DATA MINING ON MACHINE BREAKDOWNS AND EFFECTIVENESS OF SCHEDULED MAINTENANCE

By:

TAN SU SZE

(Matrix no.: 129378)

Supervisor:

Assoc. Prof. Ir. Dr. Chin Jeng Feng

May 2019

This dissertation is submitted to
Universiti Sains Malaysia
As partial fulfilment of the requirement to graduate with honors degrees in
BACHELOR OF ENGINEERING
(MANUFACTURING ENGINEERING WITH MANAGEMENT)

School of Mechanical Engineering
Engineering Campus
Universiti Sains Malaysia

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently in candidature for any degree.

Signed ……………………………………………………...…...…….. (TAN SU SZE)
Date…………………………………………………….…...…...….......

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by giving explicit references. Bibliography/references are appended.

Signed ……………………………………………………...…........... (TAN SU SZE)
Date…………………………………………………….…...…...….......

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for interlibrary loan, and for the title and summary to be made available outside organizations.

Signed ……………………………………………………...…........... (TAN SU SZE)
Date…………………………………………………….…...…...…...

Witness by

Supervisor: Assoc. Prof. Ir. Dr. Chin Jeng Feng

Signed……………………………………………………...…...….......
Date…………………………………………………….…...…...….....

ACKNOWLEDGEMENT

My project would not have completed without the aid and support from various parties. First and foremost, I would like to extend my greatest gratitude to my supervisor, Assoc. Prof. Ir. Dr. Chin Jeng Feng for his continuous assistance and guidance given throughout the whole project process. I also would like to thank my supervisor for his unconditional support and valuable information which cleared my doubts and leaded me to the right direction.

In addition, I would like to thank my parents, course mates and friends who have gave comments, suggestions and encouragement to allow me generated more ideas and be more positive throughout this journey of completion of work.

Last but not least, I would like to appreciate all who have directly or indirectly supported me throughout this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviations	Representation
AE	Acoustic Emission
ANNs	Artificial Neural Networks
CAD	Computer-Aided Design
CAM	Computer-Aided Manufacturing
CAR	Class Association Rule
CFN	*cascadeforwardnet*
CIM	Computer Integrated Manufacturing
CNC	Computer Numerical Control
CSV	Comma-separated Value
DSSA	Decision Support System Architecture
EXACTS	Expert Computer Aided Tool Selection System
GA	Genetic Algorithm
GP	Genetic Programming
GUI	Graphic User Interface
KDD	Knowledge Discovery in Databases
MES	Manufacturing Execution System
ML	Machine Learning
NCF	*newcf*
OPC-UA	Open Platform Communication-Unified Architecture
UML	Unified Modelling Language
WEKA	Waikato Environment for Knowledge Analysis

ABSTRAK

Dalam pembuatan komputer bersepadu, mesin akan dilengkapi dengan sensor dan penyimpanan ingatan. Ini menyediakan corpus maklumat yang boleh diperolehi semula untuk penemuan data dan pengetahuan. Kajian kes ini memberi tumpuan kepada penyiasatan kerosakan mesin dan keberkesanan penyelenggaraan berjadual dengan penggunaan perlombongan data. Metodologi penyelidikan melibatkan lapan langkah. Langkah pertama adalah melakukan kefahaman perniagaan bagi mendapatkan gambaran mengenai perlombongan data pada persembahan mesin. Beberapa soalan di peringkat tahap makro dan mikro telah dihasilkan untuk menyiasiat kerosakan mesin dan keberkesanan penyelenggaraan berjadual. Kedua, satu model simulasi akan dicadangkan dan direka untuk proses operasi berdasarkan penerangan pembuatan peralatan perubatan industri yang sebenar. Pengeluarannya adalah job shop di mana produk dalam kumpulan perlu melalui beberapa proses dan multi-stesen. Setiap proses mengubah ciri-ciri produk tertentu. Simulasi pengeluaran akan dibina di dalam Witness Horizon V21. Enam senario kerosakan mesin yang berbeza telah dimodelkan. Berbeza-beza strategi pemprosesan ciri akan dibuat, khususnya data berkaitan masa. Ketiga, pangkalan data hubungan dibangunkan untuk menyimpan maklumat dari simulasi. Langkah seterusnya melibatkan pra-pemprosesan data yang termasuk pemilihan data, pembersihan data dan transformasi data. Perlombongan data adalah langkah keenam di mana perisian Orange akan digunakan sebagai alat. Ketujuh, penilaian pola dibangunkan untuk membentangkan penemuan data yang membantu dalam membuat keputusan. Daripada penyelidikan, didapati terdapat sepuluh jenis kerosakan yang mempengaruhi prestasi mesin dan kebocoran cecair penyejuk adalah penyumbang utama berbanding kerosakan lain. Selain itu, kekerapan kerosakan terutamanya untuk kebocoran ceair penyejuk telah berkurangan selepas penyelenggaraan dijadualkan pada mesin. Oleh itu, perjadualan penyelenggaraan pada mesin berbukti berkesan dalam mengawal kekerapan kebocoran. Daripada keputusan, dapat menilai kerosakan mesin dan menyokong keputusan mengatur penjadualan pada mesin. Walau bagaimanapun, ia memerlukan sejumlah besar kos untuk dilaburkan dalam penyelenggaraan. Akhir sekali, beberapa tugas perlombongan data yang kompleks tidak dapat dilaksanakan kerana algoritma dan pembelajaran mesin yang terhad dalam perisian Orange.

