Economical Maintenance and Replacement Decision Making in Fleet Management using Data Mining

Siddharth Arora*, Parneet Kaur** & Prachi Arora***

*Research Scholar, Department of Computer Science & Engineering, Ambala College of Engineering & Applied Research, Devsthali, Ambala, Haryana, INDIA. E-Mail: siddhartharora10@gmail.com
**Assistant Professor, Department of Computer Science & Engineering, Ambala College of Engineering & Applied Research, Devsthali, Ambala, Haryana, INDIA.
***Assistant Professor, Department of Economics, Shivaji College, University of Delhi, New Delhi, INDIA.
E-Mail: prachi.economics@gmail.com

***Abstract*—**Recent advances in technologies allowed the development of fleet management systems that enable fleet operators and freight carriers to their fleet performance, so that optimum replacement stage can be calculated or before purchasing new vehicle by analyzing its performance which improve relevant performance as well as cost effective by intervening when such confusion occur in the mind of customer. The aim of this research work is to enhance vehicle life cycle and replacement or disposal of vehicle at the critical point so that cost-effective and good performance both can be achieved by analyzing the performance of vehicle by the use of K-Mean clustering algorithm a decision support system is designed. Analysis of the fleet performance by modeling the process of analyzing through the design and implementation of a system analysis of maintenance & replacement policies in fleet management system using data mining. The latter has three main functionalities: a) it monitors existing vehicles service records, b) it detects maintenance deviations from the normal plan and c) it detects the performance analysis and also its deviation, d) it also detect overall analysis, by the help of major affecting factor.

***Keywords*—**Data Mining, Decision Making in Fleet Management, Depreciation Analysis, Fleet Management System, K-Mean Analysis, Maintenance Analysis, Overall Analysis

***Abbreviations*—**High Maintenance (HM), Knowledge Discovery in Database (KDD), Low Maintenance (LM), Medium Maintenance (MM), Most Frequent Pattern (MFP)

I. INTRODUCTION

IT is possible to efficiently extract or mine" knowledge from large amounts of vertically partitioned data within security restrictions. Knowledge Discovery in Databases (KDD) is the term used to denote the process of extracting knowledge from large quantities of data. The KDD process assumes that all the data is easily accessible at a central location or through centralized access mechanisms such as federated databases and virtual warehouses. The automobile industry is among the most information intensive industries. Automobile information, knowledge and data keep growing on a daily basis. A formula based on the current data available, historical trends, and projections is used to estimate cars performance. The auto industry is an important sector of the global economy. By "car" we are referring to passenger cars, which are defined as motor vehicles with at least four wheels, used for the transport of passengers, and comprising no more than eight seats in addition to the driver's seat. Cars (or automobiles) make up approximately 74% of the total motor vehicle annual production in the world. The remaining 26%, not included in this statistics, is made up by light commercial vehicles and heavy trucks (motor vehicles with at least four wheels, used for the carriage of goods), buses, coaches and minibuses (comprising more than eight seats in addition to the driver's seat) [Nassar Khaled, 2007].

1.1. Cars Produced in the World

In 2012, for the first time in history, over 60 million cars passenger cars will be produced in a single year (or 165,000 new cars produced every day). Moreover in this India is the second fastest growing automobile market in the world [http://www.worldometers.info/cars/].

After a 9% decline in 2009 (due to the 2008 global financial crisis), global car production immediately jumped back the following year with a 22% increase in 2010, to then consolidate at the current 3% yearly growth rate [Michael J. Berry & Gordon S. Linoff, 1988].

Going back in history, in 2006 there were less than 50 million passenger cars produced in the world, with an increase of 6.45% over the previous year. The increase for 2007 was more modest, and 2008 showed a decline [http://oica.net/].

So, at the time of purchasing new vehicle the ability to use the performance related data of all these vehicles as above listed to extract useful information for quality purchase of new vehicle is crucial.

1.2. Goals and Objectives

The application of artificial intelligence in automobile workshop is relatively new. The aim of this research work is to show that data mining can be applied to the automobile workshop databases, which will predict or classify the data with a reasonable accuracy. For a good prediction or classification the learning algorithms must be provided with a good training set from which rules or patterns are extracted to help classify the testing dataset [Kusiak, 2006].

Main goal of this work is that at the time of purchasing new vehicles when there are a number of choices are available and customer doesn't have much of the prior knowledge regarding the performance of these vehicles in actual practice then it becomes much typical for the customer to choose the correct one in terms of quality purchase. So for this reason I had tried my best to maintain the accuracy in collecting the data related to the performance behavior of vehicles from different service agencies and various departments so that after the pre-processing of these data and applying the algorithm a quality knowledge discovery can be learned [Jiawei Han & Micheline Kamber, 2006].

