University of Louisville
ThinkIR: The University of Louisville's Institutional Repository

Electronic Theses and Dissertations

5-2018

Maintainability analysis of mining trucks with data analytics.

Abdulgani Kahraman
University of Louisville

Recommended Citation

Kahraman, Abdulgani, "Maintainability analysis of mining trucks with data analytics." (2018). *Electronic Theses and Dissertations.* Paper 2932.
https://doi.org/10.18297/etd/2932

MAINTAINABILITY ANALYSIS OF MINING TRUCKS WITH DATA ANALYTICS

By

Abdulgani Kahraman
B.S., Sakarya University, 2011

A Thesis
Submitted to the Faculty of the
J.B. Speed School of Engineering
University of Louisville
In Partial Fulfillment of the Requirements
for the Degree of

Master of Science in Computer Science

Department of Computer Engineering and Computer Science
University of Louisville
Louisville, Kentucky

May 2018

MAINTAINABILITY ANALYSIS OF MINING TRUCKS WITH DATA ANALYTICS

By

Abdulgani Kahraman
B.S., Sakarya University, 2011

A Thesis Approved on

April 24, 2018

by the following Thesis Committee:

____
Dr. Mehmed Kantardzic, Thesis Director

____
Dr. Adel Elmaghraby

____
Dr. James E. Lewis

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Mehmed Kantardzic for his great contributions and support in this study.

I would also like to thank my brother Dr. Mustafa Kahraman for his priceless support to this research study.

I take this opportunity to thank Dr. Adel Elmaghraby and Dr. James Lewis for serving on my committee, reading this thesis and providing comments during defense.

Finally, I am grateful to my family for their love and support, especially my mother.

ABSTRACT

MAINTAINABILITY ANALYSIS OF MINING TRUCKS WITH DATA ANALYTICS

Abdulgani Kahraman

April 24, 2018

The mining industry is one of the biggest industries in need of a large budget, and current changes in global economic challenges force the industry to reduce its production expenses. One of the biggest expenditures is maintenance. Thanks to the data mining techniques, available historical records of machines' alarms and signals might be used to predict machine failures. This is crucial because repairing machines after failures is not as efficient as utilizing predictive maintenance.

In this case study, the reasons for failures seem to be related to the order of signals or alarms, called events, which come from trucks. The trucks ran twenty-four hours a day, seven days a week, and drivers worked twelve-hour shifts during a nine-month period. Sequential pattern mining was implemented as a data mining methodology to discover which failures might be connected to groups of events, and SQL was used for analyzing the data.

According to results, there are several sequential patterns in alarms and signals before machine breakdowns occur. Furthermore, the results are shown differently depending on shifts' sizes. Before breakdowns occur in the last five shifts a hundred percent detection rates are observed. However, in the last three shifts it is observed less than a hundred-percentage detection rate.

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

In this study, first, the basic mining techniques' information related to the mining industry is mentioned. Second, recent technological developments and big data developments in the mining industry are presented. Third, the maintenance costs for big vehicles and how these maintenance expenses might be reduced, thanks to data mining, is considered. Lastly, as a case study, sequential pattern mining is used as a methodology on the dataset for predictive maintenance and results are shared.

1.1 Recent Technological Developments in Mining Industry

Since early civilization, individuals have utilized mining strategies to extract minerals from the soil. The citizens of ancient civilizations were all interested in mining. In the past, mining was slow-going and unsafe. As time has progressed, society has created more secure and exact strategies for finding and revealing substances found in the soil.

In the beginning, diggers employed primitive devices for burrowing. Mining shafts were burrowed out by hand, and the entire process was exceptionally long. Inevitably, individuals started using fire to clear burrows and reach more prominent profundities at a quicker rate. Amid the 1600s, diggers began utilizing explosives to break up expansive rocks. Motorized mining apparatuses, such as drills, would not be invented for a few more decades, and it was not until the Industrial Revolution began in the 1700s that mineworkers started improving the explosives they operated and created more progressive mining gear, such as drills, lifts and steam-powered pumps (www.generalkinematics.com, 2015**).**

In today's technologically-advanced society, mining strategies are continuously progressing. For instance, improving surface mining procedures, diggers are presently able to extricate over 85 percent of minerals and 98 percent of metallic minerals without digging a shaft or imperiling the lives of workers (www.generalkinematics.com, 2015**).** Newly-developed machines utilized for grinding and crushing can extricate minerals from the soil with less energy than ever before.

Miners still use several techniques, such as explosives, trucks, drills and bulldozers, particularly if they must dig deep into the soil. In any case, innovations have permitted mineworkers in uncovering minerals with more exactness and less harm to the encompassing environment. More proficient apparatuses can be utilized to decrease energy consumption and increase the sum of minerals or metals gathered from the shaft.

Mining has made the world more modern compared to the past, but the threats of mining have resulted in the deaths of numerous laborers. As innovation progresses, mining procedures have indeed become more precise and productive. In the future, with technological developments, it is possible to mine for materials with fully automatic machines thanks to the Industry 4.0 system.

Figure 1 shows several mining techniques which are separated mainly into two categories: underground mining and surface mining. While underground mining contains drift, slope, and shaft mining, surface-mining methods involve area, contour, mountaintop removal, and auger which are in Figure 1 (www.uky.edu, n.d.). Moreover, there are different kinds of machines in the mining fields, and these machines are expensive vehicles and keeping these vehicles in working conditions is vital for companies. Nowadays, thanks to the monitoring systems, companies can follow and record these machines' statuses in every moment, and it provides opportunities to take precautions before machine breakdowns occur.

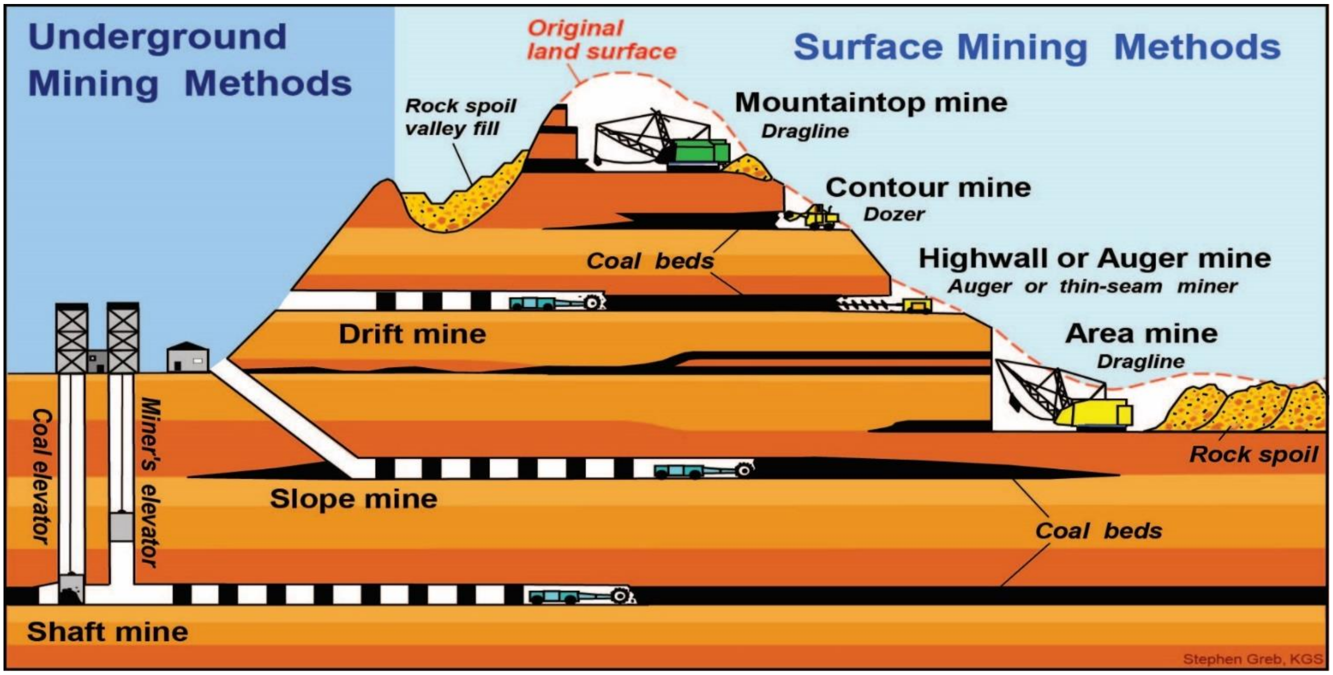


Figure 1 - Some recent mining techniques (www.uky.edu, n.d.)

1.1.1 Vehicle Health Monitoring System in Mining Industry

With technological developments companies try to record and follow their vehicles. Big vehicle production companies have their own monitoring systems for watching and taking control of their big vehicles, and these systems present a lot of advantages for mining companies (Viger, 2017). The Monitoring System is a recent technology for big and expensive vehicles, which are consistently worked for numerous hours. When the vehicles break down, it takes serious costs to repair them. Furthermore, since repairing these big machines take a significant amount of time, this completely influences the machine's availability. Hence, mining machines, especially large ones, are required to reduce the number of failures and enable operations without intrusion (Murakami, Saigo, Ohkura, Okawa & Taninaga, 2002). In order for machines to continuously work, it is essential to identify any problems and changes in status by physical examination early on that will signal the maintenance staff to take reasonable measures without waiting for total failures of valuable equipment.

Vehicle Health Monitoring System (VHMS) is one example of monitoring systems available to mining companies, and it is suitable to integrate with existing systems. With Internet of Things (IoT) every company has an opportunity to collect data, and they can analyze this data to set more efficient work schedules. Particularly, it is extremely crucial for companies which have big vehicles, because these kinds of vehicles need expensive maintenance. As a result of this, analyzing their past records and taking several precautions are vital not only for companies but also for employees because these precautions can provide safer workplaces. Currently, many companies have their own databases which were created by several monitoring systems and IoT.

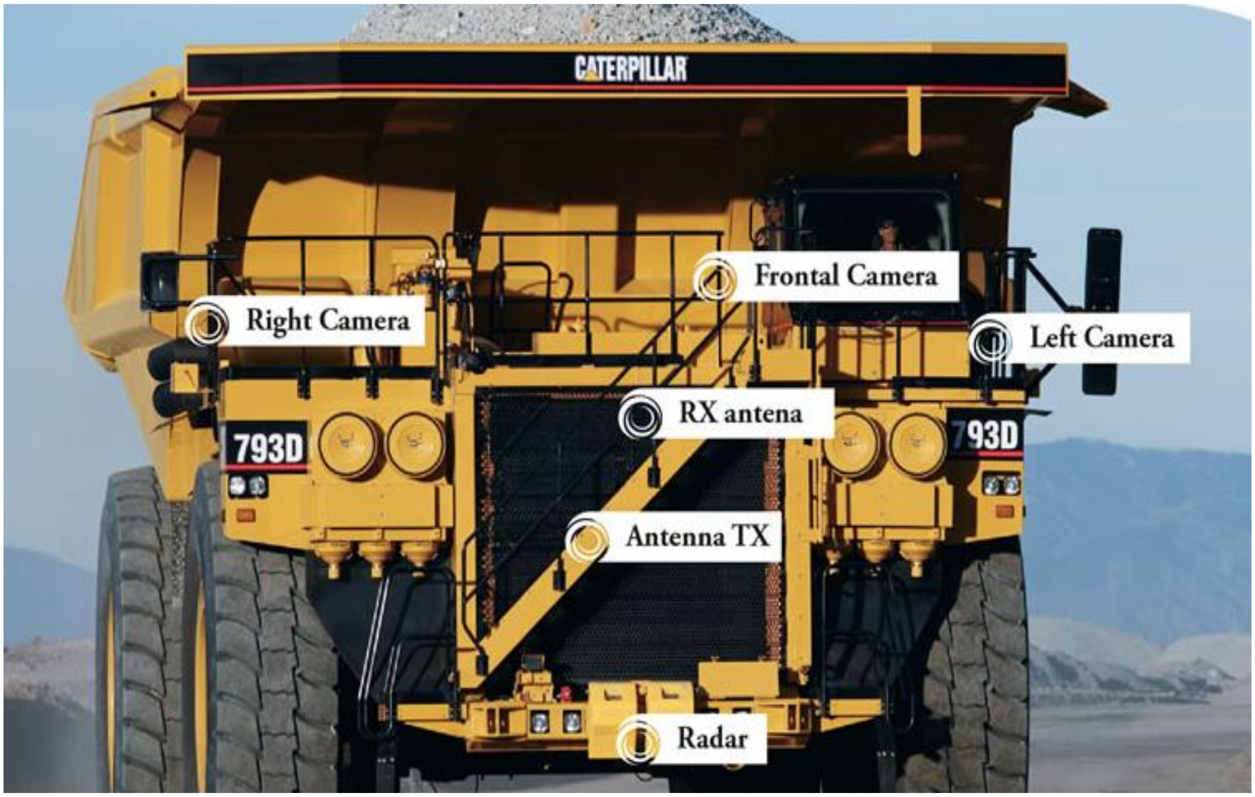


*Figure 2**- Antennas on the mining trucks [www.highservice.com, n.d.]***

This figure shows the location of antennas, radars and cameras. The last few decades of technological developments gave a priceless opportunity for companies. Currently, companies can benefit from technological devices with more affordable prices and it is possible to follow every step in the workplace. Now, expensive vehicles can be equipped with high-tech devices such as antennas, GPS, etc., and companies can remotely watch these vehicles and collect every signal and alarm from them and take crucial actions before harmful breakdowns occur. For these big trucks, left and right-side mirrors are not enough for seeing around the trucks, so several cameras are attached to the trucks. Radar and antennas provide location information and prevent accidents in the mining fields. Furthermore, there are several chips and sensors on various parts of the trucks which send data to the main monitoring data center, so that companies have enough data about their trucks' conditions.

1.2 Big Data in Mining Industry

Big data is advanced data made by the action of computers, portable phones, implanted frameworks and other organized gadgets. Such information became more predominant as innovations such as radio frequency identification (RFID) and telematics progressed (Rouse, 2014). Moreover, machine information has increased utilization of the Internet of Things(IoT), and other big data management technologies that have been developed.

Big data, where different sensors and equipment create more structured data about their operations, performances or conditions, and can be used to complete various analyses, such as process optimization, improved maintenance or machine-to machine communication (Fekete, 2015).

This thesis proposes plan maintenance and shows the importance of big data and the analysis of data. Big data is an inevitable reality for every industry, including the mining industry because prices of ores are very changeable and machines, which are used for mining, are expensive to maintain. For this reason, collecting records of production and machine status are vital for mining companies to make more efficient production plans. Therefore, it is crucial for companies' futures to record and analyze this data and arrange their maintenance and production techniques. Furthermore, almost every big mining company is using machine data, and they save all useful information for their future operations.

Today, thanks to big data and IoT, companies could collect every transaction and details such as drivers' names, vehicles' location, status, weight, speed and so on for each production step. Moreover, this data can be analyzed, and used to make better and more productive work schedules and create less risky work places.

1.3 Maintenance

There is an increasing pressure on companies, urged by worldwide competition, to streamline operations involving item and item-related manufacturing system design, item manufacturing and system maintenance. Maintenance activities are ordinarily performed first by integration of maintenance and process engineering functions, then by application of machines and hardware, and finally, through proactive actions on those machines and equipment including preventive and predictive maintenance (Bastos, Lopes, & Pires, 2012). In literature, it is possible to find three nonspecific sorts of maintenance: Corrective Maintenance, Preventive Maintenance, and Predictive Maintenance.

1.3.1 Corrective Maintenance

Corrective Maintenance(CM) actions are not schedulable, and this makes them harder to plan for and costlier to perform. It is usually not the preferred maintenance because it occurs suddenly and costs valuable money and time (Adolfsson & Dahlström, 2011). Corrective maintenance is used when a system or machine fails. It includes repair and replacement of failed parts to make machines active again.

1.3.2 Preventive Maintenance

Preventive Maintenance (PM) aims at maintaining equipment in satisfactory operating conditions, and is fulfilled by providing for systematic control, detection, and correction of incipient failures, before they cause great defects and usually a PM's planning is created according to equipment manufacturers' advices (www.revolvy.com, n.d.).

Preventive maintenance is routine and tends to follow planned schedules to prevent equipment and machinery breakdowns. The work is preemptively carried out on equipment in order to avoid its breakdown. Despite the benefits of this maintenance, it is not efficient for companies because it costs valuable money and time and does not allow for the use of parts of machines that are capable of working. Even when machines do not work, they may still contain parts that can last their full lifetimes, but preventive maintenance does not calculate for these situations.

1.3.3 Predictive Maintenance

Predictive Maintenance (PDM) has a significant difference from the other maintenance types. During regular operation to reduce failures, PDM directly monitors the status and performance of equipment and provides an opportunity to take precautions before machine failures (www.emaint.com, 2017). Although PDM is more complex compared to the others, it provides several advantages thanks to monitoring. PDM reduces maintenance cost, unnecessary preventative maintenance, unplanned maintenance and provides more efficient work.

This next figure shows three main maintenance types and their definitions. Predictive maintenance is the most efficient maintenance. However, corrective maintenance is the costliest maintenance type.

