Maintenance Method of Logistics Vehicle Based on Data Science and Quality

Yu Ping Peng[1], Shih Chieh Cheng[2](✉), Yu Tsung Huang[2], and Jun Der Leu[2]

[1] Department of Logistics Management, National Defense University, Taoyuan City, Taiwan
jason@taivs.tp.edu.com

[2] Department of Business Administration, National Central University, Taoyuan City, Taiwan
leujunder@mgt.ncu.edu.tw

Abstract. With the changes in consumption around the world, the global logistics and logistics management has been developed, which has derived the business opportunities of freight logistics and the demand for vehicles, which has led to the increase of vehicles service and service parts . Therefore, effective remaining vehicle readiness and reduction in maintain costs have become an urging subject to be solved in the industry of today. However, as in the era of big data, it is important that the enterprise makes good use of data and information to save costs, increase revenue, and ensure competitive advantages. But if we could not ensure the quality of the data, it would easily lead the analysis to the wrong decisions. Therefore, this study is based on the predictive maintenance, taking the condition of data quality considerations, and using algorithms to construct a decision support model, and proposing optimal replacement cycles and rules for vehicle components, and analyzing the impact on brands and maintenance amounts. Therefore, this study is based on maintenance history, through systematic and manual analysis, to obtain good quality data, and then use chi-square test and algorithm analysis to establish a classification model for decision support. The research department analyzes and analyzes the 3.5 ton freight vehicle maintenance and repair history of a case company from 2008–2016. After the data is cleaned and sorted, it obtains 173,693 work orders and good data quality data for 23 types of maintenance items. And the results show that: the costs contains significant divergence among brands; service parts damage is related to particular environment; we can obtain appropriate service period through proper classification rules. The decision support model constructed by this study will be improved and integrated with the actual needs of the industry on the premise of taking into account the quality of data.

Keywords: Big data · Data quality · Vehicle predictive maintenance

1 Introduction

In recent years, the popularity of Internet has led to the changes of consumption and the booming of e-commerce, and one of the main causes is commodity delivery which has derived the enormous business opportunity of freight logistics and the demand for vehicles, then rise to huge business opportunities in the logistics and freight and demand,

Y. Nah et al. (Eds.): DASFAA 2020 Workshops, LNCS 12115, pp. 131–145, 2020.
https://doi.org/10.1007/978-3-030-59413-8_11

which has gradually become one of the key development projects of industry in many countries [10, 16]. As being in the era of big data, the data and information has been exploded growth of every enterprise. Just as the logistics industry has increased its demand for vehicles as its business has grown, the maintenance of its vehicles and the wear and tear of its components have also grown simultaneously. For example, the Hsinchu Transport company has spent nearly 8 million every month in 2008 for each type of vehicle maintenance costs [33]. As the core tool of urban logistics is a vehicle, reducing vehicle maintenance costs and extending the service life are important issues that the industry needs to resolve. It is a huge and complicate data of vehicle maintenance, and it is an Operation Focus of an enterprise that how to make sure the correction and quality of data and mining, analyzing the information and knowledge behind the data [22] to ensure the vehicle has proper maintenance to maintain the work effectiveness. Because of the source and quantity of vehicle history and maintenance data were generated from different departments in enterprise, so we need effective methods to extract useful information from the complicated and huge data; data mining technology, which has been widely used in analysis of large amounts of data, is especially suitable for knowledge mining work [6]. And in the relevant researches of logistics transportation, and the maintenance and replacement rules in different fields were rarely discussed. With fewer systems and manual methods, after the data is sorted and cleaned, the data with better data quality will be analyzed and calculated.

Based on this, the purpose of this research is to verify the quality of the data by using the system and manual methods of the company's vehicle component repair and maintenance data. Then use good data to verify whether there are differences in maintenance costs between vehicle attributes and brands. And by understanding the rules of maintenance and replacement in different factories and regions, the diagnosis and decision-making of predictive maintenance of vehicle components is put forward. In order to prevent the breakdown of the vehicle during the operation, it can obtain the optimal maintenance timing. The research framework is as follows: in Sect. 2, the data quality, data mining and forms of maintenance and other literature will be described; Sect. 3 is about case study and methodology; Sect. 4 is about the data quality validation and description of the analysis of each item of data; Sect. 5 is about the discussion about the aforementioned validation; in last section Sect. 6, the findings and recommendations obtained in this study will be described.