ABSTRACT

In computer-integrated manufacturing, machines are equipped with sensors and memory storage. This provides a large corpus of information retrievable for data mining and knowledge discovery. The case study is focused on the investigation of machine breakdown and the effectiveness of scheduled maintenance with the application of data mining. The research methodology involves eight steps. The initial step is performed business understanding to gain insight of data mining on machine performances. Several questions in the stage of macro-level and micro-level are generated. Second, a proposed simulation model for the operational process was designed based on a real medical tool manufacturing plant. The production is a job shop whereby products in batch have to go through a number of processes and multi-stations. Each process alters particular attributes of the product. The production simulation would be constructed in Witness Horizon V21. Six different machine breakdown scenarios were modelled. Different feature processing strategies would be devised, in particular time-related data. Third, a relational database is developed to store the information from the simulation. The next step is involved data pre-processing which includes data selection, data cleaning and data transformation. Data mining is the sixth step in which software of Orange is used as the tool. Seventh, the pattern evaluation is developed to present the discovery of data which helps in decision-making. From the research, it is found that there have ten types of breakdowns affecting the performance of machine and the breakdown of coolant leaking is the main contributor as compared to others breakdown. Besides, the frequency of breakdown especially for coolant leaking has decreased after the maintenance is scheduled on the machine. Hence, the application of maintenance on machine is effective in controlling the frequency of breakdown. From the results of data, it is able to evaluate the breakdown of machine and support the decision on scheduling maintenance on machine. However, it requires a large amount of cost to be invested in maintenance. Last but not least, some of the complex data mining tasks are not able to perform because of the limited algorithms and machine learning in Orange software.

Chapter 1 INTRODUCTION

1.1 Overview

This is an industry case study to scrutinize the performance of machine. The main approaches used in this case study are building a database, constructing a production simulation and performing the data mining. The project is in line with the current manufacturing trend moving towards Industry 4.0. The application could be seen as the enhancement to the manufacturing executive system (MES) in term of decision making automation.

1.2 Background

In the context of Industry 4.0, intelligent and digitization of machine is an important aspect to enhance performances of the machine. The advanced technology of machine enables the information of processing recorded as purpose of review. Data mining is the process of extracting useful knowledge and pattern from large amount of data stored in databases. Vazan *et al.* (2017) defined data mining is an interdisciplinary field with the general goal of predicting outputs and uncovering relationships in data, enabling the use of automated tools and techniques, employing the sophisticated algorithms in the purpose to discover the hidden patterns, associations, anomalies from large amounts of data. Data collection from large volumes of information is one of the challenges for data mining. Sufficient historical data or real-time information of operational machine is the prerequisite for data mining upon finding variables that contain information about machine behaviour and health status.

1.3 Objective

Data mining on machine performances is conducted to study the breakdown type of machine and how often it is happening. It is also focused on the effect of breakdown on the machine which depends on the maintenance scheduled and product manufactured.

1.4 Scope

The scope of this research is focused on investigate the breakdown of the machine and the effectiveness of scheduled maintenance. A simulation study is built in WITNESS HORIZON V21 to generate a database based on the information provided by medical tool manufacturing plant. Six different machine breakdown scenarios were modelled. Data mining is performed to discover hidden pattern and information from large data masses. Orange is used in this research. It is a data mining tool consists of a collection of widgets and Python modules for visual analysis of network data. According to Ghandi *et al.* (2018), data-driven tools in data mining is used to determine the condition and health status of operational components and performed predictive maintenance.

1.5 Organization of thesis

The first part of this chapter is an overview of data mining on machine breakdown and the effectiveness of scheduled maintenance. The data mining tool used is Python Orange software. The rest of the paper is organized as follows. Chapter 2 describes a brief of literature review about the definition of data mining, application of Computer Integrated Manufacturing (CIM) incidental to Industrial 4.0 and data mining on machine performances. Chapter 3 presents the research methodology which explains the steps of data mining involved and the data mining tool used in data pre-processing process. Chapter 4 performed results and pattern evaluation based on the data selected and answering the predefined sub-questions. Chapter 5 concludes the paper with insight obtained from data mining and prospects the future work. Last but not least, Appendices contains the data for each of the questions at macro-level and micro-level with the relevant bar chart, graph, tables and descriptions.

Chapter 2 LITERATURE REVIEW

The purpose of the literature review is to build a fundamental understanding of data mining and its application on machine performances. Recent journals for the past ten years have been collected. For the introduction of data mining, a formal definition of data mining was described and how it benefits to the current trend of Industrial 4.0. Then, the common process and requirements of data mining are explained before entering the stage of data classification. The next part was explaining the uses of Computer Integrated Manufacturing (CIM) and its ultimate goals. The latest development of CIM incidental to Industry 4.0 had been discussed. After that, several automated machines with the integrated system was introduced and its applications on benefiting the production. The result and limitation obtained from the application of CIM were discussed. In consequences, ten examples of data mining on machine performance were listed. For each of the example provided, the purpose, methods or systems used, results and limitation of the application are discussed. Lastly, a table classified the data mining tool based on purpose and method was constructed.

2.1 Introduction to data mining

Data mining is an analytic process that extracts a large amount of data from different sources, which could be stored in separate locations and transforms them into a comprehensible structure with the development of information communication technology (Wang *et al*., 2018). It is considered as an analysis step of "knowledge discovery in databases (KDD)" process and identification of valuable knowledge provided by database system. The applications of data mining are wide and include manufacturing, retailing, health science etc. Common data mining steps are data collection, data pre-processing, data transformation, data reduction, feature and knowledges extraction, and visualization (Ghandi *et al*., 2018). The importance of data mining is mainly motivated by the profitability of data collection and processing, such as to discover useful patterns significant to prominent task and output result (Gullo, 2015). Generally, manufacturing industries use data mining tool to enhance the intelligence and efficiency of design, production, service process and supply chain. Through data mining, data information could generate different calculated alternatives to the problem faced. For examples, classification in data mining can help to predict a certain outcomes based on the input given (Wang *et al*., 2018). This involves computing the similarity between data sources and then separating all data sources into classes based on their similarities.