A quality correct decision can save up to 30% of the Maintenance cost and 38% of depreciation cost. A number of data mining algorithms will be used in this work to show the drawbacks and advantages. One of the tools has a built in pre-processing tool. A pre-processing tool is used to convert raw data into a format understandable by the data-mining algorithm. The rest of the tools require data to be sent to the algorithms in various formats.

1.3. Background

With the evolution of machines, we have found that some tiring and routine or complex mathematical calculations can be done using calculators, finding specific information in a large database can be done using machines fast and easily. We use machines for storing information; remind us of appointments, and so on. As the size of the data was increasing computer storage has increased. Due to the vast amount of data that was being created humans invented algorithms that produce results once a query is supplied. Although these tools perform very well, they can be used to perform only routine tasks. Automatic classifications and other machine intelligence algorithms cannot be done using standard database languages. This has led to the creation of machine intelligence algorithms that can perform tasks supplied by humans and make decisions without human supervision. From the evolution of machine intelligence came data mining. In data mining, algorithms seek out patterns and rules within the data from which sets of rules are derived. Algorithms can automatically classify the data based on similarities (rules and patterns) obtained between the training on the testing data set.

Today, data mining has grown so vast that they can be used in many applications; examples include predicting costs of corporate expense claims, in risk management, in financial analysis, in insurance, in process control in manufacturing, in healthcare, in automobile and in other fields [Harding et al., 2006]. Let us consider an example in automobile. The number of people wants to purchase car and purchasing new car are increasing proportionally. The growing number of customer indirectly increases amount of data that are required to be stored. If a small number of vehicle models are available, then customer`s decision will be able to work efficiently and provide quality decision. Now consider the case when there are a large number of vehicle models available to purchase and customer also have limited time to purchase new vehicle (as to meet out its requirement). We will find that the probably there arises more chance of customer get confused and quality of decision of the customer can decrease.

So, in such situation to take the help of computers to provide a second opinion to the customer can be a feasible solution. The computers will search for patterns within the database and will provide the customer with a fast opinion of which vehicle can have better performance.

II. LITERATURE REVIEW

The digital revolution has made digitized information easy to capture, process, store, distribute, and transmit With significant progress in computing and related technologies and their ever-expanding usage in different walks of life, huge amount of data of diverse characteristics continue to be collected and stored in databases. The rate at which such data are stored is growing phenomenally. We can draw an analogy between the popular Moore's law and the way data are increasing with the growth of information in this world of data processing applications. The advancement of data processing and the emergence of newer applications were possible, partially because of the growth of the semiconductor and subsequently the computer industry. According to Moore's law, the number of transistors in a single microchip is doubled every 18 months, and the growth of the semiconductor industry has so far followed the prediction. We can correlate this with a similar observation from the data and information domain. If the amount of information in the world doubles every 20 months, the size and number of databases probably increases at a similar pace. Discovery of knowledge from this huge volume of data is a challenge indeed. Data mining is an attempt to make sense of the information explosion embedded in this huge volume of data.

Data Mining is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. There are other definitions:

- Data Mining is the process of exploration and analysis of large quantities of data in order to discover meaningful patterns and rules from large amounts of data.
- Data Mining is the process of discovering interesting knowledge from large amounts of data stored either in databases, data warehouses, or other information repositories. Simply stated, data mining refers to extracting or "mining" knowledge from large amounts of data [Aurangzeb Khan et al., 2009].

Actually the major reason that data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years is due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. The data mining process is sometimes referred to as knowledge discovery or KDD. The term "KDD" (Knowledge Discovery in Databases) refer to the overall process of discovering useful knowledge from data. There is a difference in understanding the terms "knowledge discovery" and "data mining" between people from different areas contributing to this new field.

2.1. Data Mining Process

Generally KDD is an iterative and interactive process involving several steps. This KDD process was chosen (Figure 1) according to UNESCO definition because of its simplicity and comprehensiveness.