2 Literature Review

2.1 Data Quality and Data Mining

In the age of wisdom economy and big data, data has become an important source of growth and development of the government and enterprises, but too much data cannot be used effectively but become a hea*vy burden for enterprises and organizations.* Therefore, to convert data into useful information and knowledge as a basis for decision support reference has been apparently become an important direction of business decisions at the present time [9]. However, big data sources have a wide range of complex structures. Data often comes from the interaction and integration of different systems and periods in the organization; Therefore, the acquired data may have related quality problems such as data error, information loss, inconsistency, and noise. Therefore, data cleaning has become one of the important methods for obtaining high-quality data. The most important purpose of the so-called data cleaning or quality control is to detect and eliminate errors, inapplicability and inconsistencies in the data in order to improve its data quality [22]. However, this method often deletes 20% to 40% of the data. Therefore, in order to save data, organizations or enterprises often supplement the missing value in the original data directly through long-term data trends. In practice, this approach may increase the uncertainty of the data, and even lead to the use of problematic data or replacement, which may cause doubts about the gap [21]. Therefore, the way data is cleaned is even more important. Wang et al. [32] proposed that data cleansing can be divided into four modes: manual execution, writing special applications, data cleansing that has nothing to do with specific application areas, and solving one type of problem. This study uses manual data cleanup and other methods to clean up the data.

The good data quality data obtained after cleaning can be used for data mining. Fayyad et al. [15] proposed the so-called data mining means to have a data analysis through algorithms, and then to find out a specific pattern or model, which is also the procedure how to discover knowledge from data base. Berry and Linoff et al. [5] believed that data mining is to analyze a great amount of data in automated or half-automated ways and then from which to find out meaningful relationships or rules. To sum up, the so-called data mining is a method to derive a way of obtaining decision-making relevant information by analyzing a large number of data [27]. And today the data mining has been widely used in manufacturing, finance, service, medical and marketing and other different industries. But when it is used in the field of logistics, such as the aforementioned, Only a few researches focused on relevant issues about the maintenance of vehicles. Therefore, this study is to discuss the implementation of the logistics methods of vehicle maintenance in freight logistics and mainly to build strategic management measures for the maintenance of vehicle parts so as to get the MTBF (mean time between failures) which is anticipated to provide vehicle maintenance and management personnel with the appropriate timing of maintenance and to help establish the foundation supporting decision-making for the predicted replacement of parts.

2.2 Maintenance Types

In recent years, issues about the maintenance have been widely discussed from various fields. An appropriate maintenance has been playing an important role in work performance and cost-effectiveness of the organization, because under the conditions of increasing the reliability of equipment, to ensure normal manufacturing system processes will help enterprise's overall operating performance [31, 34]. Therefore, proper and timely maintenance is often regarded as key conditions for enterprises to keep their competitive advantage, and different forms of maintenance have become key factors affecting the effectiveness of the maintenance of organization. In general, there are four types of maintenance: corrective maintenance, preventive maintenance, predictive maintenance and proactive maintenance [12]. In the past, companies often used preventive maintenance to reduce the costs of inspection and maintenance, but if not taking into account the current state of the wear and tear of equipment components, this maintenance could easily lead to burden and waste associated with company's business costs [34], Bousdekis et al. [9] also mentioned that the implementation of preventive maintenance could result in approximately 60% of equipment which have been replaced too early, which may lead to the increase on business costs. Therefore, with the development of industry along with big data, networking and the innovation of information communication technology, the maintenance has been gradually adjusted to predictive maintenance. And with the effective predictive maintenance strategy, the results can be used as the basis for the planning of preventive maintenance period. Thus, this maintenance has been widely promoted and applied in recent years [28]. To sum up, this study focused mainly on preventive maintenance to spot out the risk of failure through model and analysis, and then found out a maintenance period and manner more cost effective and strengthened and to improve the maintenance strategy to achieve business goals with limited maintenance resources.

3 Research Method

3.1 Conceptual Framework

Based on the foregoing content and related literature, this study conducts research on the optimal maintenance replacement cycle for logistics vehicles. First, sort and sort data and clean up the cargo vehicle properties and component maintenance history of the case company, analyze and compare the differences before and after, and select appropriate data quality data to meet the application and scope. Then carry out narrative statistics and chi-square test, and verify whether there is a significant difference in maintenance costs for vehicles of the same property under different circumstances. Subsequently, a model was established using SPSS software, and a decision support model based on a decision tree was established to optimize the replacement cycle. Finally, put forward conclusions and recommendations based on empirical results. See the conceptual framework in Fig. 1.