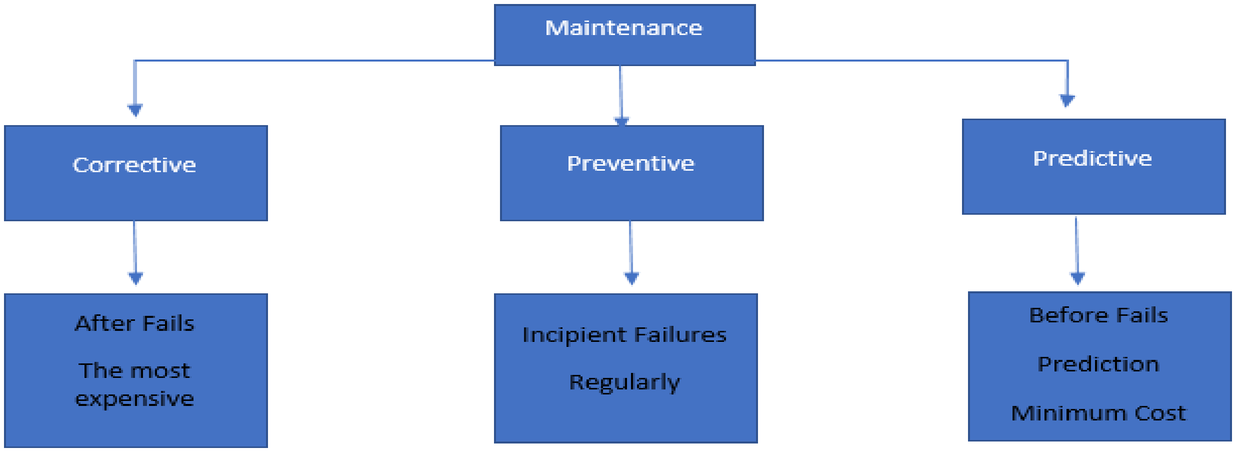


Figure 3- ***Type of maintenance***

1.3.4 The Cost of Maintenance in Mining Industry

Currently, in every industry, mechanization is used for more effective production, and some of the most expensive equipment belongs to the mining industry. For example, according to the United States Census the largest absolute increase in the mining industry's capital investments occurred from 2006 to 2015 (up $75.4 billion or 75.9 percent) (US Census Bureau., 2017). The Electric Power Research Institute (EPRI) has calculated comparative maintenance costs for different maintenance techniques in US dollars per horsepower (HP) per year. Researchers found that a preventive (scheduled) maintenance strategy is the most expensive to run at $24.00 per HP. A corrective (reactive) maintenance strategy is the second most costly at $17.00 per HP but has the additional cost of compromising safety. Maintaining a 750 HP motor with a scheduled maintenance strategy would cost approximately $18,000 per year, while a reactive maintenance strategy would cost $12,750 a year, according to the EPRI study (www.ni.com, 2015). For example, according to Caterpillar (CAT), mining trucks have between 2000 and 5000 HP (www.cat.com, n.d.). When companies calculate this maintenance cost for a mining truck according to the HP calculation, it becomes apparent how costly expenses can become for the companies. The next figure provides an example that shows the percentage of maintenance costs in pit mines for the US.

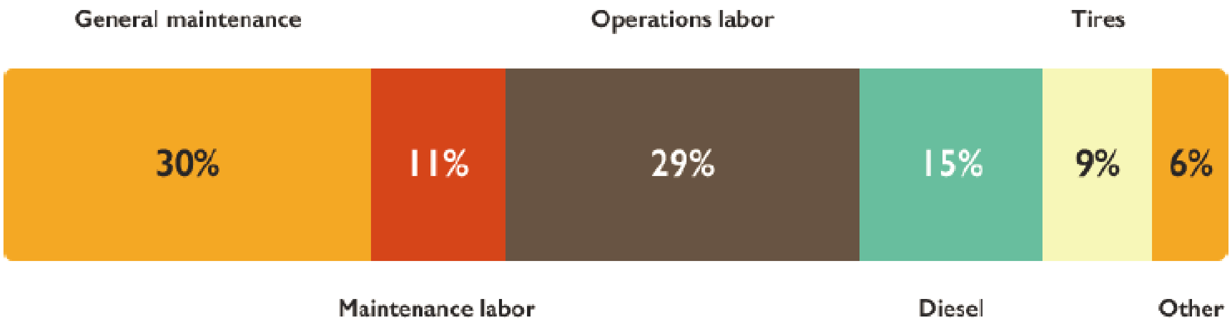


*Figure 4- **Breakdown of direct mining costs in large open pit mines in the US (Fekete, 2015)***

According to Fekete, explanation of this figure: “A mining site’s cost structure consists of three main parts: maintenance, labor and consumables. Figure 4 shows that maintenance related costs account for about one-third of total operational costs (Campbell & Reyes-Picknell, 2006), which makes maintenance the largest controllable cost. It includes items such as replacement parts, human resources, supplies and other items (Lewis & Steinberg, 2001). Though it demonstrates the expenditures of North American open pit mines, other locations and mine types show similar proportions. Mining companies can focus on improving maintenance processes with advanced technologies (such as big data and connected machines), as this area is the largest contributor for their operational expenditures. “(Fekete, 2015).

1.3.5 Predictive Maintenance and Cost

After industrial revolution, companies focus on more profit, less expenses and safer workplaces. For this reason, maintenance, which is one of the biggest expenses for companies, has become more significant. Furthermore, researchers have begun to discover more efficient maintenance techniques, and currently predictive maintenance has become more crucial for companies. Many of these crucial reasons for improving mining equipment reliability and maintainability are summarized as follows: (Peng & Vayenas, 2014).

- to maximize profit
- to reduce the cost of poor reliability/maintainability
- to reduce the use of mining equipment services in an unplanned manner because of short notice
- to provide more accurate short-term forecasts for equipment operating hours
- to overcome challenges imposed by global competition
- to take advantage of lessons learned from other industrial sectors such as aerospace, defense, and nuclear power generation
- to improve workplace safety

Bastos claims that 99 percent of machine failures are identified by a few pointers and fulfilling organizations' requirements leads to heavy expenses in maintenance systems, and maintenance, considered non-value adding, which is continuously evaluated for cost reduction, keeping the machines in excellent working condition (Bastos, Lopes & Pires, 2014). The main goal of predictive maintenance is to find the optimal time for needed maintenance before harmful events may occur and reduce maintenance cost.

Most of the failures do not happen instantaneously, and more often than not there are a few sorts of degradation processes or indications of transitions from typical states to failures. Subsequently, the genuine conditions and their trends ought to be surveyed and anticipated amid the degradation handle, and fitting maintenance actions ought to be taken some time before a breakdown occurs. This is the fundamental target of predictive maintenance.

2 BIG DATA ANALYTICS AND PREDICTIVE MAINTENANCE IN THE MINING INDUSTRY: A LITERATURE REVIEW

Machinery diagnostics and maintenance are tremendous and diverse, primarily due to a wide variety of production systems. Dozens of papers on this subject, which include theories and applicable methods, show up each year in academic journals, conferences and technical reports. In this section, several papers are summarized.

Like other companies, information technology is crucial for mining companies and recently, mining companies have collected huge data by using new high-tech devices such as GPS, monitoring systems, and fleet management systems, etc. As a result of this, currently, companies can analyze these datasets (Yildirim & Dessureault, 2007).

"Big Data in the Mining Industry" was a paper was written by Fekete (2015). In the article, Fekete discusses the challenges of the mining industry, such as, the fast dropping prices of commodities and the technological developments that have forced companies to update their structures. Thanks to the Internet of Things (IoT), several mining companies now have the opportunity to collect big datasets to improve operations and efficiency. Furthermore, huge maintenance expenses force mining companies to make more reasonable maintenance schedules. According to Fekete, after interviewing with some mining companies from Australia, it was concluded that predictive maintenance, Big Data, IoT and Data Analytics create safe workplaces, provide efficient maintenance and decrease maintenance expenditures (Fekete, 2015).

Cartella et al. (2014) defines a different approach for predictive maintenance, called Hidden Semi-Markov Models (HSMMs), and they discuss the theoretical formalization of the model, as well as a few experiments performed using both simulated and factual information with the aim of technique approval. In this paper, all performed tests are able to accurately appraise the current state of the machinery and viably foresee a predefined event with generally low normal absolute error. According to these research results, the model's appropriateness to real-world settings can be advantageous, particularly where, in real-time, the Remaining Useful Lifetime (RUL) of a machine can be calculated, and results show that HSMM would be beneficial for condition monitoring and gauging useful lifetime applications (Cartella, Lemeire, Dimiccoli & Sahli, 2014).

Sihong Peng et al. (2014) presents a study about the implementation of genetic algorithms on Underground Mining Equipment as a case study for predictive maintenance. They assumed that failures of mining equipment caused by an array of factors followed the biological evolution theory. A software was created for predictive maintenance according to their dataset and according to their opinion, these failures follow the natural advancement theory. They used several case studies to focus on practical investigations of a Load Haul Dump (LHD) vehicle with two different terms: three and six months. According to their prediction case studies, a factual test is carried out to look at the similitude between the anticipated data set with the real-life data set in the same period. This research aims at comparing real data to the prediction of this software and analyzes how successful genetic algorithms would be successful for prediction of maintenance. As a result, these two different time interval studies are investigated, and they did not show major impacts of chronological sequence in their prediction results (Peng et al. 2014).

A recent case study named "Earthmoving trucks condition level prediction using neural networks" from Greece by Marinelli et al. (2014) presents an artificial neural network

(ANN) model that predicts earthmoving trucks' condition levels using basic predictors. The results are compared to the respective predictive accuracy of the statistical method of discriminant analysis (DA). In this research, it is created an ANN-based predictive model, and used the capacity, age, kilometers travelled and maintenance level of trucks, which were collected from 126 earthmoving trucks. As a result, data processing identifies specifically a connection between kilometers travelled and maintenance level with the earthmoving trucks' condition grade. Furthermore, they found that the predictive performance of the proposed ANN model is very high for the validation process, and similar findings from the application of DA to the same data set using the same predictors. These models reached above 92 percent accuracy for prediction of trucks' condition level. As a result of that, the prediction decreases downtime, and its reverse influences earthmoving duration and cost, meanwhile increasing the maintenance and replacement policies' impressiveness. This research shows that a sound condition level prediction for earthmoving trucks is achievable through the utilization of easy to collect data and provides a comparative evaluation of the results of two widely applied predictive methods (Marinelli, Lambropoulos, & Petroutsatou, 2014).

A theoretical study by Chen et al. (2016), predicted faults from the data acquisition and fusion strategies, and it used the fault prediction method based on full-vector spectrum which belongs to Dr. Bently and Dr. Muszynska. According to this method, the uncertainty of the spectrum structure can be extracted by the designed data acquisition and fusion method. This method also shows that the reliability of the diagnosis on fault character was improved, and it gives the technical foundation for the prediction and diagnosis research of the fault characters (Chen, Han, Lei, Cui, & Guan, 2016).

Ullah et al. (2017) used one type of Machine Learning Methods for Predictive Maintenance. According to this paper, there were several reasons for increasing the internal temperature of electrical instruments, specifically contact issues, irregular loads, cracks in insulation, defective relays, terminal junctions and other similar issues. As a result of these reasons, they caused intrusive failures and potential damage to power equipment. In this paper, the authors explained the initial prevention mechanism for power substations using a computer-vision approach by taking advantage of infrared thermal images. This work included a total of 150 thermal pictures of different electrical equipment in 10 different substations in operating conditions, using 300 different hotspots. They used multilayer perceptron (MLP) to classify the thermal conditions of components of power substations into defect and non-defect classes. The performance of MLP reached 84 percent of accuracy with graph cut and this result showed the benefit of the proposed defect analysis approach (Ullah, Yang, Khan, Liu, Yang, Gao, & Sun, 2017).

3 SEQUENCE MINING METHODOLOGY FOR PREDICTIVE MAINTENANCE OF MINING TRUCKS

With recent technological developments, collecting and analyzing big data is even vital for predictive maintenance in companies. There are several techniques for analyzing big data. In this thesis, sequential pattern mining techniques is used to make predictions about maintenance timing for mining trucks. The experimental results confirm that sequential pattern mining is suitable approach for discovering information relevant to machine failures.

"Data mining is a process of discovering various models, summaries, and derived values from a given collection of data." (Kantardzic, 2005). Data mining, which is one of the basic processes of Knowledge Discovery in Database (KDD), is the procedure of extracting hidden knowledge or patterns (non-trivial, implicit, previously unknown and potentially useful) from large information warehouses (Zhao & Bhowmick -2003). The next figure shows data mining's main steps.

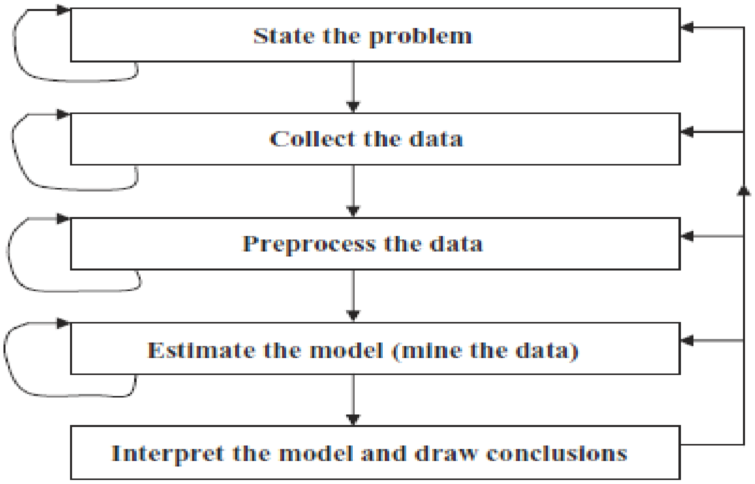

State the problem
Collect the data
Preprocess the data
Estimate the model (mine the data)
Interpret the model and draw conclusions

*Figure 5- **The Data Mining Process (Kantardzic, & Zurada, 2005)***

According to the steps in Figure 5, first state the problem which considers the reasons for machine failures. Second, data is collected for nine months from eleven mining trucks which belong to one mining company in North America. Afterwards, for more accuracy, the dataset must be cleaned, and must be eliminated from useless, duplicate data as a third step. Forth, for estimating model, which would be more useful for the dataset. Lastly, implement this technique and evaluate the results of the data mining process.

3.1 Sequential Pattern Mining

Sequential Pattern Mining is one of the most important mining techniques for analyzing big data. Sequential Pattern Mining (SPM) focuses on finding patterns that occur consecutively in a database or patterns which would be related to time or other values; the main aim is to discover related sequence patterns (Chueh, 2010). Implementation of SPM is very broad and can be used for efficient maintenance of vehicles, natural disasters, sales record analysis, marketing strategies, shopping sequences, medical treatments and DNA sequences, etc.; the subsequences and frequent relevant patterns from the given data can also be found by SPM (Ubaidulla, Sushmitha, & Vanitha, 2017). The aim of pattern mining is to discover useful, recent and unforeseen patterns in databases. A sequence database includes several sequences.

For instance, consider the following database:

Name	**Sequences**
Seq1	a,(b,c,d),(f,g)
Seq2	(a,d),c,b,(a,b,f)
Seq3	c,(a,d,e,f)
Seq4	d,g,a,e,b,b
Seq5	(c,e,g),(a,b)

This database includes four sequences which are seq1, seq2, seq3, seq4 and seq5. For this example, take into consideration that the symbols “a”, “b”, “c”, d”, “e”, “f”, “g” and “h” symbolize different items sold in a supermarket, and “a” could be an “almond”, “b” could be a “box of cereal”, etc.

Now, a sequence is an ordered list of sets of items. For this example, suppose that each sequence shows what a customer bought in a supermarket. Consider the second sequence “seq2”. This sequence shows that the second customer bought items “a” and “d” together, then bought item “c”, then bought “b”, and then bought “a”, “b”, and “f” together.

For the dataset, like this example, there are several failure codes and many alarms and signals which would be indicators named detection groups for these specific breakdowns’ codes. To try to discover several patterns called rules, and afterwards the results will be compared these patterns according to the counts for the last three and the five shifts before machine failures occur. The strength of a detection rate is measured by its support and confidence and calculation of these: (Lui Zhang-2000)

For example, an example of support, when consider two items A and B, and it can calculate the frequency of the item in the dataset. If consider a basket containing 7 items (3-oranges, 4-lemons) then support of any precise value can be calculated by the rate of number of occurrences to the total number of items in the basket (i.e., support(oranges) = 3/7).

An example of confidence, this explains how likely B is purchased when A is purchased. This defines association between two items. For example, when a person buys tea is more likely to buy sugar as well or vice versa. This is measured by the proportion of transactions with item X, in which item Y also appears. As a formulization:

“The support of a rule, X → Y, is the percentage of transactions in T that contains X ∪ Y and can be seen as an estimate of the probability, Pr(X∪Y). The rule support thus determines how frequent the rule is applicable in the transaction set T. Let n be the number of transactions in T. Let n be the number of transactions in T.

The support of the rule X → Y is calculated as follows:

$$\text{Support} = \frac{Count(X \cup Y)}{n}$$

Support is a useful measure because if it is too low, the rule may just occur due to chance. Furthermore, in a business environment, a rule covering too few cases (or transactions) may not be useful because it does not make business sense to act on such a rule (not profitable).