2.2 Machine performance and computer-integrated manufacturing

Computer-integrated Manufacturing (CIM) refers to the use of computer-controlled machines and automation systems such as Computer Numerical Control (CNC) machine aided with new capabilities of monitoring and controlling machine tools and collaboration of sensors in manufacturing products (Mourtzis, 2018). CIM combines different application of technologies like computer-aided design (CAD), computer-aided manufacturing (CAM), robotics and manufacturing resources planning to optimize the performance of the machine by automating repetitive steps(Chen, 2017). All of these components call for data storage, data processing algorithms, manipulation and retrieval mechanism, real-time sensors for sensing the current state and modify processes and etc.

The ultimate goal of CIM is to provide an error-free and automated manufacturing process. In this sense, CIM is known as "workerless" production where the industrial 4.0 transforms regular machines into self-learning and self-aware machines to offer great flexibility and quality of the product produced. The fully integrated and automated production flow significantly changes the traditional production relationship among suppliers, manufacturers, and customers as well as the interaction between workers and machines (Saurabh *et al*., 2018). Digitization and intelligentization of manufacturing process show the rapid advancement and application of CIM in the industrial 4.0 to control over the entire value chain and the life cycle of products.

One of the applications of CIM that is a hybrid system of robots and humans. For instance, instrumentation of machine tools such as computer numerical control machines, three-dimensional (3D) printers, and robots with sensors to collect the machine data for the purpose of process monitoring, control and management. Advanced robotics and programming had been used to enhance the processing flexibility of a machine which characterized by robotic automation. With sophisticated automation of robots, the speed, quality and the processing flexibility at specific machine is able to be control and thus help to accommodate product variety and lot size fluctuation with reasonable responsiveness and precision. However, sensors, data sharing and networking of machines put a risk of unprecedented cybersecurity issue for the industrial companies (Wang, 2018).

Mourtzis *et al.*(2018) presented a communication framework of Open Platform Communication-Unified Architecture (OPC-UA) to model milling and lathe CNC machine tools. The status of the machine can be monitored in real-time through well-adopted communication interfaces. The OPC-UA standard has the potential to support the integration of equipment of different vendors and architecture towards realising the vision of Industry 4.0. The first step of digitalisation is to design a model using the UML (Unified Modelling Language) and then to identify a communication standard exposing the model to cyber world. For this purpose, the model is transformed to the OPC-UA standard from UML. Data is captured from the operation of manufacturing system. To do so, sensor and wireless communications need to be installed. Therefore, information model can encapsulate the semantics related to physical object and corresponding relationship. During runtime, semantic heterogeneous data models also can be integrated into common domain model with appropriate mapping in the system in order to achieve unification. Arezoo *et al.* (2000) developed Expert Computer Aided Tool Selection System (EXACTS) to optimise the performance of simple turning operations in an automated manufacturing systems. The knowledge base of EXACTS comprises of data files about material properties, algorithmic knowledge of tool holder capabilities and tool selection guidelines. The guidelines held in rule-based form, can be changed and adapted to different machining operations.

Accorsi *et al.* (2017) introduced data mining and machine learning (ML) models for condition-based maintenance of a complex high-speed packaging machine. In grouping of machine data, clustering techniques, association rules and classification models have been exploited to detect the anomalies in the machine activity. Therefore, the maintenance in complex production systems could be predicted and improved.

Simeone *et al.* (2018) proposed an intelligent cloud manufacturing platform for increase resources of metal sheet efficiency in a manufacturing network through dynamic sharing of manufacturing services. Through cloud, users can get universal access to smart machines, production systems, sensor systems and intelligent computation. A graphic user interface (GUI) is provided to enables user to create a personal profile and entry customer and supplier instances data into the cloud platform. The data collected are the manufacturing services requests from customers and the offers from suppliers and cloud platform will utilizes a Genetic Algorithm (GA) based optimization system to find the best dimensions of metal sheets according to the demands. Hence, the utilization of sheet metal cutting services rate by CNC machine is improved based on the cloud framework.

2.3 Data mining on machine performance

Kromer *et al.* (2013) have applied an evolutionary-fuzzy rule based on genetic programming (GP) for making a quality prediction on a steel plate. The attributes include chemical properties of the raw material, density and temperature at several processing stages are collected during the production cycle. Classification on steel plate either flawless or defective was carry on according to the parameter of quality. Then, a fuzzy rule evolved by GP was generated which described the class of defective products in term of product features. Valuable information about the product features was provided based on the symbolic nature of fuzzy rules which helps to indicate the defectiveness of products. The challenge of this machine learning algorithm is to extract out the hidden data into complex data sets based on their similar characteristics and thus used for the optimised process.

According to Denkena *et al.* (2014), data mining on the machine helps to recognize unknown patterns within data. Different algorithms are used and applied to analyze manufacturing data. A cluster analysis using the k-means algorithm was applied to predict the capability of a specific process with one specific set of parameters. Afterward, a k-nearest neighbour algorithm identified the best process setting. Optimum operation regarding productivity and quality factors can be selected and also to identify alternative processes in case of missing resources or rescheduling in case of disturbance.

Du *et al.* (2015) proposed a statistical-analysis (SA)-oriented data mining technique of Bayesian approach to estimate the process control parameters on valve shell which machined by turning process. Based on the developed linear model, a two-step Bayesian method was proposed to estimate the process mean and variance. The variation of product characteristics and tool error would be generated in the

MATLAB based on the variation of fixture and measurements of valve shell loaded into programme. Therefore, the estimate process control parameter of the product was obtained and used in machining with the application of the proposed Bayesian method.