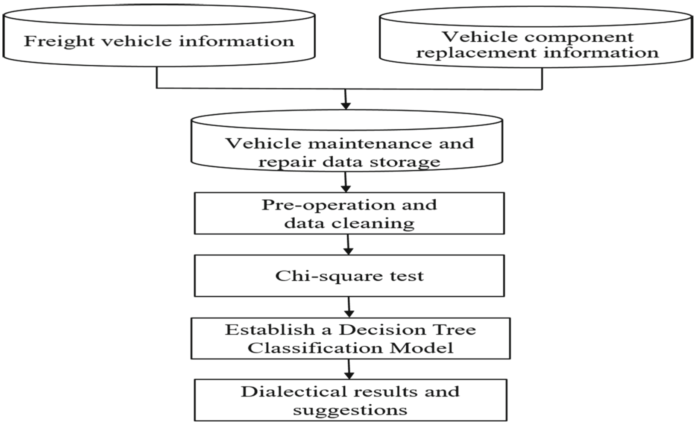

Fig. 1. Conceptual framework

3.2 Research Object

The case company in this research has been established for over 60 years and its main business is to provide Taiwan (including Islands) with cargo transportation services, which is currently one of the top five companies in the logistics industry in Taiwan. At the aspect of the business, this case company has set up satellite operating stations in various regions in order to facilitate the service areas of B2B and B2C. At the aspect of freight delivery, the categories of which include frozen products, frozen goods, food, household appliances, books, daily necessities, clothing and other diversification of commodities. This company currently used the 3.5-ton vehicle as main operating mode in B2C business model; therefore, it is much longer time when the vehicles are assigned and used are mostly need to drive between different factories and regions. Therefore, not just daily routes for vehicles are required to adjusted and planned, but also the degree of wear and tear of the vehicle parts and their replacement are much more frequent. Based on all above, this research aimed a total of 750 3.5-ton level freight vehicles of the case company's logistics services from 2008 to 2016 (total nine years), A 3.5-ton freight car used by the case company's logistics service, with a total of 750 cars (divided into four brands, represented by CT1, CT2, CT3, and CT4) was used as the research and analysis object. The original vehicle resume and maintenance information totaled 190,053. However, in order to seek data quality, quality and efficiency, the maintenance and maintenance details were processed, converted, classified, and normalized by systematic and manual methods. Analysis of quality data. Then, the decision of the optimal replacement cycle of vehicle components is formulated to obtain the appropriate maintenance timing. According to the actual business types, the factory areas of the case company were classified into five types: the coastal area, the industrial area, the metropolitan area, the agricultural area and the comprehensive area. Besides, taking the different geographical environment of the cases into accounts, vehicle parts would be worn to different degree, so this study classified the areas of business stations of the case company into the north, the central, the southern, the eastern and outlying islands areas according to the classification used by the regional meteorological observations and prediction.

4 Data Analysis and Validation

4.1 Data Cleaning and Quality Analysis

This research is based on 3.5 tons of freight vehicles used in the urban logistics in this study. A total of 190,053 vehicle resumes and historical maintenance and repair original data were analyzed. The detailed data sheet records the vehicle registration information actually repaired by each work order. The classified contents 15 types of information such as the area, the factory area, and the name of the business office, the license plate number, the date of issue, the model name, the brand name, the type, year type, tonnage level, current vehicle condition, maintenance date, vehicle components, quantity and amount, …etc. The content is to be compatible with the data cleaning and subsequent analysis. It is divided into four columns: complete, inconsistent (including format, name, description method, etc.), missing values, and other (including content errors, incorrect fields). After a unified analysis, the data integrity rate is between 71%

and 100% (see Table 1). As shown in the Table 1, inconsistent data (8.99%), missing values (0.02%), and other items (0.03%) account for about 9% of the total data; If the original data was directly brought into the algorithm analysis, it would not possible to establish effective models and classification rules due to incomplete data quality. If the direct deletion method if is used, we will aggregate 32% of parameter (61,428 data) that contains one of following incompleteness: inconsistencies, missing values, and other. Although with the direct deletion method applied and obtained high-quality data (only 68% of parameter), but it also caused large loss of historical data that could result in subsequent analysis and modeling distortion, and furthermore it may result in rules and decisions error.

Table 1. Vehicle maintenance and repair breakdown classification table (original)

Field	Information content (number of entries)				Number of data	Data integrity rate
	complete	Inconsistent	Mission value	others		
Office area	147,580	41,878	251	344	190,053	78%
Plant of Business Office	141,367	48,330	117	239		74%
Office name	159,644	30,409	42	97		84%
License number	180,693	9,266	9	85		95%
Date of issue	188,795	1,258	0	0		99%
Vehicle name	190,053	0	0	0		100%
brand name	187,901	2,135	0	17		99%
Type	182,419	7,538	61	35		96%
Year	149,565	40,488	0	0		79%
Tonnage level	190,053	0	0	0		100%
Car condition	190,053	0	0	0		100%
Maintenance date	169,880	20,173	0	0		89%
Vehicle components	135,280	54,773	0	0		71%
Quantity	190,053	0	0	0		100%
Amount	190,053	0	0	0		100%
Percentage of total	90.97%	8.99%	0.02%	0.03%	100%	

Therefore, in addition to using system analysis, this study attempts to readjust the data generated by the organization in different periods and systems through manual operations; First, filter and sort incomplete vehicle maintenance data, then proceed with data aggregation, conversion, normalization, definitely classify, and define. At the same time, the classification level of large, medium and small items of vehicle components is established in order to obtain better data quality data. The operation rules are summarized as follows:

- In terms of location, pre-adjustment was made based on the final division of business office area, plant area and name.
- Terms of location: pre-adjustment was made based on the final division of business office area, plant area and name.