The confidence of a rule, X → Y, is the percentage of transactions in T that contain X also contain Y, and can be seen as an estimate of the conditional probability, Pr(Y | X). It is computed as follows:

$$\text{Confidence} = \frac{Count(X \cup Y)}{Count\ X}$$

Confidence thus determines the predictability of the rule. If the confidence of a rule is too low, one cannot reliably infer or predict Y from X. A rule with low predictability is of limited use." (Lui- Zhang 2000).

3.2 Categories of Patterns

Sequential patterns can be divided into three main categories: periodic patterns, statistically significant patterns, and approximate patterns. However, there are more varieties of models for sequential patterns in the literature (Esmaeili, & Fazekas, 2010).

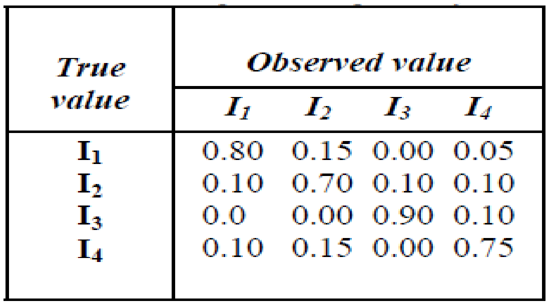
3.2.1 Periodic Patterns

According to Slimani & Lazzez, the main purpose of this model is to discover occurrences of repeated patterns in data and to try to predict future characteristics of real situations; however, this model has several disadvantages, for example, misalignment might cause us to miss some interesting and crucial patterns. Experts have shown, as a solution for this restriction, a pattern might be filled partly to make the model more flexible. As an example, in the series ({a}{b}{c}{a}{b}{c}{a}{b}{c}), the pattern {a}{b}{c} is a periodic pattern because it is repeated with a period equal to three. Each status in the pattern shows the periodicity, and this previous pattern is called a full periodic pattern. As an example, in a sequence like this ({a}{b}{c}{b}{a}{c}{a}{b}{a}{a}{c}{b}), a pattern {a}*{b} where * is a wide range of items, there is no full periodic pattern with length 3, and this is called a partial periodic pattern (Slimani & Lazzez, 2014).

3.2.2 Approximate Patterns

In the real world, there are noisy data in almost every big data, and it is necessary to reduce the effects of these kinds of data on results. Another different and more flexible method is approximate patterns, which has a specific calculation with a compatibility matrix:

“For solving the problem of finding approximate patterns, the concept of compatibility matrix is introduced [4]. This matrix provides a probabilistic connection from observed values to the true values. Based on the compatibility matrix, real support of a pattern can be computed. Table 2 gives an example of the compatibility matrix.



True value	*Observed value*			
	I_1	*I_2*	*I_3*	*I_4*
I_1	0.80	0.15	0.00	0.05
I_2	0.10	0.70	0.10	0.10
I_3	0.0	0.00	0.90	0.10
I_4	0.10	0.15	0.00	0.75

Table 2 Compatibility Matrix

For example, an observed I4 corresponds to a true occurrence of I1, I2, I3, and I4 with probability C(I1,I4)=0.05 , C(I2,I4)=0.10 , C(I3,I4)=0.10 , and C(I4,I4)=0.75 ,respectively. Compatibility matrix usually is given by some domain expert but there are some ways to obtain and justify the value of each entry in the matrix so that even with a certain degree of error contained in matrix, sequential pattern mining algorithm can still produce results of reasonable quality.

A new metric, namely match is defined to quantify the significance of a pattern. The combined effect of support and match may need to scan the entire sequence database many times. Similar to other data mining methods, to tackle this problem sampling-based algorithms can be used.

Consequently, the number of scans through the entire database is minimized." (Esmaeili, & Fazekas, 2010).

3.2.3 Statistically Significant Patterns

The calculation of support and confidence are crucial for sequential pattern mining, but just using these supporting values as a standardization measure might cause one to skip several important patterns; various data mining applications have tried to find a valid solution for this problem (Slimani & Lazzez, 2014). On the other side, according to Esmaeili & Gabor (2010), the number of occurrences (support) might be misleading, and there is no direct ratio between a repetitive number of patterns and a significance of patterns. Because of this, in several situations, many occurrences of an expected frequent pattern may not be as important as a few occurrences of an expected uncommon pattern, which is called surprising pattern instead of frequent pattern. In addition, the support threshold must be set very low to discover a small number of patterns with high information gain, and the information gain metric might be helpful to evaluate the degree of surprise of the pattern (Esmaeili, & Fazekas, 2010).

The next step is deciding k most significant patterns, and this can be easily achieved by using a threshold value and the best k patterns that have an information gain greater than the specified threshold should be returned; however, the problem of the information gain value is difficult to define the location of the occurrences of the patterns (Slimani & Lazzez, 2014). In a statistically significant method, the calculation of info gains and dataset and other values are crucial to get better results. For example:

Two input patterns' sequences such as: S1=({a}{b}{c}{b}{a}{b}{d}{c}{a}{b}{b}{d}),and S2=({b}{c}{d}{b}{a}{b}{a}{b}{a}{b}{d}{c}), then the pattern {a}{b} has the same (three times occurred) information gain in the two sequences, it is dispersed in S1 but repeats consecutively in S2 (Slimani & Lazzez, 2014).

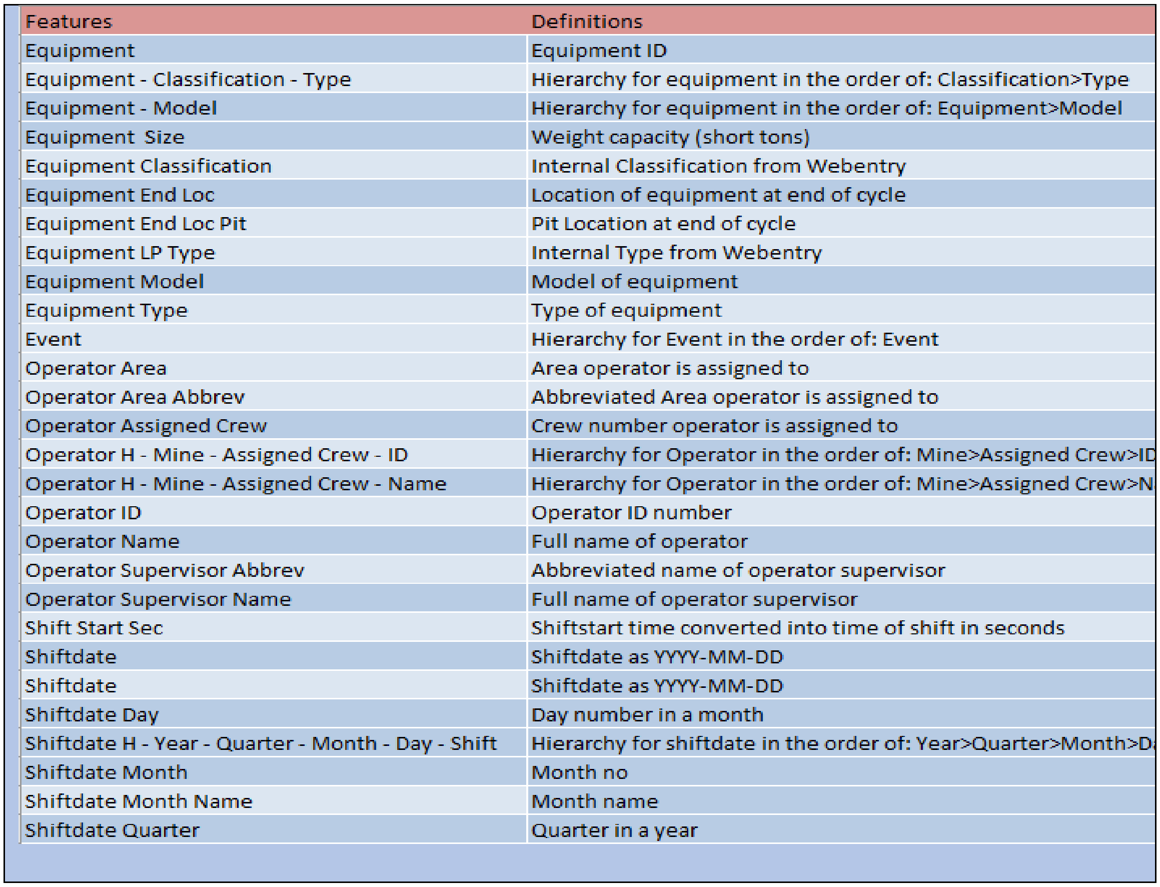
3.3 Dataset Details

In this thesis, the data was collected for nine months by eleven mining trucks which belong to a mine in North America. The dataset was provided by a mining engineer, Mustafa Kahraman, who works with the mine. In this dataset, the trucks worked twenty-four hours a day, seven days a week and every shift represented a twelve-hour period. The dataset has more than three million rows and more than one hundred columns. This dataset included three main tables: status, production, and machine health status.

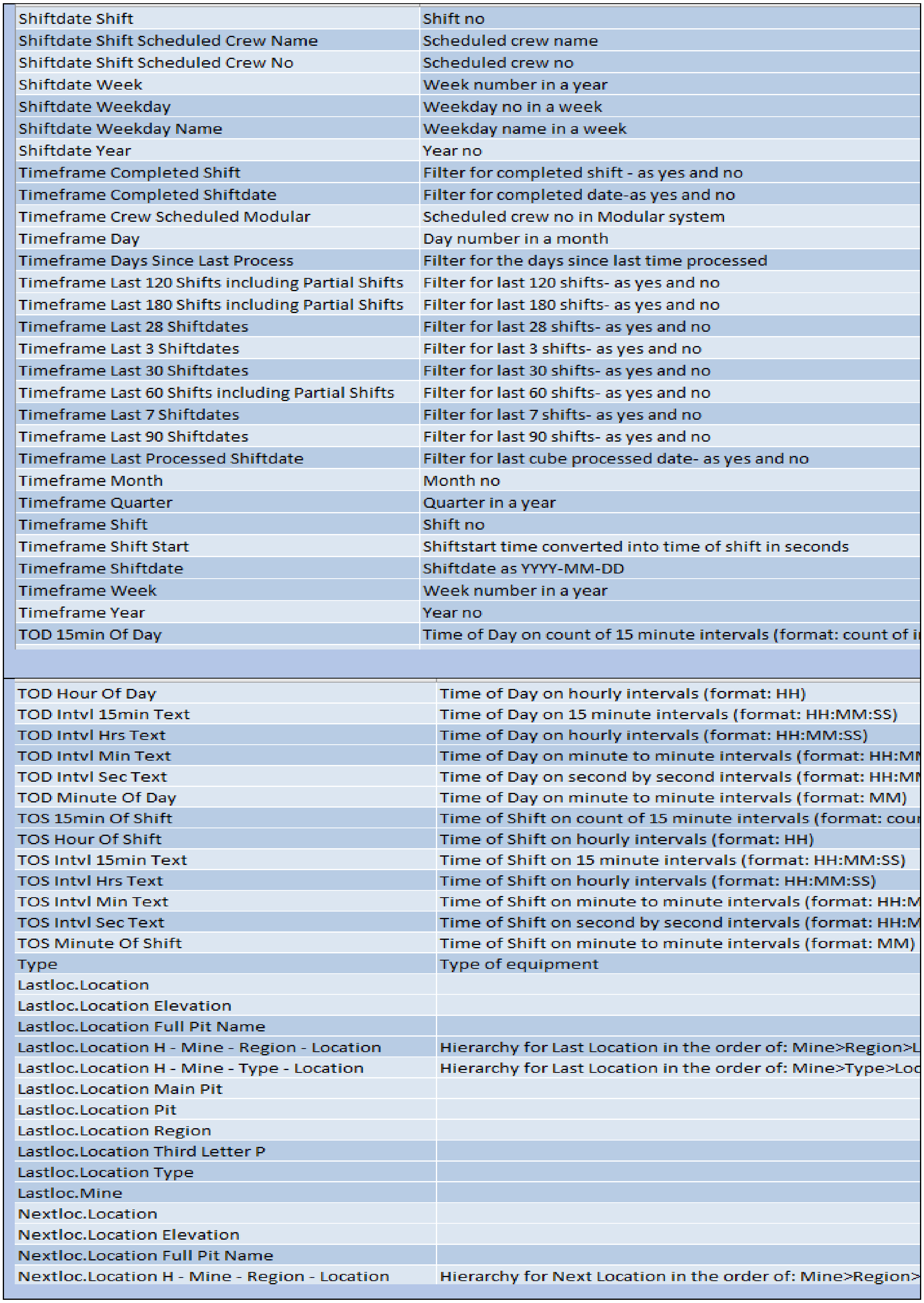
First, the status table is where equipment statuses are recorded: fail status, standby, production etc. Second, the production table is where truck cycles are saved with all associated details: shift date, driver name, material type, speed, location and so on. Lastly, the machine health status table is created by the machine health information for selected trucks generated by chips and sensors which are connected to different parts of the trucks.

3.3.1 Primary Features in Dataset and Summarization of General Processes

These next tables show the short definitions of primary features for the dataset. Experts created all these definitions and designed the database. After showing these definitions, these are next steps, making feature selection, select necessary parts of data, eliminate duplicate and missing values, and implement sequential pattern mining techniques. In all these processes, SQL Server Management Studio will be used as a tool. The next figures show some initial columns and short definitions of features for the dataset. In this thesis the main focused feature is that events column which is created by alarms and signals.



Features	Definitions
Equipment	Equipment ID
Equipment - Classification - Type	Hierarchy for equipment in the order of: Classification>Type
Equipment - Model	Hierarchy for equipment in the order of: Equipment>Model
Equipment Size	Weight capacity (short tons)
Equipment Classification	Internal Classification from Webentry
Equipment End Loc	Location of equipment at end of cycle
Equipment End Loc Pit	Pit Location at end of cycle
Equipment LP Type	Internal Type from Webentry
Equipment Model	Model of equipment
Equipment Type	Type of equipment
Event	Hierarchy for Event in the order of: Event
Operator Area	Area operator is assigned to
Operator Area Abbrev	Abbreviated Area operator is assigned to
Operator Assigned Crew	Crew number operator is assigned to
Operator H - Mine - Assigned Crew - ID	Hierarchy for Operator in the order of: Mine>Assigned Crew>ID
Operator H - Mine - Assigned Crew - Name	Hierarchy for Operator in the order of: Mine>Assigned Crew>N
Operator ID	Operator ID number
Operator Name	Full name of operator
Operator Supervisor Abbrev	Abbreviated name of operator supervisor
Operator Supervisor Name	Full name of operator supervisor
Shift Start Sec	Shiftstart time converted into time of shift in seconds
Shiftdate	Shiftdate as YYYY-MM-DD
Shiftdate	Shiftdate as YYYY-MM-DD
Shiftdate Day	Day number in a month
Shiftdate H - Year - Quarter - Month - Day - Shift	Hierarchy for shiftdate in the order of: Year>Quarter>Month>D
Shiftdate Month	Month no
Shiftdate Month Name	Month name
Shiftdate Quarter	Quarter in a year



Shiftdate Shift	Shift no
Shiftdate Shift Scheduled Crew Name	Scheduled crew name
Shiftdate Shift Scheduled Crew No	Scheduled crew no
Shiftdate Week	Week number in a year
Shiftdate Weekday	Weekday no in a week
Shiftdate Weekday Name	Weekday name in a week
Shiftdate Year	Year no
Timeframe Completed Shift	Filter for completed shift - as yes and no
Timeframe Completed Shiftdate	Filter for completed date-as yes and no
Timeframe Crew Scheduled Modular	Scheduled crew no in Modular system
Timeframe Day	Day number in a month
Timeframe Days Since Last Process	Filter for the days since last time processed
Timeframe Last 120 Shifts including Partial Shifts	Filter for last 120 shifts- as yes and no
Timeframe Last 180 Shifts including Partial Shifts	Filter for last 180 shifts- as yes and no
Timeframe Last 28 Shiftdates	Filter for last 28 shifts- as yes and no
Timeframe Last 3 Shiftdates	Filter for last 3 shifts- as yes and no
Timeframe Last 30 Shiftdates	Filter for last 30 shifts- as yes and no
Timeframe Last 60 Shifts including Partial Shifts	Filter for last 60 shifts- as yes and no
Timeframe Last 7 Shiftdates	Filter for last 7 shifts- as yes and no
Timeframe Last 90 Shiftdates	Filter for last 90 shifts- as yes and no
Timeframe Last Processed Shiftdate	Filter for last cube processed date- as yes and no
Timeframe Month	Month no
Timeframe Quarter	Quarter in a year
Timeframe Shift	Shift no
Timeframe Shift Start	Shiftstart time converted into time of shift in seconds
Timeframe Shiftdate	Shiftdate as YYYY-MM-DD
Timeframe Week	Week number in a year
Timeframe Year	Year no
TOD 15min Of Day	Time of Day on count of 15 minute intervals (format: count of i
TOD Hour Of Day	Time of Day on hourly intervals (format: HH)
TOD Intvl 15min Text	Time of Day on 15 minute intervals (format: HH:MM:SS)
TOD Intvl Hrs Text	Time of Day on hourly intervals (format: HH:MM:SS)
TOD Intvl Min Text	Time of Day on minute to minute intervals (format: HH:MM
TOD Intvl Sec Text	Time of Day on second by second intervals (format: HH:MM
TOD Minute Of Day	Time of Day on minute to minute intervals (format: MM)
TOS 15min Of Shift	Time of Shift on count of 15 minute intervals (format: cou
TOS Hour Of Shift	Time of Shift on hourly intervals (format: HH)
TOS Intvl 15min Text	Time of Shift on 15 minute intervals (format: HH:MM:SS)
TOS Intvl Hrs Text	Time of Shift on hourly intervals (format: HH:MM:SS)
TOS Intvl Min Text	Time of Shift on minute to minute intervals (format: HH:M
TOS Intvl Sec Text	Time of Shift on second by second intervals (format: HH:M
TOS Minute Of Shift	Time of Shift on minute to minute intervals (format: MM)
Type	Type of equipment
Lastloc.Location	
Lastloc.Location Elevation	
Lastloc.Location Full Pit Name	
Lastloc.Location H - Mine - Region - Location	Hierarchy for Last Location in the order of: Mine>Region>L
Lastloc.Location H - Mine - Type - Location	Hierarchy for Last Location in the order of: Mine>Type>Loc
Lastloc.Location Main Pit	
Lastloc.Location Pit	
Lastloc.Location Region	
Lastloc.Location Third Letter P	
Lastloc.Location Type	
Lastloc.Mine	
Nextloc.Location	
Nextloc.Location Elevation	
Nextloc.Location Full Pit Name	
Nextloc.Location H - Mine - Region - Location	Hierarchy for Next Location in the order of: Mine>Region>