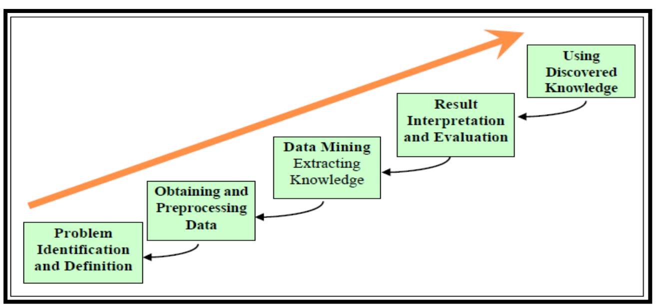

Figure 1 – The KDD Process [Aurangzeb Khan et al., 2009]

2.1.1. Problem Identification and Definition

The first step is to understand the application domain and to formulate the problem. This step is clearly a prerequisite for extracting useful knowledge and for choosing appropriate data mining methods in the third step according to the application target and the nature of data.

2.1.2. Obtaining and Pre-processing Data

The second step is to collect and reprocess the data. Today's real-world databases are highly susceptible to noisy, missing, and inconsistent data due to their typically huge size (often several gigabytes or more), and their likely origin from multiple, heterogeneous sources. Low quality data will lead to low quality mining results. Data pre-processing is an essential step for knowledge discovery and data mining. Data pre-processing include the data integration, removal of noise or outliers, the treatment of missing data, data transformation and reduction of data, etc. This step usually takes the most time needed for the whole KDD process.

2.1.3. Data Mining / Knowledge Discovery in Databases

The third step is data mining that extracts patterns and/or models hidden in data. This is an essential process where intelligent methods are applied in order to extract data patterns. In this step we have to select first data mining tasks and then data mining method. The major classes of data mining methods are predictive modeling such as classification and regression; segmentation (clustering) and association rules which are explained in detail later.

2.1.4. Result Interpretation and Evaluation

The fourth step is to interpret (post-process) discovered knowledge, especially the interpretation in terms of description and prediction which is the two primary goals of discovery systems in practice. Experiments show that discovered patterns or models from data are not always of interest or direct use, and the KDD process is necessarily iterative with the judgment of discovered knowledge. One standard way to evaluate induced rules is to divide the data into two sets, training on the first set and testing on the second. One can repeat this process a number of times with different splits, and then average the results to estimate the rules performance.

2.1.5. Using Discovered Knowledge

The final step is to put discovered knowledge in practical use. Putting the results into practical use is certainly the ultimate goal of knowledge discovery. The information achieved by data mining can be used later to explain current or historical phenomenon, predict the future, and help decision-makers make policy from the existed facts [Gardner & Xiong, 2009].

2.2. Application of Data Mining in Fleet Management

In data mining the information and knowledge gained can be used for applications ranging from automobile industry related product marketing, customer profiling and retention, identifying potential of different available models of vehicle, market segmentation, fraud detection, text and web mining, e-commerce to production control, manufacturing and science exploration.

Data mining technology in the fleet management system is a relatively universal application. No matter whether we see it in the term of analysis of prior record of vehicles or marketing of such vehicles, or in analysing the customer selection etc [Malhotra et al., 2003]. Such applications are referred to a Boundary Science, because it sets a variety of scientific theories in all. First, two basic disciplines: Information Technology and analysing market`s available choices. Another very important basis is Statistics. The charm of this area is just about the wide scope of disciplines study. Generally speaking, through the collection, processing and disposal of the large amount of information involving vehicle service record, vehicle market value prediction on the basis of that vehicle depreciation analysis helps in identifying the quality product among the number of available choices. This is the basic idea. As automation is popular in all the industry operate processes, enterprises have a lot of operational data. These data are not collected for the purpose of analysis, but come from commercial operation. Analysis of these data does not aim at studying it, but for giving business or end-user decision-maker the real valued information, in order to get profits. Commercial information comes from the market through various channels. For example, when any car comes in the workshop for repair or its regular service, we can collect the vehicle maintenance raw record data, such as brand of vehicle, model of vehicle, manufacturing year of vehicle number of kilometre it runs, prior maintenance history of vehicle etc when filling a vehicle inspection form at the time of entry of vehicle in the workshop can collect vehicle maintenance trends and frequency. In addition, enterprises can also buy a variety of customer information from other consulting firms in such field.

Performance analysis based on data mining usually can give the predetermine results based on the practically prior actual performance record of the vehicle. It should be emphasized data mining is application-oriented. There are several typical applications in marketing, decision making, banking, insurance, traffic-system, retail and such kind of commercial field. Generally speaking, the problems that can be solved by data mining technologies include: analysis of market, such as database marketing, customer segmentation & classification, profile analysis and cross-selling. And they are also used for churn analysis, credit scoring and fraud detection.