- Terms of vehicles: the format such as year, type and number is consistent with the content name.
- Terms of maintenance: in addition to the unified date format, in addition to the names of vehicle components, experts recommend using similar and synonymous names for pre-processing of data.
- In connection with the sort of missing value: directly delete to avoid the doubt of data gap.

After repeated experiments in this study, it was found that the classification model constructed was the best in terms of the quality and efficiency of the data collected at the large project level of vehicle components. It took one month to the manual process, and that is 240 man hours, if we count 8 h per day to implement data quality improvement. Therefore, a total of 173,693 maintenance work orders (93.14%) and 23 types of vehicle component items were finally summarized and sorted for the subsequent model construction and system analysis (see Table 2).

Table 2. Vehicle maintenance and repair breakdown table (revised)

Field	Minimum value	Max	average value	Standard deviation	Number of categories	Number of data
Office area	--	--	--	--	5	173,693
Plant of Business Office	--	--	--	--	5	
Office name	--	--	--	--	48	
License number	--	--	--	--	750	
Date of issue	640802	980902	934273	19370	--	
Vehicle name	--	--	--	--	1	
brand name	--	--	--	--	4	
Type	--	--	--	--	12	
Year	2008	2016	--	--	--	
Tonnage level	--	--	--	--	1	
Car condition	--	--	--	--	4	
Maintenance date	970815	1050518	--	--	--	
Vehicle components	--	--	--	--	23	
Quantity	0	851	2	15	--	
Amount	1	159,744	559	2,055	--	

4.2 Analysis of Each Vehicle Brand and the Maintenance Costs

At first, we used chi-squared test to see whether there are any significant differences of the maintenance costs between the vehicles with the same attribute but different brands. According to the five maintenance costs area as shown in Table 3, the costs of each vehicle regardless of its brand have concentrated on the range from \$176~\$335 and among them, CT3 has taken than the highest proportion(45.5%), CT1(31.0%) the second, CT2(30.7%) the third and CT4 (26.3%) has taken the lowest. The second highest proportion of range of maintenance costs of each brand is respectively from \$336~\$619

(CT1, 26.0%; CT3, 24.2%), and $ 517 or more (CT2, 23.3%) and from $120~$175 (CT4, 26.0%). According to the chi-squared test, some significant differences in the maintenance costs existed between the vehicles with the same attribute but with different brands.

Table 3. Data analysis statistical table of vehicle brands and their respective maintenance costs

Amount / Brand	$0～$199 (I)		$120～$175 (II)		$176～$335 (III)		$336～$619 (IV)		$620 or more (V)		Total
CT1	12925	12.90%	16655	16.70%	30947	31%	18819	18.80%	20560	20.60%	99906
CT2	9061	12.30%	10974	14.90%	22619	30.70%	17157	23.30%	13839	18.80%	73650
CT3	8	8.10%	7	7.10%	45	45.50%	15	15.20%	24	24.10%	99
CT4	7	18.40%	9	23.70%	10	26.30%	6	15.80%	6	15.80%	38
Total	22001(13%)		27645(16%)		53621(31%)		35997(21%)		34429(20%)		173693

4.3 Each Factory Replacement Vehicle Parts Analysis

The data mining technology was used in this research to have an inquiry in the optimum replacement cycle for vehicle parts in particular factory areas of regions. At first, a statistical data was made from the replacement data of the vehicle parts in each factory, and after the analysis, the top 5 of in each factory area have accounted for the ratio of the overall component for maintenance and replacement more than 68%, in which coastal areas have the highest 93%, and the following are comprehensive areas 88%, agricultural areas 79% industrial areas 74% and metropolitan areas 68%. Among these areas, the item of the most common maintenance in coastal areas is the switch the wiring which accounts for 31% (3,710 times); in metropolitan and comprehensive areas are general maintenance which reaches 39% (23,809 data) and 25% (13,098 data); in industrial and agricultural areas are general repair which reaches 32% (5,800 data) and 27% (4,361 data). The replacement of vehicle parts is as shown in Table 4.

Different areas may lead to differences in the usage of logistics vehicles, so this study has also taken Taiwan's geographical characteristics into consideration (divided into northern, central, southern, eastern and outlying island regions), in order to have an inquiry in the relation of the replacement for vehicle parts between different vehicle brands and different areas. Based on all above, in the study we have found that (as shown in Table 5), in coastal, metropolitan and industrial areas, these two brands CT1 and CT2 in the northern, central and southern regions have accounted for the higher proportion of the maintenance; in agricultural areas, these two brands CT1 and CT2 in the central, southern and eastern regions have accounted for the higher proportion of the maintenance; in the comprehensive areas, these two brands CT1 and CT2 have accounted for the higher proportion of the maintenance in the central and the southern regions. To sum up, although the two brands CT1 and CT2 have always accounted for the highest proportion of the times of vehicle maintenance, there are partial differences when they are in different regions.

