quality data, this dataset also contains an information log of shutdowns across a two month period in another folder named "Downtime". Once a machine learning model has been developed, this data will be used to predict shutdown events ahead of time based on readings gathered from the system.

2.2 Data Processing

This data appears to require substantial pre-processing before it can be effective for a machine learning model. While little data cleaning will be required, the formation of correlation matrices, hyperparameters and class labelling is needed to improve the readability of the data and learning outcomes of the model.

2.3 Data Exploration and Analysis

To better understand the dataset, an exploratory data analysis (EDA) will be carried out prior to the AI model being built. Visualizing the distribution of machine settings, looking at correlations between various parameters, and spotting trends or patterns that affect product quality are all part of this process. Finding outliers, missing values, and inconsistencies that could have an impact on the model's performance is another task for the analysis. These understandings will direct the selection of features most pertinent to production outcome prediction and help guarantee that the data is in the best possible condition for the training process.

2.4 Feature Selection and Engineering

Choosing the right features and applying engineering are essential to maximizing the performance of the AI model. Through the process of examining the correlations among various variables, the most pertinent characteristics that impact product quality and possible machine malfunctions will be found. To choose the most influential features, methods like feature importance ranking, correlation analysis, and domain expertise will be used. Additionally, to give the model more

insightful information, new features could be created by combining or altering already-existing data points. By ensuring that the model is trained with only the most instructive and pertinent data, this procedure will improve the model's overall efficacy and predictive accuracy.

3 Requirements

3.1 Must-have functionalities

The model should be able to predict and recommend setpoints for each machine relative to the current process values and product state.
It should also be able to predict machine breakdowns that are related to the current production and machine setpoints.
A simple user interface showing the current model recommendations, this could for example be done through the terminal.

3.2 Optional features

A nice graphical user interface with a graphical representation of the physical machines would be nice, but is perhaps beyond the scope of this project.
It would be good if the model is fed data by the machine itself and can react to changes in settings in real time.