III. PROBLEM FORMULATION

In the whole data mining process, data preparation is somehow a significant process. Some book says that if data mining is considered as a process then data preparation is at the heart of this process [Pyle, 1999]. However, nowadays databases are highly susceptible to noisy, missing, inconsistent data and due to their typically huge size (often several gigabytes or more), and their likely origin from multiple, heterogeneous sources. There are a number of data preprocessing techniques: data cleaning, data integration, data transformation and data reduction. Data cleaning can be applied to remove noise, supply missing values and correct inconsistencies in the data. Data integration merges data from multiple sources into a coherent data store, such as a data warehouse. Data transformations involve data normalization/ scaling, aggregation and feature construction. In data transformation, the data are transformed or consolidated into forms appropriate for mining. Data reduction can reduce the data size by aggregating, eliminating redundant features (feature selection), or clustering, for instance. These techniques are not mutually exclusive; they may work together. Data processing techniques, when applied prior to mining, can substantially improve the overall quality of the patterns mined and/or the time required for the actual mining [Elovici et al., 2004]. Data preprocessing step usually takes the most time needed for the whole KDD process, Researchers estimates that data preparation alone accounts for 60% of all the time and effort expanded in the entire data mining process. In following sections data cleaning, data integration, data normalization/scaling and feature construction (data transformation) are explained.

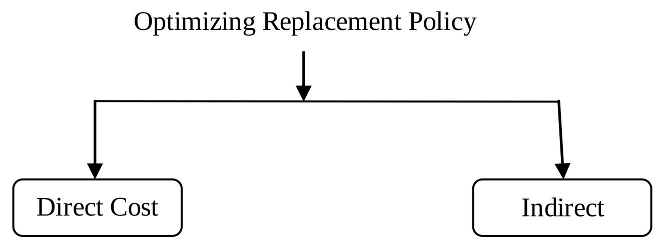

Figure 2 – Optimizing Replacement Policy

One of the most debated issues among fleet managers and leasing companies is when to sell a unit to maximize returns and minimize lifecycle costs. To perform a thoughtful analysis of this issue, it is important to consider all direct (hard) costs, including depreciation and maintenance, as well as indirect (soft) costs, including perk value, corporate image, downtime, and feature upgrades.

Depreciation, a direct cost, is the largest single controllable expense of running a fleet and can be easily tracked over the lifecycle of a vehicle. Vehicles depreciate heavily in the first couple of years, and more slowly as the vehicle ages. See the green area on the Lifecycle Costs graph, which represents depreciation cost-per-mile for atypical fleet sedan, illustrating the high depreciation in the first two years [Chiheb Saidane et al., 2007].

Table 1 – Ratio in Cost

Direct Costs	Indirect Costs
Depreciation	Perk Value
Maintenance	Corporate Image
	Downtime
	Feature Upgrades

Depreciation comprises two components: utility and prestige value. The utility component (the usefulness of the vehicle) is based entirely on mileage — a vehicle loses utility value with every mile driven. The other component, new-vehicle prestige, drops dramatically at delivery — the proverbial 30-percent loss going over the curb. Prestige value continues to drop quickly throughout the first two years, and by years four and five virtually no prestige value remains. For the balance of the vehicle life, depreciation is based solely on the amount of utility left in the vehicle.

Maintenance, another direct cost, has become somewhat easier to anticipate. Improved reliability and durability have greatly reduced the occurrence of major mechanical failures below 100,000 miles. The blue area of the Lifecycle Costs graph on the page below, which represents maintenance cost-per-mile, illustrates this. Please note that the dramatic increase in maintenance cost at about 140,000 miles is a result of a major mechanical breakdown.

Clearly, it is not financially prudent to operate extremely high-mileage vehicles because the maintenance costs eventually exceed the cost of a new vehicle. Excluding indirect costs, the best time to sell a unit is just before a major breakdown; however, the challenge lies in pinpointing when it will occur. To avoid major maintenance expense, it is recommended that passenger vehicles be replaced at a maximum of about 120,000 miles. However, when indirect costs are examined and weighed, lower replacement mileages can be rationalized. Indirect costs to factor in are:

- Perk Value: Extending the replacement life of a vehicle can negatively impact actual or perceived driver benefits, whereas a shorter replacement cycle can enhance them. • Corporate image: Fleet operators may choose to replace their vehicles well before 100,000 miles to maintain a younger, cleaner fleet that projects a desired image.
- Downtime: Lost opportunities and rental costs incurred when drivers must await repairs on aging vehicles are important factors to consider.
- Feature Upgrades: Extending the replacement life of a vehicle can delay the implementation of new features, including safety features such as side airbags and tire-pressure indicators or convenience features such as upgraded sound systems. There is no magic formula for deciding when to replace fleet vehicles, but a careful and thoughtful analysis of both the direct and indirect costs will yield the best policy for your fleet. However, when indirect costs are examined and weighed, lower replacement mileages can be rationalized. Indirect costs to factor in are:

There is no magic formula for deciding when to replace fleet vehicles, but a careful and thoughtful analysis of both the direct and indirect costs will yield the best policy for your fleet.