Packianather *et al.* (2017) applied data mining techniques such as data exploration, data segmentation, association rules, and time series to real data on the database of brick products. In data exploration, the main characteristics of products need to understand and decide the best approach to extract meaningful information. Clustering approach was used for grouping similar data points together through k-means algorithms and hierarchical clustering. Association rule was performed in numerous of ways and providing an understanding of what items customers are purchasing together. Time series forecasting was performed by exponential smoothing to observe the upward, downward, cyclic and seasonal trends for brick products. Therefore, a strategic decision making on expanding the production rate of brick at seasonal periods would be made to fulfill the demand of the market.

A framework for knowledge-driven optimization was proposed by Bandaru *et al. (*2017) which involves both online and offline elements of knowledge discovery in order to provide deeper insights about the problem to the decision maker. Descriptive statistics is the first step to obtain basic quantitative information about the data in the form of numbers by measures of central tendency, variability, shape distribution and correlation. Then, further classification as graphical, clustering-based and manifold learning methods were applied for finding the hidden structures in multivariate datasets. Lastly, supervised and unsupervised learning techniques in machine learning were used to generate knowledge in an explicit form.

Vazan *et al. (*2017) used data mining techniques of neural network (NN) to predict the manufacturing process behaviour according to the production data provided by an automotive industry component supplier. The proposed simulation model of the real manufacturing process was designed to obtain the data necessary for process control. Based on the mathematical statistical scoring method, NN method was chosen with the best total score of eight (the most successful)over other data mining techniques of random forest, boosting tree, k-nearest neighbour, multiple regression and support vector machine. In the NN method, the number of neurons in input, hidden and output layers was automatically set to the specific task, which defined by the number of input parameters, output values and the input data set. After tested, three out of the five data parameters of automotive components were classified as true which means simultaneous fulfilment of all characteristic was predicted. It is also proved that the prediction made by NN is the most accurate solution with the smallest residual error which has verified by simulation model. However, the boundaries between descriptive and predictive of data are not clear when extracting knowledge from database.

Accorsi *et al.* (2017) applied data mining and machine learning models and methods for condition-based maintenance on a complex high-speed packaging machine. In the first stage, the clustering technique was used to group machine data for detecting similar behaviours in an unsupervised way. Association rule was exploited in the second stage to discover recurring machine behaviour preceding the failure events. Finally, classification was used to exploit data and built an actual behaviour model by the classifier of decision tree, random forest and neural networks. Therefore, a classification model is built and able to predict faults of the machine and possibly identified the root causes. By comparing these three classifiers, random forest is more accurate than the decision tree and neural network.

Addona *et al.* (2017) applied feature extraction, and Artificial Neural Networks (ANNs) to characterize the condition of the wheel during grinding operations. Through a sensor monitoring system, the Acoustic Emission (AE) signal was acquired and derived a statistic from the signal. By feature extraction, the extraction of relevant signal was selected to define the intrinsic characteristics of the signal. Then, the signal feature will combine with the working parameters of grinding machine such as feed rate, depth of cut and an internal diameter of workpieces before feed into the ANNs decision-making system. Two of the ANNs model was built and relies on MATLAB function of *cascadeforwardnet* (CFN) and *newcf* (NCF). Both of the functions have been trained for making a decision on wheel condition and the need for dressing during grinding operation. ANNs was able to predict the condition of the wheel either sharp or dull in 100 % based on the condition given. Hence, the system could be employed in real time to predict the tool-wear in milling or turning operations and predict the dressing wear of grinding operations with a view of continuous improvement of machine performance.

Chaiwat and Koichi (2018) have proposed J48 classifier and CAR (Class Association Rule) to generate a predictive model and rule on the manufacturing of ceramic. The data was collected based on the properties and design of mug and presented in a relational model. WEKA (Waikato Environment for Knowledge Analysis) software was used for testing and validated the datasets prepared. Decision tree by J48 algorithm was used to determine the most important factor among product attributes while the CAR was used to understand the relationship between product attributes and customer emotions from the strongest to the weakest relationship. The predictive rule from agree to the strongest agree level also can be determined based on the dataset prepared. Hence, a predictive model of mug property was determined by J48 algorithms and CAR and thus used to evaluate customer needs and design alternatives. However, it takes a long time to identify the properties of a product into each category in classification.

Decision support system architecture (DSSA) integrated data mining techniques such as algorithm of decision tree and random forest to detect different types of defects and anomalies of equipment and to predict maintenance actions and evaluate the health status of components. The prediction of remaining useful life of the machine or components was affected by numerous sources of machine process and operating conditions. Therefore, it is important to interpret these sources to facilitate significant decision support (Ghandi*et al*., 2018).

Table 2.1: Summary table of data mining techniques/ applications in machine performance

Authors	Title	Purpose	Data Mining Techniques/ Applications	Roles of the Data Mining Techniques / Applications
Kromer *et al.* (2013)	Towards new directions of data mining by evolutionary fuzzy rules and symbolic regression	To classify data by hybrid evolutionary-fuzzy rule based on genetic programming and Boolean queries	Classification, Fuzzy Rules evolved by genetic programming	The defectiveness of products is indicated by fuzzy rules.
Denkena *et al.* (2014)	Data Mining Approach for knowledge-based process planning	Optimum operation regarding productivity and quality factors can be selected and also to identify alternative processes in case of missing resources or rescheduling in case of disturbance.	Cluster analysis of k-means algorithm, k-nearest neighbour algorithm	-Cluster analysis using k-means algorithm is used for predict the capability of a specific process with one specific set of parameters. -K-nearest neighbour algorithm is used for identified the best process setting.