Table 4. Table of replacement times of vehicle parts in each factory area (Top 5)

Item \ Area	Coastal Areas	Metropolitan Areas	Industrial Areas	Agricultural Areas	Comprehensive Areas
Vehicle parts	Wiring switch 31%(3,710)	General maintenance 25%(23,809)	General repair 27%(5,800)	General repair 32%(4,361)	General maintenance 39%(13,098)
	General repair 29%(3,421)	General repair 24%(22,757)	Fill light system 25%(4,610)	General maintenance 25%(4,087)	General repair 19%(6,434)
	General maintenance 18%(2,079)	Air conditioning repair 7%(6,694)	Intake and exhaust system 8%(1,491)	Three-stage disassembly 15%(2,405)	Three-stage disassembly 15%(5,060)
	Automotive battery 12%(1,366)	Fill light system 6%(5,668)	Three-stage disassembly 5%(928)	Compartments maintenance 7%(1,152)	Air conditioning repair 11%(3,541)
	Sheet metal 4%(481)	Radiator 5%(5,004)	Clutch 4%(804)	Generator 5%(767)	Front steel plate 3%(1,069)
Total	93% 11,057	68% 69,932	74% 13,633	79% 12,772	88% 29,202

*Remark: the percentage is the number of maintenance times of the vehicle component item in the factory area/the total number of maintenance times of vehicle parts in the factory area; the number in ( ) is the number of maintenance times of the vehicle component item.

Table 5. Maintenance statistical analysis table of vehicle brands in different areas

Factory area	Brand \ Region	Northern	Central	Southern	Eastern	Outlying Islands
Coastal area	CT1	53.26%	58.45%	40.14%	0	30.67%
	CT2	46.70%	41.55%	59.86%	0	69.33%
	CT3	0.01%	0	0	0	0
	CT4	0.03%	0	0	0	0
Metropolitan area	CT1	57.78%	74.70%	52.23%	0	0
	CT2	42.12%	25.30%	47.77%	0	0
	CT3	0.07%	0	0	0	0
	CT4	0.03%	0	0	0	0
Industrial area	CT1	58.73%	53.54%	57.54%	0	0
	CT2	41.27%	46.46%	42.29%	0	0
	CT3	0	0	0.17%	0	0
	CT4	0	0	0	0	0
Agricultural area	CT1	0	52.12%	75.18%	52.85%	0
	CT2	0	47.88%	24.82%	47.15%	0
	CT3	0	0	0	0	0
	CT4	0	0	0	0	0
Comprehens-ive area	CT1	0	68.20%	70.75%	0	0
	CT2	0	31.72%	29.25%	0	0
	CT3	0	0.06%	0	0	0
	CT4	100.00%	0.02%	0	0	0

4.4 Vehicle Component Classification Model and Performance Assessment Analysis

According to the classification rules for vehicle parts of each factory area analyzed through the model established with the decision tree, 663 rules are listed with the statistical classification (64 in coastal areas, 132 in industrial areas, 127 in metropolitan areas, 150 in agricultural areas and 190 in comprehensive areas). Excerpts from this study that meet the higher probability of occurrence in each field are as follows:

- Rule 1: If the business station is the coastal areas and located in the northern region, the vehicle brand is CT2, the replacement time is longer than 103.5 months but equal to or shorter than 103.6 months, 85.7% of the possibility is to perform general maintenance.
- Rule 2: If the business station is the industrial area and located in the southern region, the vehicle brand is CT2, the replacement time is longer than 108.6 months but equal to or shorter than 109.0 months, 85.7% of the possibility is to perform general maintenance.
- Rule 3: If the business station is the metropolitan area and located in the southern region, the vehicle brand is CT1, the replacement time is longer than 107.7 months but equal to or shorter than 107.8 months, 87.5% of the possibility is to repair air conditioning system.
- Rule 4: If the business station is the agricultural area and located in the central region, the vehicle brand is CT1, the replacement time is longer than 85.3 months but equal to or shorter than 85.4 months, 92.9% of the possibility is to perform 3-stage disassembly.
- Rule 5: If the business station is the comprehensive area and located in the central region, the vehicle brand is CT1, the replacement time is longer than 136.2 months but equal to or shorter than 136.3 months, 90.9% of the possibility is to perform 3-stage disassembly.