3.1. Vehicle Replacement: How long is too long?

Knowing in advance exactly when to replace a vehicle is sort of like knowing when to bail out of your favorite stock or commodity: it's an indefinable, unpredictable guess. Is there a formula that calculates in advance when it's no longer cost-effective to keep a company vehicle in service? No. Can you, however, collect enough data to make an educated, well informed decision to develop a standard replacement cycle for your company vehicles? Yes! In addition, by developing a comprehensive replacement program, you can make a significant contribution to your company's bottom-line profit.

3.2. Depreciation is the Single Most Important Factor

First, consider the following four characteristics of the replacement decision: Depreciation is the single-most important and significant factor for achieving the lowest possible cost. Simply put, depreciation costs decline faster than other costs increase. What is depreciation? It is the vehicle acquisition cost minus its resale value plus any cost associated with resale.

For example:

Rs. 500,000 Vehicle Acquisition Cost - Rs. 200,000 Resale Value + 5000 Sales Fee = Rs. 305,000 Depreciation.

Three significant factors affect depreciation:

1) Acquisition Cost,
2) Replacement Timing, and
3) Mileage.

Table 2 – Factors to Consider

S. No	Costs to Consider	Guess Factor
1	Acquisition	
2	Resale Value	Future unknown, unpredictable used vehicle market, high impact
3	Resale Fees	
4	Interest	Future unknown, somewhat predictable, less impact
5	Fuel	Future unknown, somewhat predictable, less impact
6	Maintenance and Repairs	Future unknown, somewhat predictable, less impact
7	Insurance	
8	Title and License	

3.3. Maintenance: The Second-Largest Vehicle Cost

Maintenance directly affects reliability, which can conceal significant "hidden" or "soft" costs: downtime, lost time, lost business, lost sales, diminished productivity, plus "hard" costs such as repair and car rental expenses. Vehicle condition affects resale value, which in turn affects depreciation. Monitoring the condition of vehicles through written condition reports, vehicle inspections, and increased driver responsibility will help improve the vehicle condition and resale value results. Consider requiring two signatures on the condition report, both the driver and the supervisor. Review condition reports at least twice a year. Use the odometer reading to update your records and monitor the vehicle for replacement.

3.4. Extending Vehicle Life

3.4.1. Advantages

Table 3 – Advantage (Extending Vehicle Life)

S. No	Advantages
1.	Reduces depreciation
2.	Lower taxes and license fees
3.	Lower collision repair costs

3.4.2. Disadvantages

Table 4 – Disadvantage (Extending Vehicle Life)

S. No	Disadvantages
1.	Higher maintenance costs
2.	Diminished productivity downtime
3.	Older vehicle technology, higher fuel costs
4.	Lower employee morale
5.	Safety risks associated with vehicle breakdown, parts failures, and older technology
6.	Values decline at a faster pace during weak used-vehicle market conditions

An analysis of your past replacement cycles and costs weighed against those non-financial, non-qualitative factors and your company's unique requirements should provide a clear indication of the most appropriate months-in-service and miles-in-service for your company vehicles. A flexible replacement policy will allow you to adapt to changing conditions in both the used-vehicle market and manufacturers' new model year availability.

3.5. Focusing Area of Research

This research work attempts to explore the possibility to use data mining K-Mean clustering algorithms and techniques which helps to make quality decision of field survey data by providing the result in the form of graphical clustered representation. The actual practical raw data is gathered from various authorized service centers of various brands. We had firstly pre-processed the data from the raw collected data. Based upon this genuine practical data we had designed and developed a system by applying data mining K-Mean clustering technique to cluster the data in the form of graphical clustered view by the use of various selected important traits.