Du *et al.* (2015)	Engineering model-based Bayesian monitoring of ramp- up phase of multistage manufacturing process	To estimate the process control parameters on valve shell which machined by turning process.	Bayesian approach	-Two-step Bayesian method is proposed to estimate the process mean and variance.
Packianather *et al*. (2017)	Data mining techniques applied to a manufacturing SME	To observe the upward, downward, cyclic and seasonal trends for brick products. A strategic decision making on expand the production rate of brick at seasonal periods could be made.	Data exploration, Data clustering, k-means algorithms, Association rules, and Exponential smoothing	-Data exploration is used for understand the main characteristics of products -Clustering approach is used for grouping similar data points together through k-means algorithms and hierarchical clustering. -Association rule is performed in a numerous of ways to provide an understanding of what items customers are purchasing together. -Time series forecasting is performed by exponential smoothing to observe the upward, downward, cyclic and seasonal trends for brick products.

Bandaru *et al. (*2017)	Data mining methods for knowledge discovery in multi-objective optimization: Part A- A survey	To provide deeper insight about the problem to the decision maker.	Descriptive statistics, Clustering-based, Manifold learning methods and Machine learning	-Descriptive statistics is used for obtaining basic quantitative information about the data in the form of numbers by measures of central tendency, variability, shape distribution and correlation. -Classification as graphical, clustering-based and manifold learning methods is applied for finding the hidden structures in multivariate datasets.
Vazan *et al. (*2017)	Using data mining methods for manufacturing process control	To predict the manufacturing process behaviour according to the production data.	Neural network method (NN)	NN method is used to classify the new data set and make prediction analysis based on the data collected.

Accorsi *et al.* (2017)	Data mining and machine learning for condition-based maintenance	To predict the actual behaviour of machine and the probability of failure event of the system	Clustering, Association rule, Decision tree, Random forest and Neural network	-Clustering techniques is used to group machine data for detecting similar behaviours in an unsupervised way. -Association rule is used for discover recurring machine behaviour preceding the failure events. -Classification approach is used to predict faults of machine and possibly identify the causes that caused malfunction of machine.
Addona *et al.* (2017)	Prediction of dressing in grinding operation via neural networks	To predict the tool-wear in milling or turning operations and predict the dressing wear of grinding operations.	Feature extraction and Artificial Neural Networks (ANNs)	-Feature extraction is used for define the intrinsic characteristics of signal. -ANNs is used for predict the condition of wheel either sharp or dull.

Chaiwat and Koichi (2018)	Application of Kansei engineering and data mining in the Thai ceramic manufacturing	To generate a predictive model with desired characteristics from customer.	J48 classifier and CAR (Class Association Rule)	-Decision tree by J48 algorithm is used to determine the most important factor among product attributes -CAR is used to understand the relationship between product attributes and customer emotions from strongest to the weakest relationship.
Ghandi *et al.* (2018)	Toward data mining based decision support in manufacturing maintenance	1. To detect the different types of defects and anomalies of an equipment 2. To predict maintenance actions and evaluate health status of components	Decision tree and Random forest	-Decision tree is used for detect the anomalies of equipment -Random forest is used for predict maintenance action and evaluate health status of components.

Chapter 3 RESEARCH METHODOLOGY

3.1 Overview

This chapter focuses on fundamental data mining concepts and techniques to discover patterns from a large amount of data. In section 3.1, an objective will be defined after studying the condition of the factory from a manufacturing industry perspective. Section 3.2 introduces how the data was collected. Then, data pre-processing is to be carried out, which include the steps such as data preparation (Section 3.3), data selection (Section 3.4), data cleaning (Section 3.5), data transformation (Section 3.6), data mining and the tool used (Section 3.7). Finally, Section 3.8 reveals the discovered functional patterns from database which could be used in decision making.

Start from business understanding, information understanding is obtained from a section of production running with machines in a medical tool manufacturing plant. With a defined objective, an appropriate data mining problem is determined. Data collection will be done in a simulation study which created in WITNESS HORIZON V21. The result of the simulation will document in Microsoft Access. Then, the process proceeds to the data pre-processing stage. In data preparation, a prepared dataset that matches to the data needs are stored in "Comma-separated value" (CSV) format where each row is representing an instance and each column is representing an attribute of the data (Diego *et al*., 2019). Incomplete data or missing values will be recovered in this phase (Zhang *et al.,* 2003). In data selection, the relevant data type and source will determine and they would present in numerical figures or text form. From the dataset, filtering is applied to eliminate the duplicate data. In data cleaning, the integrity of data is checked by smoothing and eliminating the noise which presented in data (Gandhi *et al.,* 2018). It is performed in linear projection widget by Orange. In data transformation, the data are transformed into an appropriate standard structure form which allows the data mining tool to interpret. It is performed in the Microsoft Excel. After that, data mining is performed in Orange software, one of the open source data mining tool. According to the Packianather (2017), the process of data mining is known as the process of discovery of hidden trends, patterns, relationships and insights through a large number of datasets where can predict the future occurrences. Lastly, the pattern evaluation is performed to identify an interesting pattern which represents knowledge and correlation of data based on the interaction of interest measure with the data mining modules.

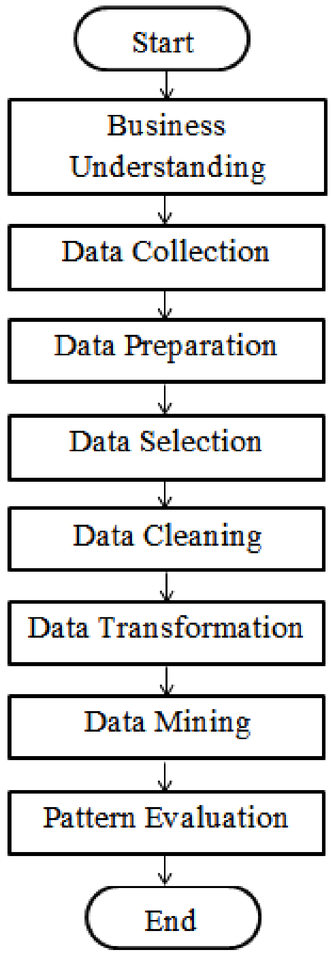


Figure 3.1: Flow chart of data mining process

3.2 Business Understanding

The initial step focuses on understanding the objectives and demands from a manufacturing industry perspective. Information would be obtained from a medical tool manufacturing plant. A section of the production running intensively using machines or equipment would be studied. An appropriate data mining problem definition could be determined through exploration, classification, prediction and modelling. From the finding, a data mining goal is created. Each sub-question would be addressed by using a specific data mining tool and filtered datasets.