Figure 6- ***Definitions of Features***

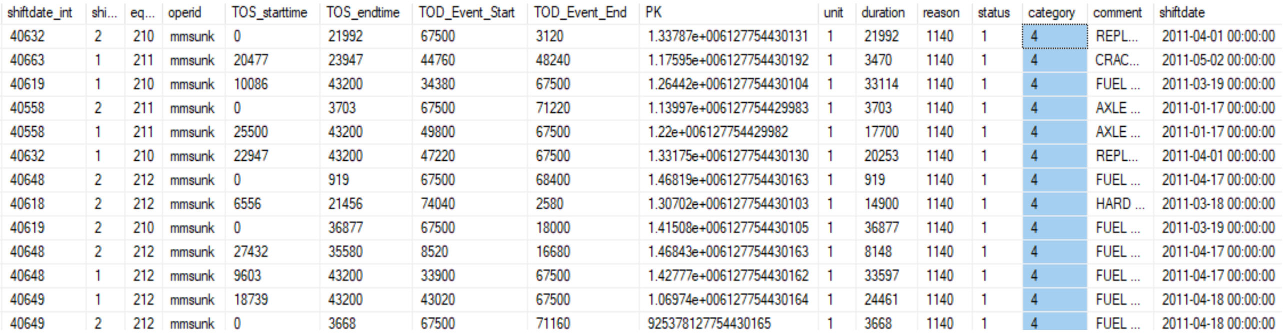
3.3.2 Preprocessing of Dataset and Features Selection

Before analyzing the data, as preprocessing steps, first removed most of the features that were not going to be used in the analysis. Furthermore, there are Non-Applicable (N/A) values and missing values. These would not be useful for the analysis and removed this data. In addition, duplicate data was deleted before implementing data mining techniques.

In making the decision to select table features, it was consulted an expert who worked in this mining company and is knowledgeable about these trucks and the database. According to the expert, there would be a lot of distinctive features which may have an impact on breakdowns of trucks, such as driver mistakes, overloading, locations, and so on.

Furthermore, according to this expert, apart from these reasons there are several chips and sensors which record the trucks' status, signals, changes and alarms. After receiving this information, it was decided to analyze more mechanical failures which may be indicative of alarms and signals, which are called events in the database. Other identifying features need more in-depth research, truck expertise, and workplace-specific knowledge.

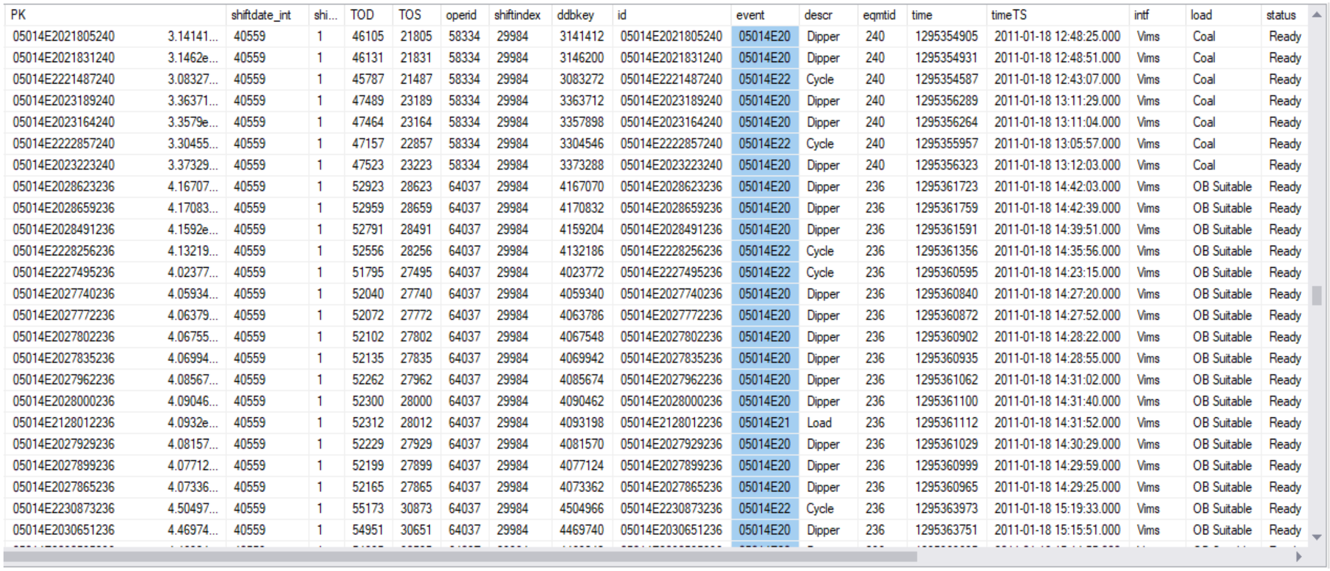
First, it will be used the Status Table which includes the shift number, time, date, reason number, category and so on. The category column shows the machine status and when it shows a breakdown status (when category equals 4), as an example to select one reason code, for example 1140, and a related time, date, and shift number. Afterwards, it will be combined this data with the events column from the Machine Health Status. Figure 6 is as an example of the Status Table:



shiftdate_int	shi...	eq...	operid	TOS_starttime	TOS_endtime	TOD_Event_Start	TOD_Event_End	PK	unit	duration	reason	status	category	comment	shiftdate
40632	2	210	mmsunk	0	21992	67500	3120	1.33787e+006127754430131	1	21992	1140	1	4	REPL...	2011-04-01 00:00:00
40663	1	211	mmsunk	20477	23947	44760	48240	1.17595e+006127754430192	1	3470	1140	1	4	CRAC...	2011-05-02 00:00:00
40619	1	210	mmsunk	10086	43200	34380	67500	1.26442e+006127754430104	1	33114	1140	1	4	FUEL ...	2011-03-19 00:00:00
40558	2	211	mmsunk	0	3703	67500	71220	1.13997e+006127754429983	1	3703	1140	1	4	AXLE ...	2011-01-17 00:00:00
40558	1	211	mmsunk	25500	43200	49800	67500	1.22e+006127754429982	1	17700	1140	1	4	AXLE ...	2011-01-17 00:00:00
40632	1	210	mmsunk	22947	43200	47220	67500	1.33175e+006127754430130	1	20253	1140	1	4	REPL...	2011-04-01 00:00:00
40648	2	212	mmsunk	0	919	67500	68400	1.46819e+006127754430163	1	919	1140	1	4	FUEL ...	2011-04-17 00:00:00
40618	2	212	mmsunk	6556	21456	74040	2580	1.30702e+006127754430103	1	14900	1140	1	4	HARD ...	2011-03-18 00:00:00
40619	2	210	mmsunk	0	36877	67500	18000	1.41508e+006127754430105	1	36877	1140	1	4	FUEL ...	2011-03-19 00:00:00
40648	2	212	mmsunk	27432	35580	8520	16680	1.46843e+006127754430163	1	8148	1140	1	4	FUEL ...	2011-04-17 00:00:00
40648	1	212	mmsunk	9603	43200	33900	67500	1.42777e+006127754430162	1	33597	1140	1	4	FUEL ...	2011-04-17 00:00:00
40649	1	212	mmsunk	18739	43200	43020	67500	1.06974e+006127754430164	1	24461	1140	1	4	FUEL ...	2011-04-18 00:00:00
40649	2	212	mmsunk	0	3668	67500	71160	925378127754430165	1	3668	1140	1	4	FUEL ...	2011-04-18 00:00:00

Figure 7- An Example of Breakdowns on the Status Table

As a next step, it will be selected related events' column data from the Machine Health Status Table, which was collected from signals and chips before and during the same failures' status. In this table, the events column is the most important column for the analyses because after selecting related events, it will be preprocessed this data in an attempt to discover some patterns. Figure 7 is as an example of the Machine Health Status table:



PK		shiftdate_int	shi...	TOD	TOS	operid	shiftindex	ddbkey	id	event	descr	eqmtid	time	timeTS	intf	load	status
05014E2021805240	3.14141...	40559	1	46105	21805	58334	29984	3141412	05014E2021805240	05014E20	Dipper	240	1295354905	2011-01-18 12:48:25.000	Vims	Coal	Ready
05014E2021831240	3.1462e...	40559	1	46131	21831	58334	29984	3146200	05014E2021831240	05014E20	Dipper	240	1295354931	2011-01-18 12:48:51.000	Vims	Coal	Ready
05014E2221487240	3.08327...	40559	1	45787	21487	58334	29984	3083272	05014E2221487240	05014E22	Cycle	240	1295354587	2011-01-18 12:43:07.000	Vims	Coal	Ready
05014E2023189240	3.36371...	40559	1	47489	23189	58334	29984	3363712	05014E2023189240	05014E20	Dipper	240	1295356289	2011-01-18 13:11:29.000	Vims	Coal	Ready
05014E2023164240	3.3579e...	40559	1	47464	23164	58334	29984	3357898	05014E2023164240	05014E20	Dipper	240	1295356264	2011-01-18 13:11:04.000	Vims	Coal	Ready
05014E2222857240	3.30455...	40559	1	47157	22857	58334	29984	3304546	05014E2222857240	05014E22	Cycle	240	1295355957	2011-01-18 13:05:57.000	Vims	Coal	Ready
05014E2023223240	3.37329...	40559	1	47523	23223	58334	29984	3373288	05014E2023223240	05014E20	Dipper	240	1295356323	2011-01-18 13:12:03.000	Vims	Coal	Ready
05014E2028623236	4.16707...	40559	1	52923	28623	64037	29984	4167070	05014E2028623236	05014E20	Dipper	236	1295361723	2011-01-18 14:42:03.000	Vims	OB Suitable	Ready
05014E2028659236	4.17083...	40559	1	52959	28659	64037	29984	4170832	05014E2028659236	05014E20	Dipper	236	1295361759	2011-01-18 14:42:39.000	Vims	OB Suitable	Ready
05014E2028491236	4.1592e...	40559	1	52791	28491	64037	29984	4159204	05014E2028491236	05014E20	Dipper	236	1295361591	2011-01-18 14:39:51.000	Vims	OB Suitable	Ready
05014E2228256236	4.13219...	40559	1	52556	28256	64037	29984	4132186	05014E2228256236	05014E22	Cycle	236	1295361356	2011-01-18 14:35:56.000	Vims	OB Suitable	Ready
05014E2227495236	4.02377...	40559	1	51795	27495	64037	29984	4023772	05014E2227495236	05014E22	Cycle	236	1295360595	2011-01-18 14:23:15.000	Vims	OB Suitable	Ready
05014E2027740236	4.05934...	40559	1	52040	27740	64037	29984	4059340	05014E2027740236	05014E20	Dipper	236	1295360840	2011-01-18 14:27:20.000	Vims	OB Suitable	Ready
05014E2027772236	4.06379...	40559	1	52072	27772	64037	29984	4063786	05014E2027772236	05014E20	Dipper	236	1295360872	2011-01-18 14:27:52.000	Vims	OB Suitable	Ready
05014E2027802236	4.06755...	40559	1	52102	27802	64037	29984	4067548	05014E2027802236	05014E20	Dipper	236	1295360902	2011-01-18 14:28:22.000	Vims	OB Suitable	Ready
05014E2027835236	4.06994...	40559	1	52135	27835	64037	29984	4069942	05014E2027835236	05014E20	Dipper	236	1295360935	2011-01-18 14:28:55.000	Vims	OB Suitable	Ready
05014E2027962236	4.08567...	40559	1	52262	27962	64037	29984	4085674	05014E2027962236	05014E20	Dipper	236	1295361062	2011-01-18 14:31:02.000	Vims	OB Suitable	Ready
05014E2028000236	4.09046...	40559	1	52300	28000	64037	29984	4090462	05014E2028000236	05014E20	Dipper	236	1295361100	2011-01-18 14:31:40.000	Vims	OB Suitable	Ready
05014E2128012236	4.0932e...	40559	1	52312	28012	64037	29984	4093198	05014E2128012236	05014E21	Load	236	1295361112	2011-01-18 14:31:52.000	Vims	OB Suitable	Ready
05014E2027929236	4.08157...	40559	1	52229	27929	64037	29984	4081570	05014E2027929236	05014E20	Dipper	236	1295361029	2011-01-18 14:30:29.000	Vims	OB Suitable	Ready
05014E2027899236	4.07712...	40559	1	52199	27899	64037	29984	4077124	05014E2027899236	05014E20	Dipper	236	1295360999	2011-01-18 14:29:59.000	Vims	OB Suitable	Ready
05014E2027865236	4.07336...	40559	1	52165	27865	64037	29984	4073362	05014E2027865236	05014E20	Dipper	236	1295360965	2011-01-18 14:29:25.000	Vims	OB Suitable	Ready
05014E2230873236	4.50497...	40559	1	55173	30873	64037	29984	4504966	05014E2230873236	05014E22	Cycle	236	1295363973	2011-01-18 15:19:33.000	Vims	OB Suitable	Ready
05014E2030651236	4.46974...	40559	1	54951	30651	64037	29984	4469740	05014E2030651236	05014E20	Dipper	236	1295363751	2011-01-18 15:15:51.000	Vims	OB Suitable	Ready

Figure 8- An example of Events Column on the Machine Health Status Table

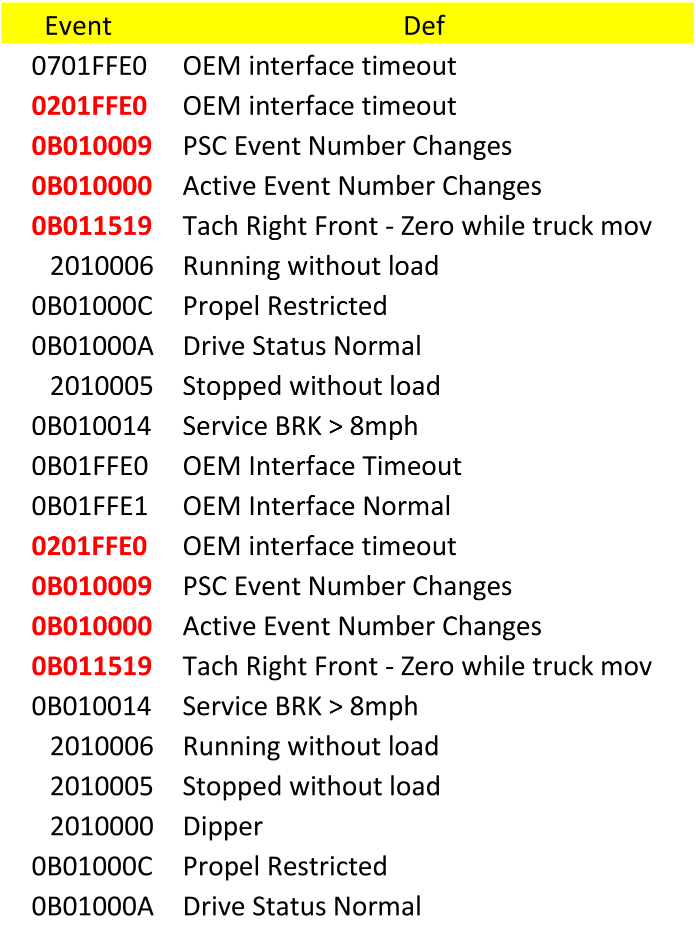
There are 35 different defined machine breakdown reasons in the dataset, and it will be implemented the same processes for three different failure codes which are related to more mechanical problems according to the expert. Next, it will be discovered several patterns between the same breakdown codes consecutively, and after that it will be discovered several patterns in the last three and five shifts' events column data occurring before the shift in which machine failures occur for the same failure codes. After that, it will be analyzed how many times these patterns occurred and compare pattern numbers with the last three and five shifts, which would be related failures according to these patterns' confidence calculation values.

3.4 Implementation of the Sequential Pattern Mining

Discovering unexpected and useful patterns in databases is the fundamental data mining task. In recent years, a trend in data mining has been to design algorithms for discovering patterns in sequential data. One of the most popular data mining techniques for finding patterns is sequential

pattern mining. It consists of discovering interesting subsequence patterns in a set of sequences, where the remarkable subsequences can be measured in terms of various criteria such as their occurrence frequency, length, and so on (Viger, 2017).