The central task is to derive conclusions on the basis of the K-Mean clustering data mining techniques and algorithms that will be used for the quality decision making by the end user of the system. Certainly, in order to achieve this many tests should be performed and necessary analysis should be carried out on their results. Implementation of an application that would provide the needed data mining apparatuses to perform the tests is also a significant part of this research work. Special attention is dedicated to constructing clustered graphical results that would be useful for the end user. The classifiers by ensuring that there is enough information for them to find the patterns underlying the data. The main issues to be solved in order to have a successful approach toward the problem solution are:

- ***Maintenance Analysis***: When you look to buy a car in the market, you will be able to find numerous options in front of you. Therefore, if you have to obtain the best results, it is very important to analyse the various factors as which are helpful in determining before buying the behaviour of the car performance. So, we had determined the factors which are important in analyzing the car performance and on the basis of that we had designed and developed a user friendly system so that on the basis of the prior actual collected data information our designed system end-user can analyse the performance of the vehicle with the help of various important traits. So by analysing through our system it will help a lot to the end-user in quality decision making. In maintenance analysis the factors which are responsible for affecting the maintenance of the vehicle are analysed and on the basis of these factors a system is designed which contributes to the clustered graphical results on the basis of the selected parameters. Main motto behind the design of this system is to priory analyse the performance of the vehicle as will be performed by the vehicle in general conditions. Also in addition to this it will also help in deciding which vehicle a new purchaser should buy as end user can guess the maintenance related performance details of various vehicles before buying.
- ***Depreciation Analysis***: One of the factors in optimisation is the right depreciation. Depreciation analysis is also one of the important factors in quality decision making regarding the performance of the vehicle.

  Depreciate too fast and you will lose money; depreciate too slowly and you will still lose money because in such situation maintenance cost will be increase a lot. We had designed the system and expertise to find the optimal rate that will enable you to benefit from a balanced cash flow and to better allocate your funds towards what is most important to you.

  However a proper analysis on the physical meaning of the data should be done in advance. The classification of wrong and right data is not trivial and a number of parameters should be included in the data sets in order to make the classification process easier and more accurate.
- ***Overall Analysis***: Auto customization is earning a lot of popularity in the market these days. If we walk on the road, we shall be able to see different types of custom cars ripping the road. However, here we are trying to analyze the overall performance of the car .in the overall analysis we will analyse the different parameters on the basis of those the overall cluster analysis results in the graphical form can be achieved. Moreover, by getting the overall analysis performance result end-user will be able to analyse the ability and performance of various vehicle's options which are available.
- ***Overall Report Generation***: Overall Report generation helps to easily design interactive reports and connect them to virtually data source. Users can benefit from on-report sorting and filtering – giving them the power to execute decisions instantly. This provide end-user overall prior record of the selected vehicle in the form of mathematical format.
- ***Using the Best Data Mining Practices***: The data mining process of building a model is done by following a specific protocol. There are several possibilities of using them as number of techniques are available, therefore necessary tests should be done to pick the most reliable and safe technique which is to be used . Reliability of the performance of the technique which we are going to used should be checked by ensuring that the result of the technique which we are going to be use are in accordance to our requirement and would be as accurate as possible.

- ***Selecting the Best Parameters***: Judging about performance of a certain vehicle for a given model and compare it with others is always done with the aim of picking the best performing parameters which should be analyzed. However this becomes complicated when the results of many parameters are close to each other. Thus several measurements of the performance of each parameter should be considered to make a decision for the best one. So parameters are selected after the discussion with logistic managers along with trial verification and on the basis of those these parameters are included in the system and a clear analysis is made.

IV. Research Methodology

Qualitative approach is employed for the evaluation of the data mining methods. Here particular features of them are analyzed. Also the interrelations among the attributes are taken into consideration. A thorough and comprehensive literature study is carried out prior to the developmental and analytical work. The literature encompasses peer-reviewed articles, journals, books and discussion with logistic managers. First, the analysis of the data mining algorithms and their applicability to the performance analysis set is done. Afterwards, a data warehouse is designed and filled with the data of the actual vehicle collected records. This step includes an assessment of the quality of the data. Then, the studied data mining methods are applied to the warehouse [Creswell, 2002].