5 Discussion

Vehicle parts in this study the construction of predictive maintenance decision model, through the Department of digging a lot of maintenance records, And after data cleaning, to obtain data of appropriate data quality, after research and analysis, the results of maintenance costs for different brands, chi-square inspection, component classification rules, and test data set classification performance were obtained, itemized description as follows:

Terms of Data Quality and Data Scrubbing: Good data quality must possess characteristics such as consistency, correctness, and completeness [2]. Data scrubbing is to remove dirty data and improve its data quality. Although the industry has many tools for data extraction, transformation, and installation, but lacks industry expertise and scalability applications [18]. Based on this, this study found that if didn't scrubbing dirty data from the original data as follow: inconsistent data, missing values, and other, it will not be applicate efficiently. And if the direct deletion method is adopted, the amount of data will be reduced by 61,428, which may lead to distortion of model construction and wrong analysis and decision. Therefore, in this study, a systematic and manual job analysis method was adopted, and also implements relevant expert expertise to data cleaning and sorting operations, which took approximately 240 man hours to process, and after integration, 173,693 suitable data quality data (data totals) were obtained. (Amount 93.14%). On the whole, in the modern era of rapid development of artificial intelligence, the use of workers' wisdom in research methods highlights the fact that enterprises are prolific in organizing data in different systems and periods, coupled with different professional knowledge in various industries. Effective use, most small and medium-sized enterprises, even large organizations, can only adopt the original operation mode, plus the current calculation, analysis, and operation mode, and carry out analysis and research in order to obtain core wisdom and give full play to data. Therefore, in the future, the industry can partially update the system and perform professional manpower operations, which will effectively use the data quality data obtained, and save labor costs and obtain data benefits.

The Vehicle Brands and the Maintenance Costs: Chernatony and McWilliam et al. [11] considered that the brand is consistent with the commitment to quality and also serves as a support for making a purchase decision. Owning the vehicles with reliable performance will be able to save considerable maintenance costs for the enterprise. In this study, we found that there is a significant difference between maintenance costs of the vehicles with the same attribute but not the same brands, On the whole, the maintenance costs of freight logistics vehicles mainly are concentrated on these three ranges \$176–\$335 (31%), 336–\$619 (21%) and \$620 or more (20%), which have accounted for 72% of the proportion of overall maintenance times, which shows that the high maintenance costs account for the most proportion of the overall maintenance costs in the case company so there is still more for improvement and review. As far as the brands are concerned, the main maintenance costs of CT1 and CT3 are mostly in the third and the fourth ranges, accounting for respectively 51.6% and 69% of the overall maintenance times, which shows that the maintenance costs of these two brands in the company belong to the medium or higher price; CT2 are mostly in the third and the fifth ranges, accounting for 54% of the overall maintenance times, which shows that the maintenance cost of this brand in the company belongs to the high price; CT4 are mostly in the second and the third ranges, accounting for 50% of the overall maintenance times, which shows that the maintenance cost of this brand in the company belongs to the medium or lower price. On the whole, the comprehensive analysis above has showed that there is a significant difference of the amount of maintenance between the vehicle brands of the case company and the sequence of the maintenance costs of the vehicle brands from high to low is respectively CT2, CT1, CT3 and CT4. Based on this, the organization can review the number, the service life and the dispatch frequency and other performance in the future through the costs and the frequency of the maintenance so as to achieving the purpose of saving maintenance costs and increasing the efficiency of transportation.

The Replacement of Vehicle Parts in Each Factory Area: Generally speaking, in order to prevent the wear, deformation or damage of vehicle parts from leading to problems and concerns the service, which may result in issues about safety and customer complaints and other issues, the general preventive maintenance, repair and other operating items are often scheduled. In this study, we have found that due to the complete government regulations and investment and construction planning, the industrial areas (Industrial parks) have the better and flatter regional roads and customers are more concentrated and thereby the general preventive maintenance and repair and other needs for replacing component are relatively lower (general repair accounted for only 27%); the general maintenance and the general repair and other maintenance items in other areas are listed in the replacement of main components accounting for more than 45% (47% in the coastal areas, 49% in the metropolitan areas, 57% in the agricultural areas and 58% in the comprehensive areas), which has revealed that in order to save the replacement material effectively and to reduce the maintenance costs, the company should take the general maintenance and repair aa the main contents for review and adjustment. And in the coastal areas, because the vehicles were running on the coastal areas with relatively high salinity and humidity, in addition to the general maintenance and repair, switch wiring (31%) has accounted for more than 30% of the replacement needs for components, which has revealed that in the coastal areas, the company should pay more attention to the maintenance against the rust damage and influences on the vehicles due to the salinity and the humidity. And about the maintenance times of vehicles, we have found out that in each factory and area, if taking the brands into consideration, CT1 and CT2 vehicle had higher proportion of maintenance and repair, in which especially the maintenance data of the freight vehicles CT1 in the central metropolitan areas (74.70%), in the southern agricultural areas (75.18%) and in the southern comprehensive areas (70.75%) and other regions were higher than 70%, which has revealed that the maintenance of CT1 was mainly concentrated in the central and southern regions, but which of CT4 was primarily concentrated in the northern comprehensive areas.