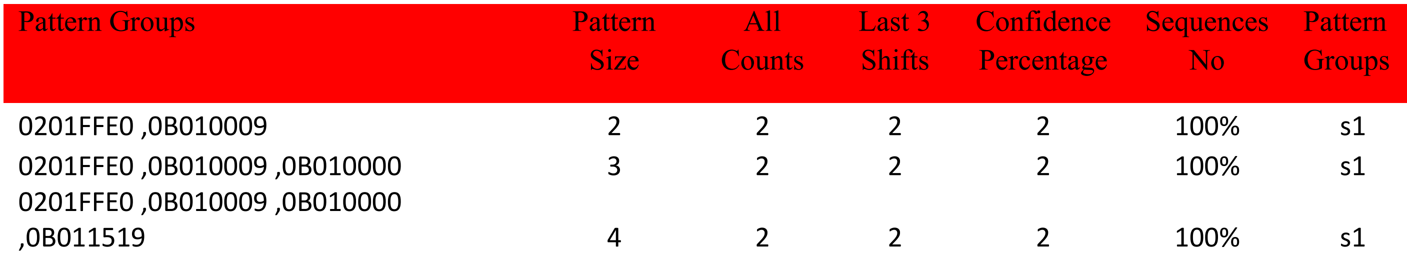
After deciding which sequential pattern mining technique would be a more appropriate technique for analyzing the alarms/signals, SQL Management Studio will be used as a tool, which has several functions to find patterns with T-SQL. It will be tried to discover some patterns within each breakdown reason after ordering them based on their date. For example, figure X shows the short part of event data between two 1140 failure codes, and it will be checked this data to find 2,3,4, or 5 groups of sequential patterns. Next, it will be compared it with the last three and five shifts' patterns. If there are any specific patterns in these last three shifts and five shifts, then it would be a reasonable indicator of this specific failure code. Additionally, the same processes will be implemented for four different breakdown codes.



Event	Def
0701FFE0	OEM interface timeout
0201FFE0	OEM interface timeout
0B010009	PSC Event Number Changes
0B010000	Active Event Number Changes
0B011519	Tach Right Front - Zero while truck mov
2010006	Running without load
0B01000C	Propel Restricted
0B01000A	Drive Status Normal
2010005	Stopped without load
0B010014	Service BRK > 8mph
0B01FFE0	OEM Interface Timeout
0B01FFE1	OEM Interface Normal
0201FFE0	OEM interface timeout
0B010009	PSC Event Number Changes
0B010000	Active Event Number Changes
0B011519	Tach Right Front - Zero while truck mov
0B010014	Service BRK > 8mph
2010006	Running without load
2010005	Stopped without load
2010000	Dipper
0B01000C	Propel Restricted
0B01000A	Drive Status Normal

*Figure 9- **Small part of event data columns after preprocessing for reason 1140***

Figure 8 shows a small part of the data which comes from the preprocessed data. The first column is the main column, named the event column, which shows codes for events. The second column, named def, illustrates definitions of events. In the event column, the rows which are bold and red in color, illustrate a simple example of one pattern for this data. Normally there are more than seven thousand rows of data after preprocessing between the same reason codes (code 1140). However, this example is a small part of this data and contains one pattern example. An example of finding a pattern for this small data result is illustrated in the next table. These patterns are a combination of common elements with varied sizes. These patterns occur in the same place and in the same shifts before the same failure code.



Pattern Groups	Pattern Size	All Counts	Last 3 Shifts	Confidence Percentage	Sequences No	Pattern Groups
0201FFE0 ,0B010009	2	2	2	2	100%	s1
0201FFE0 ,0B010009 ,0B010000	3	2	2	2	100%	s1
0201FFE0 ,0B010009 ,0B010000 ,0B011519	4	2	2	2	100%	s1

*Figure 10- **An example of patterns for breakdown reason 1140***

In this table, the first column shows elements of patterns, the second column is a pattern size, meaning how many events are included in this pattern, and the third column belongs to the last three shifts before machine failure occurs. For the fourth column, it shows how many times this pattern occurred in the last five shifts before the machine failure. The fifth column shows how

many times, this pattern is found between the same failure reasons. The sixth column is the confidence value which is calculated by dividing the last three or five shifts' number of patterns by the "all count" pattern size. Lastly, the seventh column shows that all these patterns include common events which means they occurred in the same place in the dataset, and they have varied sizes and elements. It will be made sequential groups for them, as illustrated by the last column. After this process, a new table is to show which sequence groups are related to which number of failures. The failure code 1140 occurred five times, and after finding patterns, it will be separated every group of sequence related to their common elements. It will be done the same process for four different breakdowns, after which will be illustrated the results according to their confidence values.

4 EXPERIMENTAL RESULTS: FREQUENT SEQUENCES

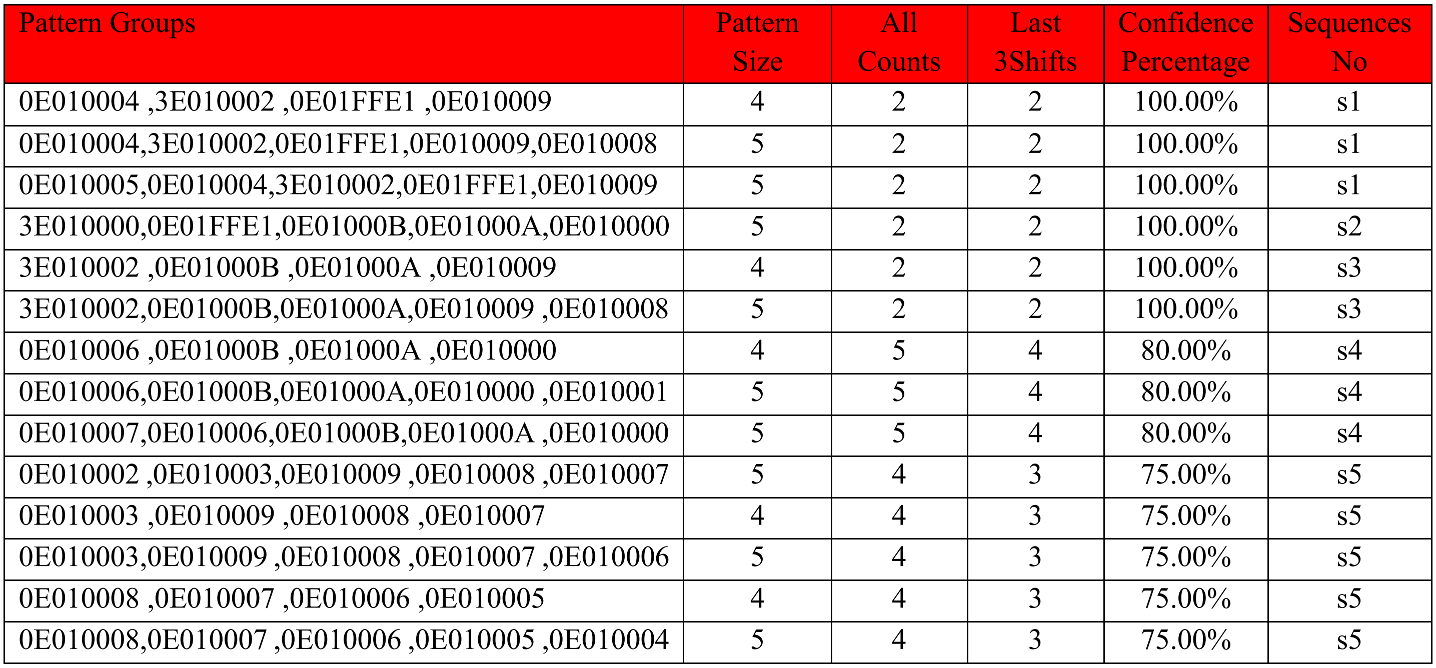
In this section, it will be shared the pattern results with several graphical illustrations. It will be done the same processes for three different failure codes which are more related to mechanical breakdowns apart from other causes**.** Mechanical breakdowns have more expensive maintenance costs and are more related to alarms and signals which referred to as events in the dataset. The codes are 1104,1140 and 1143. First of all, it will be shown the groups' patterns data which are above a minimum of 70 percent confidence. Next, it will be illustrated these results as graphs according to percentages.

4.1 Results for Failure code 1104

It will be separated each failure code result into two parts according to the last shift numbers, which are the last three and last five shifts.

4.1.1 Pattern Results for Last 3 Shifts

The next figure belongs to the last three shift patterns and is between two of the same failure codes called "all counts". It was calculated these pattern groups' confidence percentages and ordered them from largest to lowest values for 1104 breakdown codes.



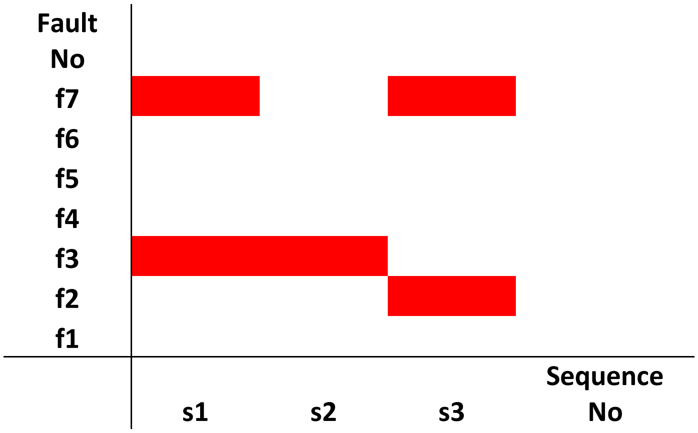
Pattern Groups	Pattern Size	All Counts	Last 3Shifts	Confidence Percentage	Sequences No
0E010004 ,3E010002 ,0E01FFE1 ,0E010009	4	2	2	100.00%	s1
0E010004,3E010002,0E01FFE1,0E010009,0E010008	5	2	2	100.00%	s1
0E010005,0E010004,3E010002,0E01FFE1,0E010009	5	2	2	100.00%	s1
3E010000,0E01FFE1,0E01000B,0E01000A,0E010000	5	2	2	100.00%	s2
3E010002 ,0E01000B ,0E01000A ,0E010009	4	2	2	100.00%	s3
3E010002,0E01000B,0E01000A,0E010009 ,0E010008	5	2	2	100.00%	s3
0E010006 ,0E01000B ,0E01000A ,0E010000	4	5	4	80.00%	s4
0E010006,0E01000B,0E01000A,0E010000 ,0E010001	5	5	4	80.00%	s4
0E010007,0E010006,0E01000B,0E01000A ,0E010000	5	5	4	80.00%	s4
0E010002 ,0E010003,0E010009 ,0E010008 ,0E010007	5	4	3	75.00%	s5
0E010003 ,0E010009 ,0E010008 ,0E010007	4	4	3	75.00%	s5
0E010003,0E010009 ,0E010008 ,0E010007 ,0E010006	5	4	3	75.00%	s5
0E010008 ,0E010007 ,0E010006 ,0E010005	4	4	3	75.00%	s5
0E010008,0E010007 ,0E010006 ,0E010005 ,0E010004	5	4	3	75.00%	s5

Figure 11- ***Patterns for failure code 1104, comparing last three shifts and confidence percentages***

After making groups for these patterns, it will be created a table for sequence groups which have a hundred percent confidence and related failure times. For example, groups of s1 occurred before the third and seventh breakdown for 1104**.** It will be shown this whole groups of sequences and their related order of failures in the next two figures. The first one will be illustrated it as a table, and second one will be as a graphical according to failure timing and sequence numbers.

Faults No	100%
f1	
f2	s3
f3	s1,s2
f4	
f5	
f6	
f7	s1,s3

Table



Graph

*Figure 12- **Sequence group distributions according to breakdown orders for failure code 1104 in last three shifts***

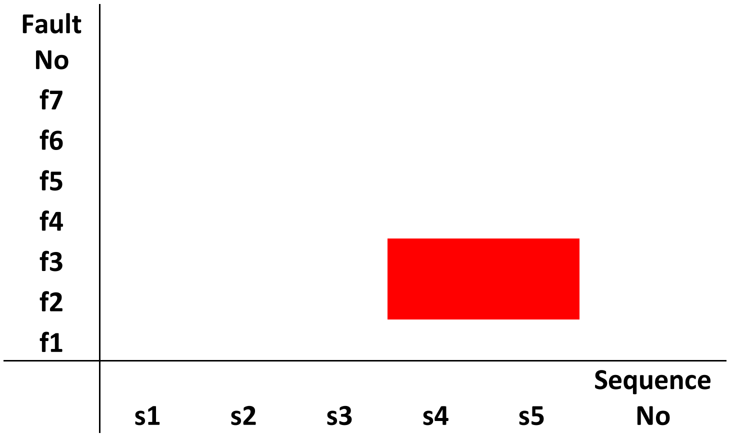
In this table, it can be seen there are seven faults for reason 1104. Before failures occur, there are three different sequential pattern groups. S3 occurred before the second failure and s1, s2 before the third breakdown and s1, s3 before the seventh breakdown in last three shifts. It can be seen these information as a table and graph in Figure 12.

When detection rate is calculated which equals by dividing counts of filled rows in the table by fault occurring times. For this example, detection rate equals the count of filled rows in the table which is 3 divided by total faults' count (7), so it comes to a 43% detection rate. This percentage means that the system can detect 4 out of 10 breakdowns situations.

For more accuracy, decreasing the confidence value to 70%, which might increase the detection rate. The next figure shows sequence groups between 70% and 100% and a distribution of these groups as a table and graph.

Faults No	>70%
f1	
f2	s4,s5
f3	s4,s5
f4	
f5	
f6	
f7	

Table



Graph

*Figure 13- **Sequence group distributions according to breakdown orders for failure code 1104 in last three shifts above 70%***

From this table it can be seen s4 and s5 occurred before the second and third failures, and this did not change the detection rate. As it can be seen from Table in Figure 12; s1, s2 and s3 occurred before faults 1 and 2.

The below graph of these patterns was created according to percentages of confidence.

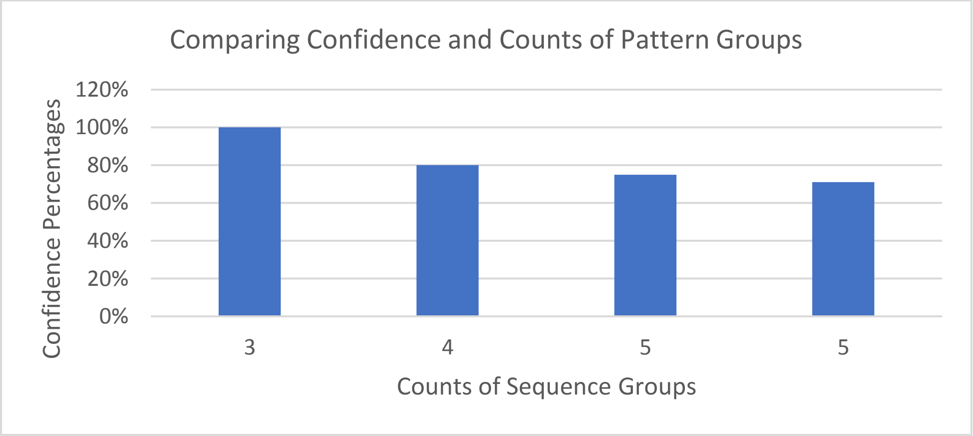


*Figure 14- **Graphical illustration of patterns and confidence for failure code 1104 in last three shifts***

For confidence values from 100% to 70%, the count of patterns slightly increased because when the confidence percentage threshold is decreased, it increased the number of patterns.

4.1.2 Pattern Results for Last 5 Shifts

The next table belongs to the last five shift patterns and is between two of the same failure codes called “all counts”, and it is calculated these pattern groups’ confidence percentages and ordered them from largest to lowest values for 1104 breakdown codes. This time, obviously, there are more pattern groups because of the last shift size. When increasing the shift size, it increased the pattern counts.

This table shows the pattern groups’ counts and confidence values for the last five shifts.