3.3 Data Collection

Collecting a sufficiently large dataset related to breakdown and maintenance from the manufacturing industry could be hard. A manufacturing industry is commonly stable as machines are built for heavy usage. Under such circumstance, certain breakdowns may not happen regularly. Data collection therefore requires significantly long period of time.

Under such situation, a computational simulation was used as an alternative to generate data. The simulation model of manufacturing system would be built in WITNESS Horizon V21. WITNESS is a system modelling tool of which system is created in a graphical form on the computer screen with the assistance of elements. The elements of Witness include variables, functions, pie charts, time series, and shift (Runciman, 1997). The accuracy of the models is controlled with a large number of predefined and user defined functions. There are the input and output rules controlling the way of parts (Spedding and Sun, 1999). In simulation study, this process is known as model translation which covered the model into a programming language. After that, the verification and validation can be done to verify the accuracy of system to the real system. Lastly, result of the simulation is documented and analysed. The data generated through simulation would be stored in database developed in Microsoft Access. Microsoft Access database is created based on the Access Jet Database Engine.

3.4 Data Preparation

Data preparation is the process of collecting, cleaning and consolidating data into a single file or data table. It is the fundamental stage for data analysis as real-world data may be incomplete, noisy and inconsistent. Incomplete data or missing values would be recovered also in this phase (Zhang *et al.*,2003). When the data is properly prepared, a dataset that match to data needs would be generated and stored in "Comma-separated value" (CSV) format where each row is representing an instance and each column is representing an attribute of the data (Diego *et al.,* 2019).

3.5 Data Selection

The purpose of data selection is to determine the data type and source from database. The data type and resource can be represented in a variety of ways such as numerical figures or text form. Filtering would be applied to select relevant data and eliminate the duplicate data in the dataset.

3.6 Data Cleaning

Data cleaning is the process that checks the integrity of data and consistency, smoothing and eliminating the noise in data (Gandhi *et al.*, 2018). Data cleaning would be performed in Orange by linear projection widget. In linear projection widget, the projection of circular placement is selected and used for detect the outliers of data. An outlier widget would be used if any outlier or anomalies of data is found in dataset. Then, the data table would be connected to the widget of outliers in order to identify the individual error and removed it from dataset.

3.7 Data Transformation

In data transformation, the data are transformed into an appropriate form readable by data mining tool. The strategies for data transformation include smoothing, attribute construction, aggregation, normalization and discretization. Therefore, new attributes would be obtained after reconstructing the current structure of attributes from the given data. The ultimate goal of data transformation is to form a standard data structure to make data mining application workable. These strategies would largely be performed in the Microsoft Excel using formulated functions for data classification.

3.8 Data Mining

Data Mining process is the discovery of hidden trends, patterns, relationships and insights through a large number of datasets where can predict the future occurrences (Packianather, 2017). Typically, data mining involves in data manipulation and classification by applying the statistical and mathematical concepts. In this research, Orange is used to explore and analysis data.

Orange is one of the open source tools available to perform data mining especially for novice and expert. It is a component-based platform for data visualization, machine learning, data mining and data analysis. A front-end visual programming or Python scripting is performed for rapid qualitative data analysis with simple visualization. A set of components known as widgets in Orange is work for data pre-processing, feature scoring and filtering, modelling, exploration and model evaluation (Bansal, 2014). The data and visualization widgets are widely-used for data exploratory in this research. The basic statistical tools used for data inspection are file, selected column and data table. They are used for loaded selected data before starting processing data. Linear projection and outliers are used for detect anomalies of data from dataset. The tool used for data visualization is distribution graph, scatter plot and linear regression. From the distribution graph, it helps to listing all the attributes values and how often they are occurred. Scatter plot is used to determine relationship between two variables and how much one variable is affected by another. The relationship between two variables also known as correlation which is similar to linear regression, predicting the value of dependent variables based on the value of independent variables.

3.9 Pattern Evaluation

The pattern evaluation is performed to identify an interesting pattern which represents knowledge and correlation of data based on the interaction of interest measure with the data mining modules. In the first phase of evaluation, the model result will be evaluated in the context of business goals and requirements. High-level graphical user interfaces and visualization tools would be integrated with existing domain-specific data and information systems to searching for patterns, interpreting and visualizing discovered patterns. At the end, a good decision from the discovered knowledge could be made based on the needs and goals of business.

Chapter 4 RESULT AND DISCUSSION

4.1 Company background

B. Braun Corporation was originated in 1839 where it started as a pharmacy in Melsungen, to sell medical herbs by mail to customers in Germany. It is one of the most successful global players with more than 170 years of innovative history in the health care industries. In 14th September 1972, B. Braun founded their production area in Malaysia which is located in Bayan Lepas Free Industrial Zone, Penang. B.Braun Medical Industries (BMI) in Penang is the subsidiary of B. Braun and it is also one of the largest manufacturing sites in Asia Pacific Region other than Europe. The operational business of B.Braun is led from its headquarters in Melsungen, Germany. Now, B.Braun supplies globally to the healthcare market with products for intensive medicine, anesthesia, extra corporeal blood treatment, cardiology and surgery, as well as services for hospitals, medical industries, general practitioners and the homecare sector.