Vehicle Parts Classification Model: In this study, we followed the decision tree to analyze and took the areas, factory areas, brands, items, possibility, cycle and other key elements about replacement of vehicle parts into consideration so as to construct an optimal decision support model. In thus study, we have found that the often damages of freight logistics vehicle parts in specific factory areas and regions are related to each other. According to the predictive replacement model, we can obtain the rules and probability of maintenance and replacement in different factory area and different regions. Based on the Decision Support Model, we have constructed a total of 663 classification rules combined with the aforementioned research analysis of vehicle brands, maintenance costs and replacement of vehicle parts to propose the main rules for maintenance and repair. The brands CT1 and CT2 were found to be majority; with the classification in each factory area and region and in comparison with replacement times of vehicle parts, we have found out that the replacement probability of components is higher than 60%, partially more than 90%, which has showed the classification rules through the decision-making model can effectively improve the replacement benefits of vehicle parts.

In summary, in the past, only a few researches were made about predictive vehicle parts replacement decision support for the freight logistics vehicle parts. This study has particularly focused on the maintenance data for all 3.5 ton level freight vehicles to discuss about the rules for maintenance and replacement of vehicle parts then to conduct the data analysis. At the same time, according to the study and analysis of the results of the classification rules, we have verified that the vehicle parts of some particular brands in some specific regions and factory areas were often damaged and in bad condition. The decision support model developed will help the company to establish a reference for the company to create a decision support model to build a predictive optimal replacement cycle. In practice it will help the case company to acquire the appropriate maintenance timing for vehicle parts and the appropriate maintenance frequency as well, and thereby to reduce the maintenance costs and to solve the problem about library reserve stock of vehicle parts, in order to ensure the best condition of operating vehicles and to improve their management and maintenance; on the other hand, it can help the company to determine if the company should continue to maintain these vehicles or scrap the old vehicles or purchase new vehicles, etc., and can clarify the condition that specific vehicle parts in the business station of the case company are often damaged in different factory areas and different regions. On the whole, the study results can help vehicle managers and technicians to obtain maintenance experience; moreover, with the knowledge sharing and storage, the company will be able to solve the problems about uneven quality of maintenance and about technicians' insufficient maintenance ability. With the decision support for an optimum replacement cycle of vehicle parts, the organization can save the costs related to maintenance.

6 Conclusion and Suggestions

With the growing of economy, the logistics industry has been committed to achieving the "Last mile" like physical transportation and distribution services, which is the important connotation of supply chain activities [30]. Therefore, the competitiveness among the logistics industries has become very fierce, so the poor efficiency of goods distribution will represent resources wastes and the weak competitiveness. Only a few researches were made about predictive vehicle parts replacement decision support for the freight logistics vehicle parts. The case company in this study accounted for one of top five proportion of Taiwanese freight logistics market share. Use its complete vehicle maintenance information to obtain good data quality data through system and manual data cleaning methods, And combined with data mining classification methods for analysis, and then according to the classification rules and the analysis results to verify that the vehicle parts in some specific regions and factory areas were often damaged and in bad condition when being maintained or replaced. The decision support model developed in this study will help the company to establish a reference for the company to create a decision support model to build a predictive optimal replacement cycle. This will help the organization to acquire the appropriate maintenance timing for vehicle parts and to confirm the appropriate maintenance frequency, it has also taken the corporate business model into account to provide vehicle maintenance technicians and management personnel maintenance knowledge and ability which are different from the previous standard of care standards or don't only rely on rules of thumb. Therefore, through this study, the company can reduce the maintenance costs and to solve the problem about library reserve stock of vehicle parts, in order to ensure the best condition of operating vehicles and to improve their management and maintenance. On the other hand, it can help the company to determine if the company should continue to maintain these vehicles or scrap the old vehicles or purchase new vehicles or other decisions.

According to the study, we have found out that there are some significant differences of maintenance costs between the vehicles of the case company with the same attribute but different brands; the vehicle parts of some particular brands in some specific regions and factory areas were often damaged and in bad condition. The decision support model developed in this study will help the company to establish a reference for the company to create a decision support model to build a predictive optimal replacement cycle. On the whole, the study results can help supervisors and technicians to obtain maintenance experience; moreover, with the knowledge sharing and storage, the company will be able to solve the problems about uneven quality of maintenance and about technicians' insufficient maintenance ability. Meanwhile, with the decision support for the optimum replacement cycle of vehicle parts. In addition, through the data quality cleanup operation, we can learn that the company's data can be used to complete the uniform, standardized and unified use of the system platform interface, in addition to saving labor and operation costs, and can effectively use high-performance data quality management methods, so that the data generated within the organization can be fully utilized to achieve comprehensive results, and then achieve business optimization, cost energy efficiency and utility optimization of corporate goals, so as to achieve enhanced competitive advantage To ensure the goal of sustainable operation. In summary, in the future we can develop a more appropriate predictive replacement model of the vehicle parts combined with the model of freight vehicles with different types and tonnages level, the drivers' driving habits, the vehicle running state and other information.