Pattern Groups	Pattern Size	All Counts	Last 5Shifts	Confidence Percentage	Sequences No
0E010004,3E010002,0E01FFE1,0E010009	4	2	2	100%	s1
0E010004,3E010002,0E01FFE1,0E010009 ,0E010008	5	2	2	100%	s1
0E010005,0E010004,0E010000,0E010001 ,0E010002	5	2	2	100%	s1
0E010005,0E010004,0E010007,0E010006 ,3E010002	5	2	2	100%	s1
0E010005,0E010004,3E010002,0E01FFE1 ,0E010009	5	2	2	100%	s1
3E010000,0E01FFE1,0E01000B,0E01000A,0E010000	5	2	2	100%	s2
3E010002,0E01000B,0E01000A,0E010009	4	2	2	100%	s3
3E010002,0E01000B,0E01000A,0E010009,0E010008	5	2	2	100%	s3
0E010000,0E010001,3E010001	3	2	2	100%	s4
0E010001,3E010001	2	2	2	100%	s4
0E010002,0E010003,0E010009,0E010008 ,0E010007	5	4	4	100%	s5
0E010003,0E010009,0E010008,0E010007	4	4	4	100%	s5
0E010003 ,0E010009 ,0E010008,0E010007 ,0E010006	5	4	4	100%	s5
0E010008,0E01000B,0E01000A ,0E01FFE0	4	2	2	100%	s6
0E010009,0E010008,0E01000B,0E01000A,0E01FFE0	5	2	2	100%	s6
0E01000A,0E010009 ,0E010008 ,0E010007	4	2	2	100%	s7
0E01000A,0E010009,0E010008,0E01000B ,0E010000	5	2	2	100%	s7
0E01000B,0E01000A,0E010009,0E010008 ,0E010007	5	2	2	100%	s7
0E01FFE0 ,0E01000A	2	2	2	100%	s8
0E01FFE0,0E01FFE1,0E010009,0E010008,3E010002	5	2	2	100%	s9
0E01FFE0 ,3E010002 ,0E01FFE1	3	2	2	100%	s10
0E01FFE1 ,0E010008 ,0E010009 ,0E010000	4	2	2	100%	s11
0E01FFE1 ,0E01000A	2	2	2	100%	s12
3E010002 ,0E010002	2	2	2	100%	s13
3E010002 ,0E010002 ,0E010003	3	2	2	100%	s13
0E010009 ,0E010008 ,0E010007	3	8	7	88%	s14
0E010009 ,0E010008 ,0E010007 ,0E010006	4	7	6	86%	s14
0E010006 ,0E01000B ,0E01000A ,0E010000	4	5	4	80%	s15
0E010006,0E01000B,0E01000A,0E010000,0E010001	5	5	4	80%	s15
0E010007,0E010006,0E01000B,0E01000A,0E010000	5	5	4	80%	s15
0E010008 ,0E010007	2	9	7	78%	s16
0E010004 ,0E010000 ,0E010001	3	4	3	75%	s17
0E010006,0E010002,0E010003 ,0E010009 ,0E010008	5	4	3	75%	s18
0E010006 ,0E010009 ,0E010008 ,0E010005	4	4	3	75%	s18
0E010006,0E010009,0E010008 ,0E010005 ,0E010004	5	4	3	75%	s18
0E010007 ,0E010005	2	4	3	75%	s19
0E010008 ,0E010007 ,0E010006	3	8	6	75%	s19
0E010008 ,0E010007 ,0E010006 ,0E010005	4	4	3	75%	s19
0E010008,0E010007,0E010006 ,0E010005 ,0E010004	5	4	3	75%	s19

0E010009,0E010008,0E010007 ,0E010006 ,0E010005	5	4	3	75%	s19
0E01000A ,0E010009 ,0E010008 ,0E01000B	4	4	3	75%	s19

*Figure 15****- Patterns for failure code 1104, comparing last five shifts and confidence percentages***

After making groups for these patterns, it is created another table (Figure 16) for these sequence groups which have a hundred percent confidence and related failure times. For example, groups of s1 occurred before the third and seventh breakdown for 1104.

Faults No	100%
f1	s8,s13
f2	s3,s4,s5,s10,s12,s13
f3	s1,s2,s5,s7,s12
f4	s8
f5	s4,s9,s11
f6	s6,s10,s11
f7	s1,s3,s6,s9

Figure 16- ***Sequence group distributions according to breakdown orders for failure code 1104 in last five shifts***

In this table, it can be seen there are seven faults for reason code 1104. However, there are 13 different sequential pattern groups because of shift sizes. Comparing the last three shift patterns, the number of patterns increased because this time it was used the last five shifts in the dataset. The next figure shows sequence group distributions as a graph.

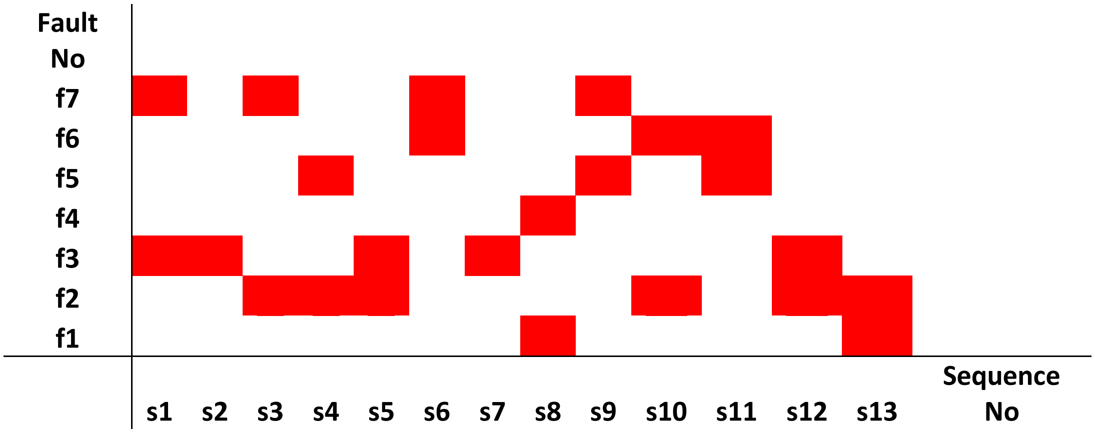


Figure 17- ***Graphical representation of sequence group distributions according to breakdown orders for failure code 1104 in last five shifts***

When calculate the detection rate which equals by dividing counts of filled rows in the table by faults occurring times. For this example, the detection rate is equal to the count of filled rows in the table, which is 7 divided by all fault counts which is 7, so it comes to a 100% detection rate. This percentage means that the system can detect 10 out of 10 breakdown situations.

The graph below shows all these patterns according to percentages of confidence:

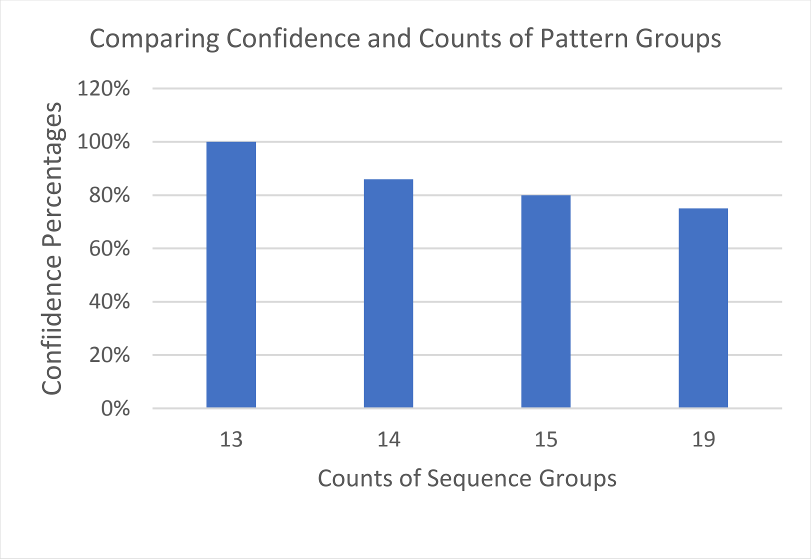


*Figure 18- **Graphical illustration of patterns and confidence for failure code 1104 in last five shifts***

As it can be seen before for the last 3 shifts, the same thing can be seen for the last 5 shifts as well. For confidence values 100% to 70%, the count of patterns slightly increased because when decreased the confidence percentages' threshold, it increased the number of patterns.

4.2 Results for Failure code 1140

It will be separated each failure code result into two parts according to the last shift numbers which are the last three and last five shifts before related machine failures occur.

4.2.1 Pattern Results for Last 3 Shifts

The next figure belongs to the last three shift patterns and is between two of the same failure codes called “all counts”, and it was calculated these pattern groups’ confidence percentages and ordered them from largest to lowest values for 1140 breakdown codes.

Pattern Groups	Pattern Size	All Counts	Last 3Shifts	Confidence Percentage	Sequences No
0201FFE0,0B010009	2	2	2	100%	s1
0201FFE0,0B010009,0B010000	3	2	2	100%	s1
0201FFE0,0B010009,0B010000,0B011519	4	2	2	100%	s1
0201FFE1,16016465,0201FFE0,0B010009	4	2	2	100%	s1
0201FFE1,16016465,0201FFE0,0B010009,0B010000	5	2	2	100%	s1
16016464,0201FFE1,16016465,0201FFE0,0B010009	5	2	2	100%	s1
16016465,0201FFE0,0B010009	3	2	2	100%	s1
16016465,0201FFE0,0B010009,0B010000	4	2	2	100%	s1
16016465,0201FFE0,0B010009,0B010000,0B011519	5	2	2	100%	s1
0B010011,160169DD,16016A41	3	2	2	100%	s2
0B010011,160169DD,16016A41,160169DC	4	2	2	100%	s2
0B010011,160169DD,16016A41,160169DC,16016A40	5	2	2	100%	s2
160169DD,16016A41,160169DC	3	3	3	100%	s2
160169DD,16016A41,160169DC,16016A40	4	3	3	100%	s2
16016A41,160169DC	2	3	3	100%	s2
16016A41,160169DC,16016A40	3	3	3	100%	s2
2010000,0B01000C	2	2	2	100%	s3
16016465,160178B5,160178B4,16016464,2010006	5	2	2	100%	s4
0201FFE1,0201FFE0	2	4	3	75%	s5
160169DC,16016A40	2	4	3	75%	s6

*Figure 19- **Patterns for failure code 1140, comparing last three shifts and confidence percentages***

Table

Graph

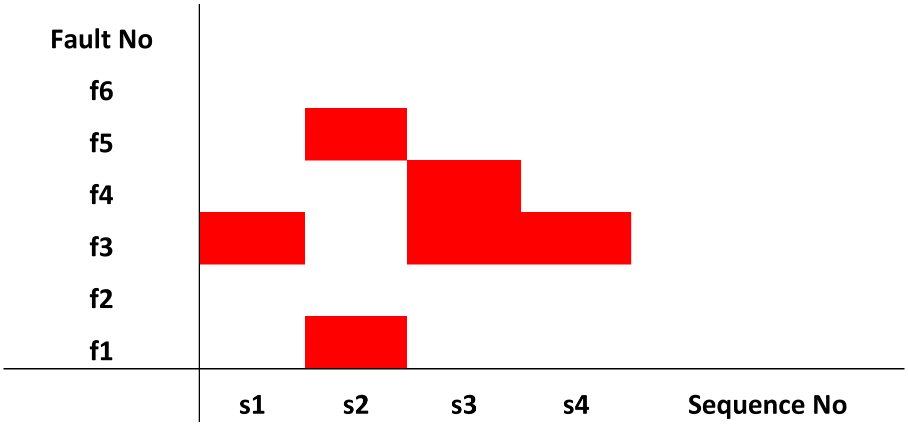


*Figure 20- **Sequence group distributions according to breakdown orders for failure code 1140 in last three shifts***

After making groups for these patterns, it will be created a table for these sequence groups which have a hundred percent confidence and related failure times. For example, groups of s1 occurred before the third breakdown for 1140.

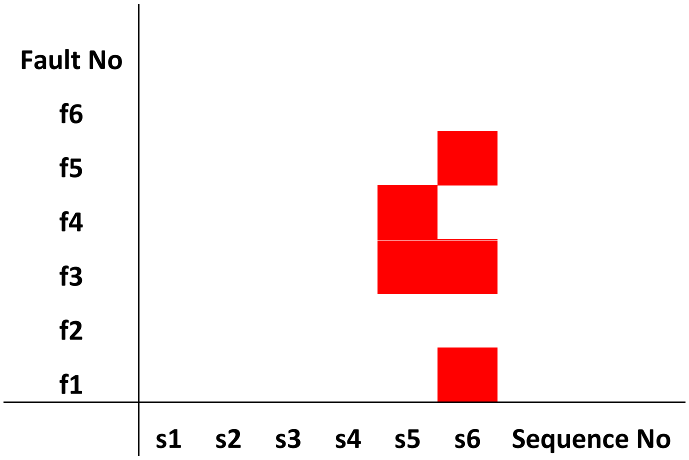
In this table, it can be seen there are six faults for reason 1140. Before failures occur, there are four different sequential pattern groups. S2 occurred before the first failure and s1, s3, and s4 occurred before the third breakdown, s3 before the fourth breakdown, and s2 again occurred before the fifth breakdowns in last three shifts. It can be seen these information as a table and graph in the Figure 20.

When calculate detection rate which equals by dividing counts of filled rows in the table by faults occurring times. For this example, detection rate equals count of filled rows in the table which is 4 dividing by total faults' counts (6), so it comes to a 66% detection rate. This percentage means that system can detect 6 out of 10 breakdowns situation.

For more accuracy, decreased the confidence value to 70%, which might increase the detection rate. The next figure shows sequence groups between 70% and 100% and a distribution of these groups as a table and graph.

FaultNo	Seq Groups
F1	s6
F2	
F3	s5,s6
F4	s5
F5	s6
F6	

Table



Graph

*Figure 21- **Sequence group distributions according to breakdown orders for failure code 1140 in last three shifts above 70%***

From this table it can be seen s5 and s6 occurred before the same failures, and it does not change the detection rate. As it can be seen from Figure 20's Table; s1, s2, s3 and s4 occurred before the same failure numbers.

The graph below of these patterns is created according to percentages of confidence:

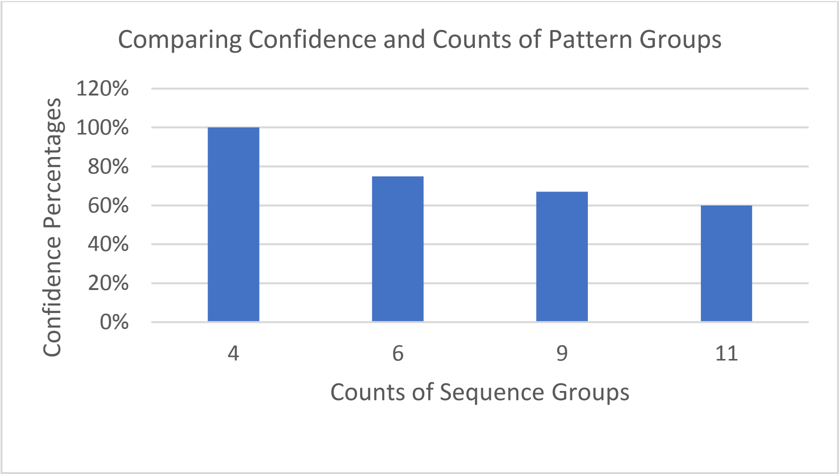


*Figure 22- **Graphical illustration of patterns and confidence for failure code 1140 in last three shifts***

For confidence values from 100% to 60%, count of patterns slightly increased because when decreased confidence percentages' threshold, it increased the number of patterns.

4.2.2 Pattern Results for Last 5 Shifts

The next table belongs to the last five shift patterns and is between two of the same failure codes called “all counts”, and it was calculated these pattern groups’ confidence percentages and ordered them from largest to lowest values for 1140 breakdown codes. This time, obviously, there are more pattern groups because of the last shift size. When increased the shift size, it increased the pattern counts.

This table shows the pattern groups’ counts and confidence values for the last five shifts.

Pattern Groups	Pattern Size	All Counts	Last 5Shifts	Confidence Percentage	Sequences No
0201FFE0 ,0B010009	2	2	2	100%	s1
0201FFE0 ,0B010009 ,0B010000	3	2	2	100%	s1
0201FFE0 ,0B010009 ,0B010000 ,0B011519	4	2	2	100%	s1
0201FFE1 ,16016465,0201FFE0 ,0B010009	4	2	2	100%	s1
0201FFE1,16016465,0201FFE0,0B010009,0B010000	5	2	2	100%	s1
16016464,0201FFE1,16016465,0201FFE0,0B010009	5	2	2	100%	s1
16016465,0201FFE0 ,0B010009	3	2	2	100%	s1
16016465,0201FFE0 ,0B010009 ,0B010000	4	2	2	100%	s1
16016465,0201FFE0,0B010009,0B010000 ,0B011519	5	2	2	100%	s1
0B010011 ,160169DD ,16016A41	3	2	2	100%	s2
0B010011 ,160169DD ,16016A41 ,160169DC	4	2	2	100%	s2
0B010011 ,160169DD ,16016A41 ,160169DC ,16016A40	5	2	2	100%	s2
160169DD ,16016A41 ,160169DC	3	3	3	100%	s2
160169DD ,16016A41 ,160169DC ,16016A40	4	3	3	100%	s2
16016A41 ,160169DC	2	3	3	100%	s2
16016A41 ,160169DC ,16016A40	3	3	3	100%	s2
2010000,0B01000C	2	2	2	100%	s3
16016465,160178B5,160178B4,16016464,2010006	5	2	2	100%	s4
0201000A,0B01000C,2010002,2010004,0B010014	5	4	4	100%	s5
0201FFE1 ,0201FFE0	2	4	4	100%	s6
070108CC ,160178B5	2	2	2	100%	s7
070108CC ,160178B5 ,160178B4	3	2	2	100%	s7
070108CD ,070108CC ,160178B5	3	2	2	100%	s7
070108CD ,070108CC ,160178B5 ,160178B4	4	2	2	100%	s7
0B011519 ,0B010000	2	2	2	100%	s8
0B010009 ,0B011519 ,0B010000	2	2	2	100%	s8
0B01000A ,0B010009 ,0B011519 ,0B010000	4	2	2	100%	s8
0B011519 ,0B01000C ,0B01000A	3	2	2	100%	s9
0B010009 ,0B011519 ,0B01000C	3	2	2	100%	s9
0B010009 ,0B011519 ,0B01000C ,0B01000A	4	2	2	100%	s9
0B01000A ,160165F4	2	2	2	100%	s10
0B01000A ,160165F4 ,1601607C	3	2	2	100%	s10
2010006,16016465,2010000,160178B5 ,2010005	5	2	2	100%	s11
160178B5 ,160178B4 ,2010005,2010006	4	2	2	100%	s12
3E010000 ,0B010002	2	4	5	80%	s13
160169DC ,16016A40	2	3	4	75%	s14
160178B4 ,2010005,2010006	3	3	4	75%	s15
2010005,0B010009 ,0B0114B5	3	3	4	75%	s16
3E010000 ,0B010002 ,0201FFE1	3	3	4	75%	s17

Figure 23- ***Patterns for failure code 1140, comparing last five shifts and confidence percentages***

After making groups for these patterns, it is created another table for these sequence groups which have a hundred percent confidence and related failure times. For example, groups of s1 occurred before the third breakdown for 1140.