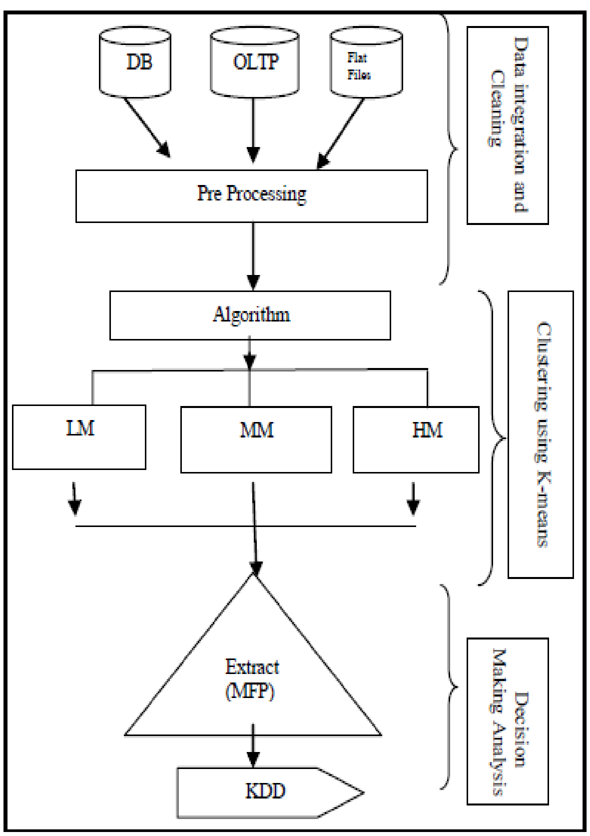

Figure 3 – Model of Research Work

Finally, the indications of the data mining models obtained in the previous phase are used to generalize the knowledge extracted from the dataset. we proposed a model for mining patterns of huge stock data to analyse the factors affecting the fleet`s maintenance and depreciation. In the first phase, we filter out the data from the database, OLTP and flat files in three different clusters by applying some pre-processing filters. On the basis of the pre-processing filter result we will design the algorithm to get the results in different clusters i.e. Low-Maintenance (LM), Medium - Maintenance (MM) and High–Maintenance (HM) using K-Means algorithm. In the second phase we have proposed Most Frequent Pattern (MFP) algorithm to find frequencies of property values of the corresponding items to analyse the factors such as Maintenance system and Depreciation system so that on the basis of the results we can easily analyse and can get the cost effective decision making.

4.1. Data Collection and Description

After designing the most suitable proposed model, it is necessary to decide on how the empirical data will be collected. Secondary data can be classified as internal and external. The internal data are those generated within the organization for which the research is being conducted and the external data are those generated by sources outside the organization. Vehicles historical maintenance data is collected from organization's database (internal databases) and some additional information such as vehicle category, depreciation cost, market share of a given car etc are gathered from external databases.

As mentioned above, for data mining purpose secondary data are used. The information at the level of the individual user is typically used to construct a response model. Response models are typically built from historical purchase data. For data mining purposes, the maintenance history can often be translated into features based on measures of service kilometer and monetary values. This work was mainly based on secondary data. Internal secondary data were gathered from various authorized workshops of various brands for building a model.

For conducting this study three types of data were collected from Parsian bank's databases: Customer historical purchase data and demographic, Customer transaction data and Campaigns Data.

4.1.1. Vehicle Historical Purchase Data

For building a response model, three sets of variables, Recency- Frequency and Monetary, must be used. In order to build these variables, there is a need for vehicle historical purchase data. Thus vehicle historical purchase data were collected from authorized agency. This data includes such information as vehicle brand and model, unique engine number and chassis number, date of sale of vehicle. This information is firstly confirmed from the agency from where that proper vehicle has been sold.

4.1.2. Vehicle Maintenance Data

For building the RFM variables, besides historical purchase data, vehicle maintenance data is also need. Maintenance data includes the service / repair kilometer every time when that vehicle comes in the workshop, date of service / repair and amount of money that the customer pays for the service.

4.1.3. Miscellaneous Data

For building target variable in addition to customer history we need campaign information. As I explained above depreciation cost of vehicle is decided based on the market response in response to that particular model in the market (Considering the average condition of the vehicle w.r.t. manufacturing year).Miscellaneous data includes the date of the evaluation of cost and the evaluation of the cost of the vehicle in the market w.r.t that particular vehicle.

Historical, old vehicle sale data is also collected from different sources so that it may help in finding the actual market value of the vehicle from time to time.

4.2. Data Processing

Data pre-processing is an often neglected but important step in the data mining process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects. As we gathered the data so, in the collected data It is required there may be some missing value related to any parameter or may be presence of noise in the data or irrelevant field of value. So, data-gathering methods are often loosely controlled, resulting in out-of-range values (e.g., cost of new car: −100), impossible data combinations (e.g., Type of Vehicle: Car, Category- two wheeler: Yes), missing values, etc. Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running an analysis [Aurangzeb Khan et al., 2009]. There may be any situation arise when our dataset possess the following error;

- Incomplete: Lacking attribute values, lacking certain attributes of interest, or containing only aggregate data
    - e.g., Color of Car = " "
- Noisy: Containing errors or outliers
    - e.g., kilometer Run = "-100"
- Inconsistent: Containing discrepancies in codes or names
    - e.g., Number of year old = "05" year of Manufacturing of car = "2014"
    - Category of vehicle = "car", Fuel used: milk.