References

1. Agresti, A., Kateri, M.: Categorical data analysis. In: Lovric, M. (ed.) International Encyclopedia of Statistical Science, pp. 206–208. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-04898-2_161
2. Aebi, D., Perrochon, L.: Towards improving data quality. In: Sarda, N.L. (ed.) Proceedings of the International Conference on Information Systems and Management of Data, Delhi, pp. 273–281 (1993)
3. Berry, M.J., Linoff, G.: Data Mining Technique: For Marketing, Sales, and Customer Relationship Management. Wily Press, New York (1997)
4. Bose, I., Mahapatra, R.K.: Business data mining-a machine learning perspective. Inf. Manag. **39**(3), 211–225 (2001)
5. Bousdekis, A., Papageorgiou, N., Magoutas, B., Apostolou, D., Mentzas, G.: Enabling condition-based maintenance decisions with proactive event-driven computing. Comput. Ind. **100**, 173–183 (2018)
6. Brachman, R.J., Khabaza, T., Kloesgen, W., Piatetsky-Shapuro, G., Simoudis, E.: Mining business database. Commun. ACM **39**(11), 42–48 (1996)
7. Chen, M.C., Hsu, C.L., Chang, K.C., Chou, M.C.: Applying Kansei engineering to design logistics services–a case of home delivery service. Int. J. Ind. Ergon. **48**, 46–59 (2015)
8. Chernatony, L.D., McWilliam, G.: Branding terminology the real debate. Mark. Intell. Plan. **7**(7), 29–32 (1989)
9. De Faria, H., Costa, J.G.S., Olivas, J.L.M.: A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis. Renew. Sustain. Energy Rev. **46**, 201–209 (2015)
10. Farid, D.M., Zhang, L., Rahman, C.M., Hossain, M.A., Strachan, R.: Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. Expert Syst. Appl. **41**(4), 1937–1946 (2014)
11. Fayyad, U., Poatersky-Shapiro, G., Symth, P.: From data mining to knowledge discovery in databases. AI Mag. **17**, 37–54 (1996)
12. Ghajargar, M., Zenezini, G., Montanaro, T.: Home delivery services: innovations and emerging needs. IFAC-PapersOnLine **49**(12), 1371–1376 (2016)

13. Han, J., Pei, J., Kamber, M.: Data mining: concepts and techniques. Elsevier, Amsterdam (2011)
14. Galhardas, H., Florescu, D., Shasha, D., Simon, E., Saita, C.A.: Declarative data cleaning: language model and algorithms. In: Proceedings of the 27th International Conference on Very Large Databases (VLDB 2001), pp. 371–380 (2001)
15. Jim, W.: The butterfly effect of data quality. In: The Fifth MIT Information Quality Industry Symposium (2011)
16. Lee, X., Massman, W., Law, B.: Handbook of Micrometeorology. A Guide for Surface Lux Measurement and Analysis. Springer, Dordrecht (2005). https://doi.org/10.1007/1-4020-2265-4
17. Cai, L., Zhu, Y.: The challenges of data quality and data quality assessment in the big data era. Data Sci. J. **14**(2), 1–10 (2015)
18. Ohbyung, K., Namyeon, L., Bongsik, S.: Data quality management, data usage experience and acquisition intention of big data analytics (2014)
19. Schmid, H.: Probabilistic part-ofispeech tagging using decision trees. In: New Methods in Language Processing, p. 154. Routledge (2013)
20. Shaw, M.J., Subramaniam, C., Tan, G.W., Welge, M.E.: Knowledge management and data mining for marking. Decis. Support Syst. **31**(1), 127–137 (2001)
21. Shin, J.H., Jun, H.B.: On condition based maintenance policy. J. Comput. Des. Eng. **2**(2), 119–127 (2015)
22. Taleb, I., Dssouli, R., Mohamed, A.S.: Big data pre- processing: a quality framework. In: 2015 IEEE International Congress on, pp. 191–198 (2015)
23. Tilahun, N., Thakuriah, P.V., Li, M., Keita, Y.: Transit use and the work commute: analyzing the role of last mile issues. J. Transp. Geogr. **54**, 359–368 (2016)
24. Waeyenbergh, G., Pintelon, L.: A framework for maintenance concept development. Int. J. Prod. Econ. **77**(3), 299–313 (2002)
25. Wang, Y.F., Zhang, C.Z., Zhang, B.B., et al.: A survey of data cleaning. New Technol. Libr. Inform. Serv. **12**, 50–56 (2007)
26. Wu, I.C., Li, R.J., Chen, T.L.: A vehicular maintenance and replacement decision support system in distribution services: a data mining technique. J. Manag. Syst. **21**(1), 111–137 (2014)
27. Zheng, Z., Zhou, W., Zheng, Y., Wu, Y.: Optimal maintenance policy for a system with preventive repair and two types of failures. Comput. Ind. Eng. **98**, 102–112 (2016)