Faults No	100%
F1	s2,s5,s7,s10
F2	s7
F3	s1,s3,s4,s6,s9,s11,s12
F4	s3,s6,s8,s9
F5	s2,s5
F6	s5,s10

*Figure 24- **Sequence group distributions according to breakdown orders for failure code 1140 in last five shifts***

In this table, it can be seen there are six faults for reason code 1140. However, there are 12 different sequential pattern groups because of shift sizes. Comparing the last three shifts patterns, the number of patterns increased because this time it was used the last five shifts dataset. Next figure shows sequence group distributions as a graph.

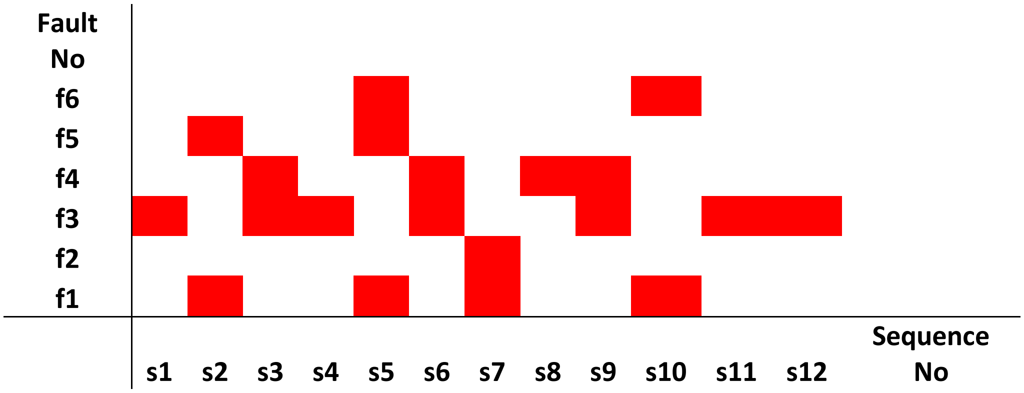


*Figure 25- **Graphical representation of sequence group distributions according to breakdown orders for failure code 1140 in last five shifts***

When calculate detection rate which equals dividing counts of filled rows in the table by faults occurring times. For this example, detection rate equals count of filled rows in the table which is 6 dividing by all faults' counts which is 6, so it equals 100% detection rate. This percentage means that system can detect 10 out of 10 breakdowns situation.

The graph below shows all these patterns according to percentages of confidence:

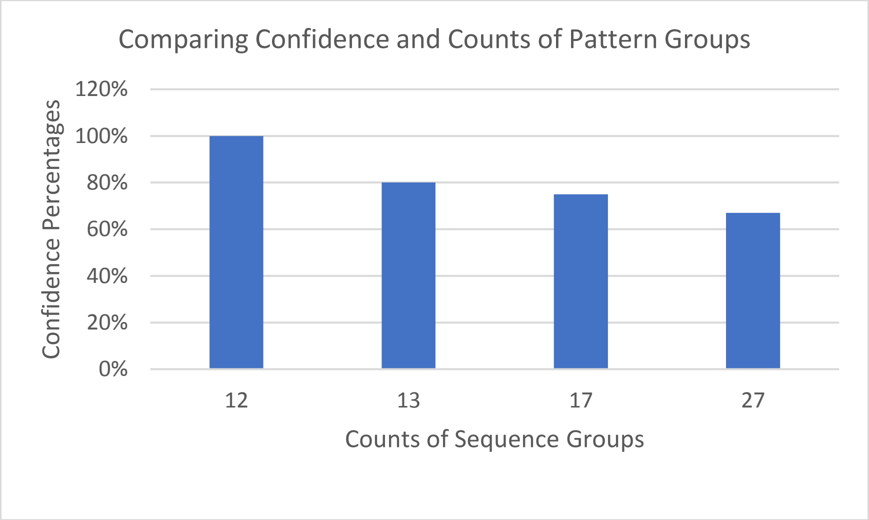


*Figure 26- **Graphical illustration of patterns and confidence for failure code 1140 in last five shifts***

For confidence values 100% to 60%, the count of patterns slightly increased because when decreased confidence percentages' threshold, it increased number of patterns.

4.3 Results for Failure code 1143

It will be separated each failure code result into two parts according to the last shift numbers which are the last three and last five shifts.

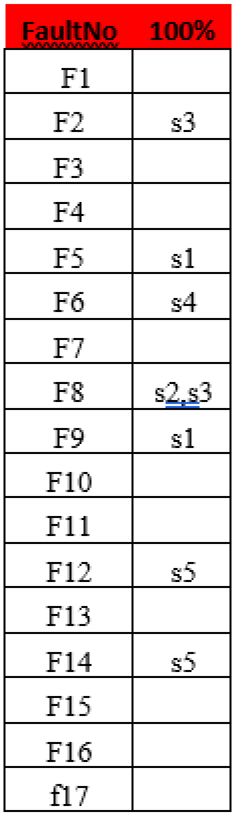
4.3.1 Pattern Results for Last 3 Shifts

The next figure belongs to the last three shift patterns and is between two of the same failure codes called "all counts", and it was calculated these pattern groups' confidence percentages and ordered them from largest to lowest values for 1143 breakdown codes.

Pattern Groups	Pattern Size	All Counts	Last 3Shifts	Confidence Percentage	Sequences No
05010ADD ,05010ADC	2	2	2	100%	S1
050114B4 ,050114B9 ,050114B8	3	2	2	100%	S2
050114B4 ,050114B9 ,050114B8 ,5010905	4	2	2	100%	S2
050114B4,050114B9,050114B8,5010905,5010904	5	2	2	100%	S2
050114B8 ,5010905	2	2	2	100%	S2
050114B8 ,5010905,5010904	3	2	2	100%	S2
050114B9 ,050114B8 ,5010905	3	2	2	100%	S2
050114B9 ,050114B8 ,5010905,5010904	4	2	2	100%	S2
050116CA ,050114B5	2	2	2	100%	S3
050119A0 ,050116CB	2	2	2	100%	S4
050119A0 ,050116CB ,050116CA	3	2	2	100%	S4
050119A1 ,050119A0 ,050116CB	3	2	2	100%	S4
050119A1 ,050119A0 ,050116CB ,050116CA	4	2	2	100%	S4
3E010002 ,3E010001 ,3E010002	3	2	2	100%	S5
050114B4 ,050114B9	2	3	2	67%	s6
050117B8 ,3E010002	2	3	2	67%	s7
050117B9 ,050117B8 ,3E010002	3	3	2	67%	s7
0501199C ,050119CF	2	3	2	67%	s8
0501199C ,050119CF ,050119CE	3	3	2	67%	s8
050119CE ,3E010002	2	3	2	67%	s9

*Figure 27- **Patterns for failure code 1143 with comparing last three shifts and confidence percentages***

After making groups for these patterns, it will be created a table for these sequence groups which have a hundred percent confidence and related faults time. For example, groups of s1 occurred before the fifth and ninth breakdowns for 1143.



FaultNo	100%
F1	
F2	s3
F3	
F4	
F5	s1
F6	s4
F7	
F8	s2,s3
F9	s1
F10	
F11	
F12	s5
F13	
F14	s5
F15	
F16	
f17	

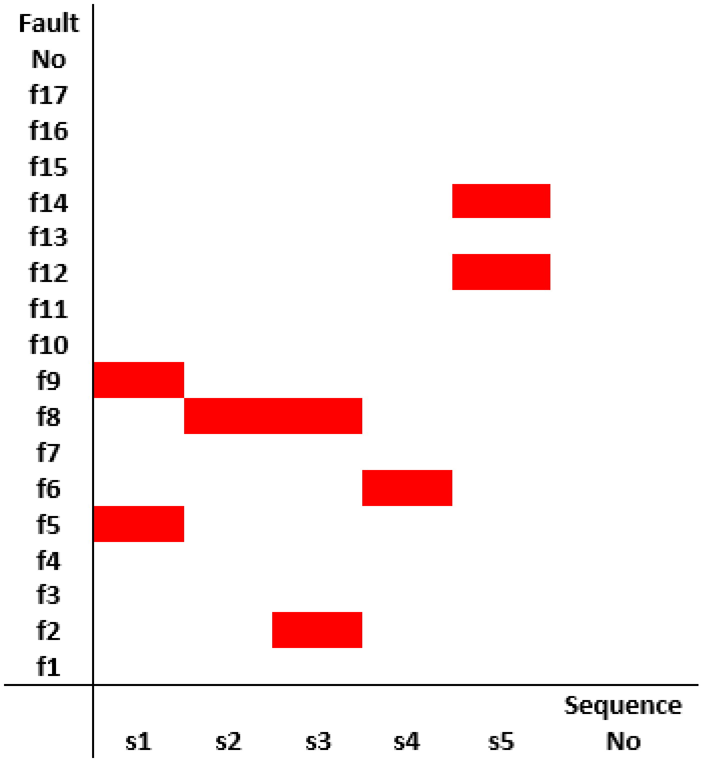


Table **Graph**

Figure 28- Sequence group distributions according to breakdown orders for failure code 1143 in last three shifts

In this table (Figure 28), it can be seen there are 17 faults for reason 1143. Before failures occur, there are four different sequential pattern groups. S3 occurred before the second failure, s1 occurred before the fifth, s4 occurred before the sixth, s2 and s3 occurred before the eighth, s1 again occurred before the ninth, and lastly s5 occurred before the twelfth and fourteenth breakdown and within the last three shifts. This is provided in the table and graph in Figure 28.

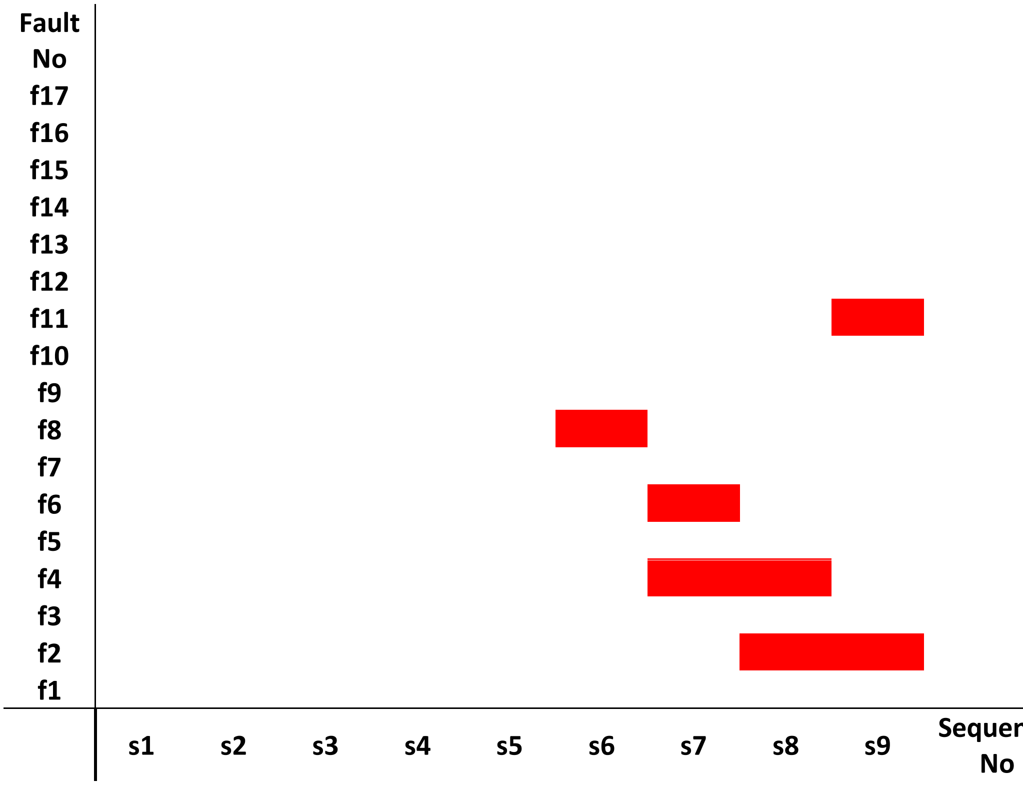
It was calculated the detection rate by dividing counts of filled rows in the table by the fault occurring times. For this example, the detection rate is equal to the count of filled rows in the table, which is 7 divided by all fault counts (17), so it comes to about 41% detection rate. This percentage means that system can detect 4 out of 10 breakdown situations.

For more accuracy, decreased the confidence value to 60%, and it might be also increased the detection rate. The next figure shows sequence groups between 60% and 100% and the distribution of these groups as a table and graph are illustrated in the next figure.

Increasing accuracy is crucial because it means catching more failures' patterns before machine breakdowns occur. However, for more accuracy it is necessary to have more wide and clean dataset. Nine months and 11 trucks information would not be enough for making exact decision, but this thesis would be a good example for predictive maintenance.

FaultNo	Above 60%
F1	
F2	s8,s9
F3	
F4	s7,s8
F5	
F6	s7
F7	
F8	s6
F9	
F10	
F11	s9
F12	
F13	
F14	
F15	
F16	
f17	

Table



Graph

*Figure 28- **Sequence group distributions according to breakdown orders for failure code 1143 in last three shifts above 60% confidence***

From this table it can be seen s7 and s8 occurred before the fourth breakdown, and s9 occurred before the eleventh breakdown. These fault numbers are different from the table in Figure 28 and increased the detection rate. The detection rate is equal to the count of filled rows in the table (9) totally, after decreasing confidence, it increased the filled rows and divided the fault counts (17), so coming to about 52% detection rate. This percentage means that the system can detect 5 out of 10 breakdown situations.

The graph below shows these all patterns according to percentages of confidence:

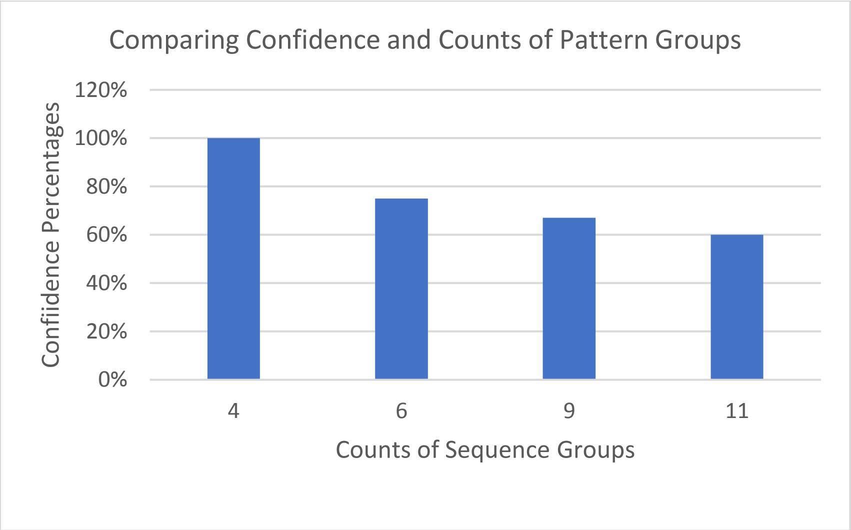


Figure 29- ***Graphical illustration of patterns and confidence for failure code 1143 in last three shifts***

As it can be seen before for last 3 shifts, same thing can be seen for last 5 shifts as well. For confidence values from 100% to 60%, count of patterns slightly increased because when decreased confidence percentages' threshold, it increased the number of patterns.

4.3.2 Pattern Results for Last 5 Shifts

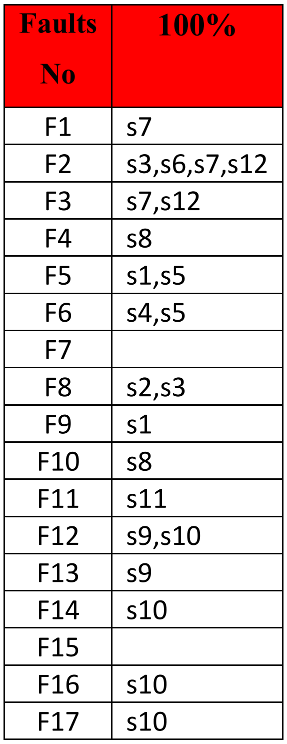
The next table belongs to the last five shift patterns and is between two of the same failure codes called “all counts”, and it was calculated these pattern groups’ confidence percentages and ordered them from largest to lowest values for 1143 breakdown codes. This time, obviously, there are more pattern groups because of the last shift size. When increased the shift size, it increased the pattern counts.

This table shows the pattern groups’ counts and confidence values for the last five shifts.