There are a total of 5 plants in BMI Penang plants which comprised of AESCULAP, Medical Plant, Pharmaceutical Plant, Needle Plant and Plastic Technology which are focused on manufacturing products of IV Catheters, Hypodermic Needles, Special Needles, Elastomeric Pumps, Pharmaceutical Solutions and Surgical Instruments. In the Aesculap plant of B.Braun, the production area of CNC machines is located at the third floor of building together with 13 units of CNC Milling machines located. The model of CNC machine is known as MILLTAP 700 which developed by GILDEMEISTER AG (DMG). The MILLTAP 700 is able to achieve absolute top values in all processes with its combination of impressive work capacity and chipping performances. The adapted Siemens 840D solution line with Sinamics 120 compact inverter and 10” monitor systems provides high dynamic contour accuracy and efficient program sequences. Furthermore, PROGRESS line system can also be equipped in machine which helps to displays the remaining time and quantity of complete machining order as status control.

4.2 Business Understanding

The main goal of this study is to look into the performances of CNC machines in medical tool manufacturing plant. It focuses on identifying the type of breakdown and how often they are happened when the machine is operated and the effect to production after maintenance is applied into schedule. Based on the business goal, the study starts with problems faced on the production of medical products. Several questions in the stage of macro-level and micro-level is generated and used for investigate the breakdown and performances of the machine. In macro-level, CNC machine of CK6166-11 was selected for analysis the data while in micro-level, three specific CNC machines which are CK6166-13, CK6166-23 and CK6166-33 have been selected for investigate the relationship between product and breakdown. Another five CNC machines which are CK6166-11, CK6166-12, CK6166-13, CK6166-14 and CK6166-15 were selected for study the relationship between breakdown type and maintenance.

4.2.1 Macro-level

In macro level, it considers the full panorama of business with a defined objective. Less specific and philosophical analysis is constructed in this phase.

1. Breakdown
   - i. Type and frequency of breakdown
   - ii. Duration of repair time based on the breakdown type
   - iii. Cost spent on the breakdown in three year periods
   - iv. Comparison on type and frequency of breakdown before maintenance and after maintenance
   - v. Duration of repair time is influenced by maintenance schedule
   - vi. Comparison on cost spent of before maintenance and after maintenance
2. Products
   - i. Type and Frequency of breakdown based on the product
   - ii. Effect on product and breakdown based on the product characteristic

4.2.2 Micro-level

At the micro level, the investigation focuses on small details and individual perspective, e.g. how certain attributes are affected by the others and the interaction between them.

1. Breakdown type based on the product
   i. Frequency of breakdown type against product
2. Breakdown type based on the maintenance
   i. The maintenance schedule is affecting the frequency of breakdown
   ii. The breakeven of breakdown over maintenance schedule

4.3 Data Collection

In the medical product plant, there is a production area for Computer Numerical Controlled (CNC) machines to produce metal parts in lines. Each CNC machine is controlled and monitored by a single operator. The main components of automated CNC machines include operating panel, magazine tool, coolant lubricant system, motor, spindle, chip conveyor and others. A pre-programmed file like G-code program would be inserted into the machine before start processing. Each event happens in the machine would be recorded into log file in term of process and time. The status of machine either run, stop or idle will be signalled by the light towel of machine. The time for processing a metal part is calculated and recorded at the screen of control panel. Therefore, the condition of machine and the daily output of machine could be known from the system. A similar production condition is constructed in Witness software. A real-time setting where program are run in Witness enables data generated in simulation.

A simulation model was constructed using WITNESS Horizon V2.1 discrete-event simulation software. A hypothetical case is built. As shown in Figure 4.1, the model consists of a single machine (Mc) a single buffer (B1) and five types of products (P1 to P5). Each product inherent three attributes which believed to affect breakdowns. Information regarding ten types of common breakdown and five scheduled maintenance policies happened to CNC machines in the company were identified. As shown in Figure 4.2, they were embedded individually into the simulation. Specifically, mean time between failure (MTBF) and mean time to repair

(MTTR) were defined in Truncated Normal distributions for each breakdown. A truncated normal distribution requires five numerical variables: mean, standard deviation, min, max and random seed. The mean and the standard deviation are used to characterize normal distribution; the min and the max delineate the lower bound and upper bound of the MTBF generated. The random seed is a number meant to initialize pseudorandom number generator. Default values have been given to these five numerical variables at the beginning of the program. Each breakdown type has a distinctive set of default values. The machine is broken down only during processing (busy time). Only one breakdown type could take place at a time. Repair time is measured as soon as the breakdown is signalled. Production halts during repair and is resumed immediately after repair. For model simplification, labour is considered always available.

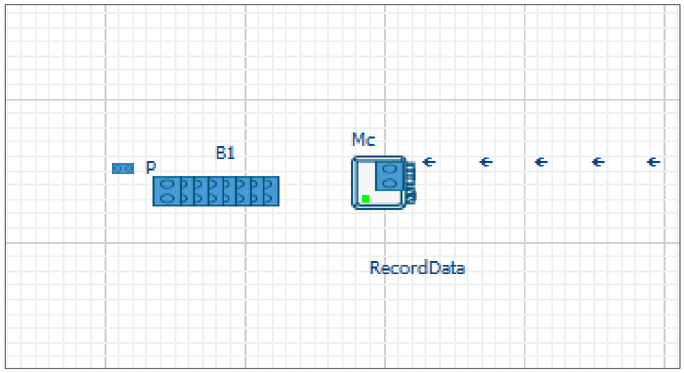
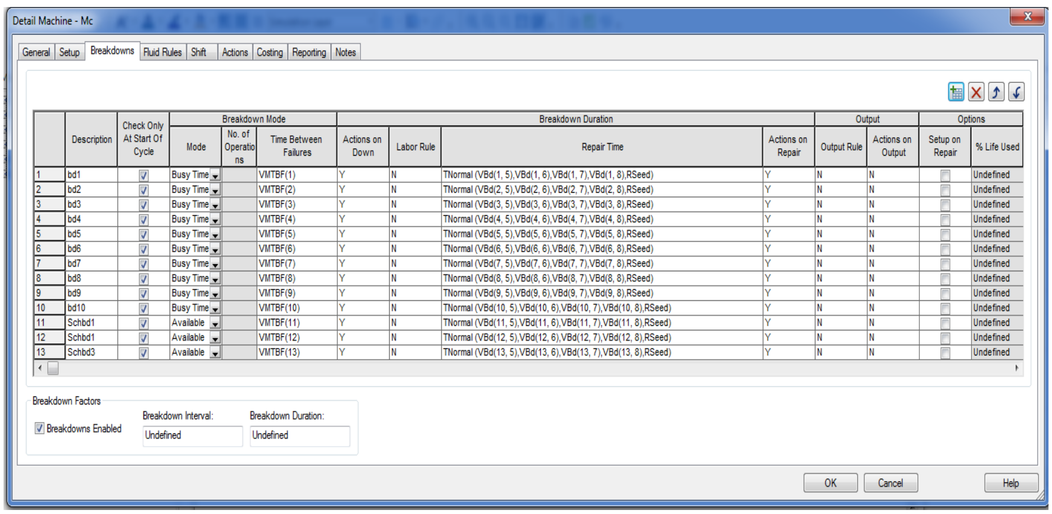


Figure 4.1: The screenshot of the simulation model interface



	Description	Check Only At Start Of Cycle	Breakdown Mode			Breakdown Duration				Output		Options	
			Mode	No. of Operations	Time Between Failures	Actions on Down	Labor Rule	Repair Time	Actions on Repair	Output Rule	Actions on Output	Setup on Repair	% Life Used
1	bd1	✓	Busy Time		VMTBF(1)	Y	N	TNormal (VBd(1, 5),VBd(1, 6),VBd(1, 7),VBd(1, 8),RSeed)	Y	N	N		Undefined
2	bd2	✓	Busy Time		VMTBF(2)	Y	N	TNormal (VBd(2, 5),VBd(2, 6),VBd(2, 7),VBd(2, 8),RSeed)	Y	N	N		Undefined
3	bd3	✓	Busy Time		VMTBF(3)	Y	N	TNormal (VBd(3, 5),VBd(3, 6),VBd(3, 7),VBd(3, 8),RSeed)	Y	N	N		Undefined
4	bd4	✓	Busy Time		VMTBF(4)	Y	N	TNormal (VBd(4, 5),VBd(4, 6),VBd(4, 7),VBd(4, 8),RSeed)	Y	N	N		Undefined
5	bd5	✓	Busy Time		VMTBF(5)	Y	N	TNormal (VBd(5, 5),VBd(5, 6),VBd(5, 7),VBd(5, 8),RSeed)	Y	N	N		Undefined
6	bd6	✓	Busy Time		VMTBF(6)	Y	N	TNormal (VBd(6, 5),VBd(6, 6),VBd(6, 7),VBd(6, 8),RSeed)	Y	N	N		Undefined
7	bd7	✓	Busy Time		VMTBF(7)	Y	N	TNormal (VBd(7, 5),VBd(7, 6),VBd(7, 7),VBd(7, 8),RSeed)	Y	N	N		Undefined
8	bd8	✓	Busy Time		VMTBF(8)	Y	N	TNormal (VBd(8, 5),VBd(8, 6),VBd(8, 7),VBd(8, 8),RSeed)	Y	N	N		Undefined
9	bd9	✓	Busy Time		VMTBF(9)	Y	N	TNormal (VBd(9, 5),VBd(9, 6),VBd(9, 7),VBd(9, 8),RSeed)	Y	N	N		Undefined
10	bd10	✓	Busy Time		VMTBF(10)	Y	N	TNormal (VBd(10, 5),VBd(10, 6),VBd(10, 7),VBd(10, 8),RSeed)	Y	N	N		Undefined
11	Schbd1	✓	Available		VMTBF(11)	Y	N	TNormal (VBd(11, 5),VBd(11, 6),VBd(11, 7),VBd(11, 8),RSeed)	Y	N	N		Undefined
12	Schbd1	✓	Available		VMTBF(12)	Y	N	TNormal (VBd(12, 5),VBd(12, 6),VBd(12, 7),VBd(12, 8),RSeed)	Y	N	N		Undefined
13	Schbd3	✓	Available		VMTBF(13)	Y	N	TNormal (VBd(13, 5),VBd(13, 6),VBd(13, 7),VBd(13, 8),RSeed)	Y	N	N		Undefined

Figure 4.2: Ten types of machine breakdowns (bd) and three types of schedule maintenance (Schbd)

A simple randomized algorithmic mechanism (as shown in Figure 4.3) is built-in to variously and dynamically influence the breakdown rate in accordance to the product attributes and scheduled maintenance practices. For the former, when the value of product attribute exceeds six sigma of its value range, the MTBF of certain breakdown types would be reduced by a small amount based on probability. This in effect increases the frequencies of the relevant breakdown types. However, the MTBF would be reverted to default after repair. For the latter, a scheduled maintenance would increase the MTBF of selected breakdown types, making the breakdown less likely to occur. Aforementioned, some randomness is embedded in the algorithm, the effect may be subtle and hence less predictable. The information is not revealed to operation analyst for this research. The operation analyst's data mining task therefore is akin to an investigator to find clues from the data presented.