Pattern Groups	Pattern Size	All Counts	Last 5Shifts	Confidence Percentage	Sequences No
05010ADD ,05010ADC	2	2	2	100%	s1
050114B4 ,050114B9	2	3	3	100%	s2
050114B4 ,050114B9 ,050114B8	3	2	2	100%	s2
050114B4 ,050114B9 ,050114B8 ,5010905	4	2	2	100%	s2
050114B4 ,050114B9 ,050114B8 ,5010905,5010904	5	2	2	100%	s2
050114B5 ,050114B4 ,050114B9	3	2	2	100%	s2
050114B8 ,5010905	2	2	2	100%	s2
050114B8 ,5010905,5010904	3	2	2	100%	s2
050114B9 ,050114B8 ,5010905	3	2	2	100%	s2
050114B9 ,050114B8 ,5010905,5010904	4	2	2	100%	s2
050116CA ,050114B5	2	2	2	100%	s3
050119A0 ,050116CB	2	2	2	100%	s4
050119A0 ,050116CB ,050116CA	3	2	2	100%	s4
050119A1 ,050119A0 ,050116CB	3	2	2	100%	s4
050119A1 ,050119A0 ,050116CB ,050116CA	4	2	2	100%	s4
050119CA,050116CB ,050116CA ,050119CD ,050119D1	5	2	2	100%	s4
0501199B ,0501199A ,3E010002	3	2	2	100%	s5
050114B5 ,050114B9 ,050114B7 ,050114B6	4	2	2	100%	s6
050114B5 ,050114B9 ,050114B7 ,050114B6 ,050114B4	5	2	2	100%	s6
050114B9 ,050114B7 ,050114B6 ,050114B4	4	2	2	100%	s6
0501199C ,5010249	2	3	3	100%	s7
0501199C ,5010249,5010248	3	3	3	100%	s7
0501199C ,5010249,5010248,050114B9	4	2	2	100%	s7
0501199C ,5010249,5010248,050114B9 ,050114B8	5	2	2	100%	s7
0501199D ,0501199C ,5010249	3	2	2	100%	s7
0501199D ,0501199C ,5010249,5010248	4	2	2	100%	s7
050119CE ,050119CB	2	2	2	100%	s8
050119CF ,050119CE ,050119CB	3	2	2	100%	s8
3E010001 ,050119CB ,050119CA ,0501177F	4	2	2	100%	s9
3E010001 ,050119CB ,050119CA ,0501177F ,0501177E	5	2	2	100%	s9
3E010002 ,3E010001 ,3E010002	3	4	4	100%	s10
3E010002 ,3E010001 ,3E010000 ,0501177F	4	2	2	100%	s11
3E010002 ,3E010001 ,3E010000 ,0501177F ,0501177E	5	2	2	100%	s11
5010248,050114B9 ,050114B8	3	2	2	100%	s12
5010249,5010248,050114B9 ,050114B8	4	2	2	100%	s12
050114B9 ,050114B7 ,050114B6	3	3	4	75%	s13
0501015A ,050119CB	2	3	4	75%	s14

*Figure 30- **Patterns for failure code 1143, comparing last five shifts and confidence percentages***

After making groups for these patterns, it was created another table for these sequence groups which have a hundred percent confidence and related failure times. For example, groups of s1 occurred before the fifth and ninth breakdown for 1143.



Faults No	100%
F1	s7
F2	s3,s6,s7,s12
F3	s7,s12
F4	s8
F5	s1,s5
F6	s4,s5
F7	
F8	s2,s3
F9	s1
F10	s8
F11	s11
F12	s9,s10
F13	s9
F14	s10
F15	
F16	s10
F17	s10

*Figure 31- **Sequence group distributions according to breakdown orders for failure code 1143 in last five shifts***

In this table, it can be seen there are 17 faults for reason code 1104. However, there are 12 different sequential pattern groups because of shift sizes. Comparing the last three shifts patterns, the number of patterns increased, because this time it was used the last five shifts dataset. Next figure shows sequence group distributions as a graph.

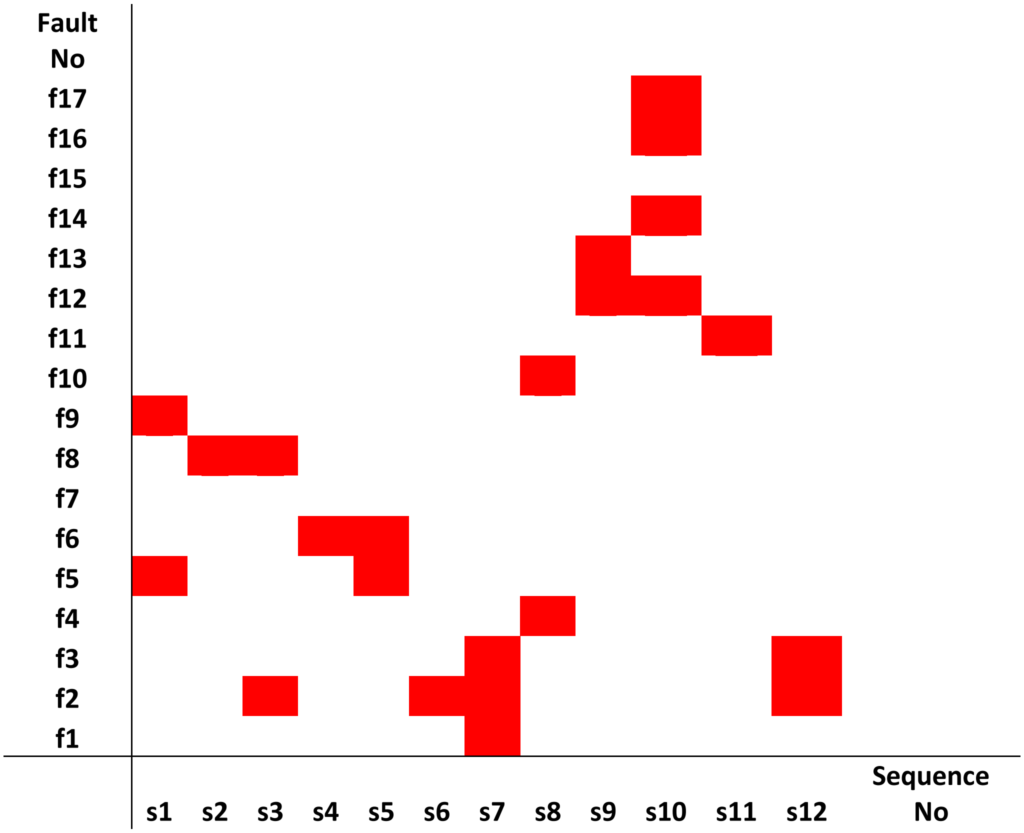


*Figure 32- **Graphical representation of sequence group distributions according to breakdown orders for failure code 1143 in last five shifts***

When calculate detection rate which equals dividing counts of filled rows in the table by faults occurring times. For this example, detection rate equals count of filled rows in the table which is 15 dividing by all faults' counts which is 17, so it equals about 90% detection rate. This percentage means that system can detect 9 out of 10 breakdowns situation.

The graph below shows all these patterns according to percentages of confidence:

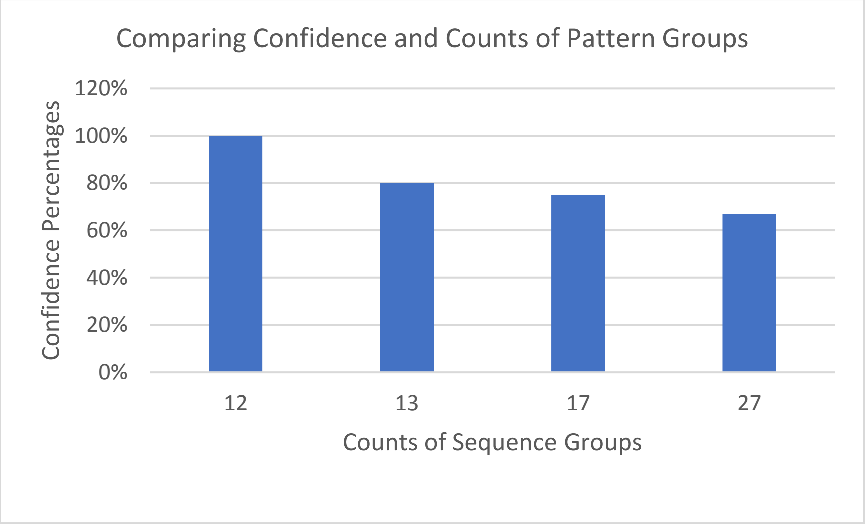


Figure 33- ***Graphical illustration of patterns and confidence for failure code 1143 in last five shifts***

For confidence values 100% to 70% count of patterns slightly increased because when decrease confidence percentages' threshold, it increases number of patterns.

5 CONCLUSION

Nowadays, with the increasing presence of technological devices, every company tries to follow their vehicles, especially those that have expensive prices and maintenance costs. The mining industry has several expensive vehicles, and companies record their statuses regularly. In this study, there is one big dataset belonging to one North American mining company. The dataset includes records from eleven trucks for nine months, coming to more than three million rows, and more than a hundred columns.

It is focused on the sequences of alarms and signals, which might be related to reasons for breakdowns. However, there are several other features which may have an impact on the breakdown of machines, such as driver mistakes, machine loading status, speed, or road conditions. For example, there are several threshold speed limits for mining trucks depending on road slopes, and the dataset does not show these kinds of thresholds. Another example is the various loading rules for these mining trucks. When trucks are loaded more than load limits allow, it may have an impact on the machine's health status, but there are not this kind of information and did not calculate for those factors. These features need more in-depth research and expert knowledge for a future research project.

In this thesis, it was selected three various kinds of breakdown codes which cause more expensive maintenance costs and mechanical failures. Mechanical breakdown reasons are more related to alarms and signals than other machine failures' reasons. It was selected three of them, but for more accurate results, it must implement more than three breakdown reasons. According to the three machine failure codes, there are specific relationships between events and breakdown codes. First, it was discovered several patterns between two same failure codes and counted them. Afterwards, it was discovered various patterns in the last three and five shifts before breakdowns occur. Lastly, it was calculated confidence values for the last three and five shifts and illustrated them in tables and graphs.

The results showed that when implement the sequential pattern algorithm before machine failures, it is possible to discover several patterns which may indicate breakdowns. However, for more accurate results, it is necessary to have cleaner and larger datasets, and additionally, time records beyond nine months. Despite these missing values, results indicate a detection rate of more than 90% in the last five shift events, which shows several specific and identifiable groups of patterns before machine breakdowns occur. However, the results do not show high detection rates for the last three shifts' alarms and signals before machine breakdowns occur. For future work, a more wide and clean dataset would be more accurate in discovering mining trucks failure reasons.

REFERENCES

Adolfsson, E., Dahlström, T. (2011). Efficiency in corrective maintenance (Master's Thesis), Department of Technology Management and Economics Division of Logistics and Transportation CHALMERS UNIVERSITY OF TECHNOLOGY Gothenburg, Sweden, Report No. E2011:096

Bastos, P., Lopes, I., & Pires, L. (2013). Application of data mining in a maintenance system for failure prediction. *Safety, Reliability and Risk Analysis,* 933-940. doi:10.1201/b15938-138

Bastos, P., Lopes, R., & Pires, L. (2012). A Maintenance Prediction System using Data Mining Techniques. *2012 Proceedings of the World Congress on Engineering*.

Cartella, F., Lemeire, J., Dimiccoli, L., & Sahli, H. (2015). Hidden Semi-Markov Models for Predictive Maintenance. *Mathematical Problems in Engineering, 2015*, 1-23. doi:10.1155/2015/278120

Chen, L., Han, J., Lei, W., Cui, Y., & Guan, Z. (2016). Full-Vector Signal Acquisition and Information Fusion for the Fault Prediction. *International Journal of Rotating Machinery, 2016*, 1-7. doi:10.1155/2016/5980802

Chueh, H. (2010). Mining Target-Oriented Sequential Patterns With Time-Intervals. *International Journal of Computer Science and Information Technology, 2*(4), 113-123. doi:10.5121/ijcsit.2010.2410

Esmaeili, M., & Fazekas, G. (2010). Finding Sequential Patterns from Large Sequence Data. *International Journal of Computer Science* Issues, (7).

Fekete J.A. (2015). Big Data in Mining Operations (Master's Thesis), MSc Business Administration and Information Systems(IT Management and Business Economics), Copenhagen Business School, Copenhagen, Denmark.

Kantardzic, M. (2011). *Data Mining: Concepts, Models, Methods, and Algorithms*. Wiley-IEEE Press.

Liu, B. (2011). Association Rules and Sequential Patterns. In: Web Data Mining. Data-Centric Systems and Applications. doi:https://doi.org/10.1007/978-3-642-19460-3_2

Marinelli, M., Lambropoulos, S., & Petroutsatou, K. (2014). Earthmoving trucks condition level prediction using neural networks. *Journal of Quality in Maintenance Engineering, 20*(2), 182-192. doi:10.1108/jqme-09-2012-0031

Murakami, T., Saigo, T., Ohkura, Y., Okawa, Y., & Taninaga, T. (2002). Development of Vehicle Health Monitoring System (VHMS/WebCARE) for Large-Sized Construction Machine (Vol. 48, Tech.). Komatsu.

Peng, S., & Vayenas, N. (2014). Maintainability Analysis of Underground Mining Equipment Using Genetic Algorithms: Case Studies with an LHD Vehicle. *Journal of Mining, 2014*, 1-10. doi:10.1155/2014/528414

Rouse, M., (2014). What is machine data? - Definition from WhatIs.com. (n.d.). Retrieved March 28, 2018, from http://internetofthingsagenda.techtarget.com/definition/machine-data

Slimani, T., & Lazzez, A. (2014). SEQUENTIAL MINING: PATTERNS AND ALGORITHMS ANALYSIS. International Journal of Information Technology and Computer Science, 6(3). doi:10.5815/ijitcs.2014.03.09

Ubaidulla, D., Sushmitha, B. S., & Vanitha, T. (2017). A STUDYONMINING SEQUENTIAL PATTERN IN TIME SERIES DATA. International Journal of Latest Trends in Engineering and Technology. Retrieved from https://www.ijltet.org/journal/151065801484.pdf.

Ullah, I., Yang, F., Khan, R., Liu, L., Yang, H., Gao, B., & Sun, K. (2017). Predictive Maintenance of Power Substation Equipment by Infrared Thermography Using a Machine-Learning Approach. *Energies, 10*(12), 1987. doi:10.3390/en10121987

US Census Bureau. (2017, May 03). Library. Retrieved March 28, 2018, from https://www.census.gov/library/publications/2017/econ/2017-csr.html

Viger, P.F. (2017). An Introduction to Sequential Pattern Mining. (2018, January 02). Retrieved March 28, 2018, from http://data-mining.philippe-fournier-viger.com/introduction-sequential-pattern-mining/

Yildirim, M., & Dessureault, S. (2007). Truck assignment performance evaluation by using data mining techniques. *Data Mining Mine Data: Truck-shovel Fleet Management Systems*. doi:10.1201/9781439833407.ch12

Zhao, Q., & Bhowmick, S. S. (2003). Sequential Pattern Mining: A Survey. *Technical Report, CAIS, Nanyang Technological University, Singapore, No. 2003118*.

www.highservice.com , FleetSafety® – Highservice. (n.d.). Retrieved April 07, 2018, from http://www.highservice.com/highservice/en/highservice-technology-eng/fleetsafety-collision-avoidance-system/

www.uky.edu, Coal Mining, Kentucky Geological Survey. (n.d.). Retrieved April 04, 2018, from http://www.uky.edu/KGS/coal/coal-mining.php

www.modularmining.com , Modular Mining Systems | Mine Management Solutions. (n.d.). Retrieved April 02, 2018, from http://www.modularmining.com/wp-content/uploads/Mining-Magazine__Healthy-and-Wise.pdf

https://www.emaint.com , What is Predictive Maintenance & How to Get Started. (2017, December 19). Retrieved April 02, 2018, from https://www.emaint.com/what-is-predictive-maintenance/

www.generalkinematics.com , A Brief History of Mining: The Advancement of Mining Techniques and Technology. (2015, December 22). Retrieved March 28, 2018, from https://www.generalkinematics.com/blog/a-brief-history-of-mining-and-the-advancement-of-mining-technology/

www.revolvy.com , "Preventive maintenance" on Revolvy.com. (n.d.). Retrieved March 28, 2018, from https://www.revolvy.com/main/index.php?s=Preventive maintenance

www.ni.com , The Cost of Mining Equipment Mismanagement. (n.d.). Retrieved March 28, 2018, from http://www.ni.com/white-paper/52608/en/

www.cat.com , Mining Trucks Mining Trucks. (n.d.). Retrieved March 28, 2018, from https://www.cat.com/en_US/products/new/equipment/off-highway-trucks/mining-trucks.html

CURRICULUM VITAE

NAME: Abdulgani Kahraman

ADDRESS: Department of Computer Engineering and Computer Science
University of Louisville
Louisville, KY 40292

EDUCATION: M.S. Computer Science
University of Louisville
2015-2017
B.S. Computer Engineering
Sakarya University
2007-2011

EXPERIENCE: Software Developer
Innova IT Solutions, Istanbul, Turkey
2011-2014

AWARDS: Study Abroad Scholarship
Ministry of Education, Turkey
2014