4.3. Algorithm Used

4.3.1. The K-Means Algorithm

The K-Means method is one of the most widely used clustering algorithms, drawing its popularity from its speed in practice. Recently, however, it was shown to have exponential worst-case running time. In order to close the gap between practical performance and theoretical analysis, the K-Means method has been studied in the model of smoothed analysis [Jin et al., 2006].

The K-Means algorithm is a simple iterative method to partition a given dataset into a user specified number of clusters, k. This algorithm has been discovered by several researchers across different disciplines. A detailed history of K-Means along with descriptions of several variations is given in Gray & Neuhoff (1998), provides a nice historical background for K-Means placed in the larger context of hill-climbing algorithms.

The algorithm operates on a set of d-dimensional vectors, $D = \{x_i | i = 1, \ldots, N\}$, where $x_i \in \Re^d$ denotes the ith data point. The algorithm is initialized by picking k points in _d as the initial k cluster representatives or "centroids". Techniques for selecting these initial seeds include sampling at random from the dataset, setting them as the solution of clustering a small subset of the data or perturbing the global mean of the data k times. Then the algorithm iterates between two steps till convergence:

Step 1: Data Assignment. Each data point is assigned to its closest centroid, with ties broken arbitrarily. This results in a partitioning of the data.

Step 2: Relocation of "means". Each cluster representative is relocated to the center (mean) of all data points assigned to it. If the data points come with a probability measure (weights), then the relocation is to the expectations (weighted mean) of the data partitions.

The algorithm converges when the assignments (and hence the c_j values) no longer change.

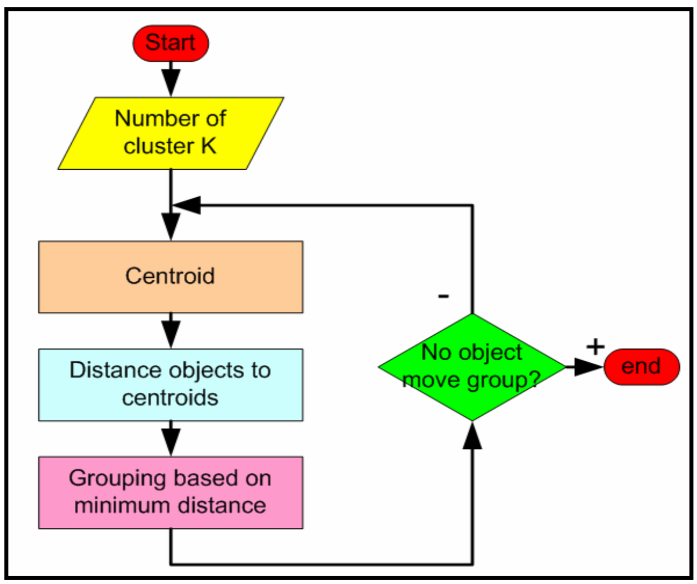

Figure 4 – K-Mean Algorithm Flow Chart

The algorithm execution is visually depicted in figure 4. Note that each iteration needs *N*×*k* comparisons, which determines the time complexity of one iteration. The number of iterations required for convergence varies and may depend on *N*, but as a first cut, this algorithm can be considered linear in the dataset size. One issue to resolve is how to quantify "closest" in the assignment step. The default measure of closeness is the Euclidean distance, in which case one can readily show that the non-negative cost function,

$$\mathop{}_{i=1}^{N}\left(argmin \left\|x_i - c_j\right\|^2\right)$$

"j" will decrease whenever there is a change in the assignment or the relocation steps, and hence convergence is guaranteed in a finite number of iterations. The greedy-descent nature of K-Means on a non-convex cost also implies that the convergence is only to a local optimum, and indeed the algorithm is typically quite sensitive to the initial centroid locations.

4.4. Cluster Analysis

Cluster analysis is a multivariate procedure for detecting natural groupings in data. The grouping is based on the scores of several measures (e.g. Service Kilometer and Service Cost). Vehicle category/ model [Index] Manufacturing year [Index].

4.4.1. Goals in Conducting Cluster Analysis

Elements within a group should be as similar as possible.

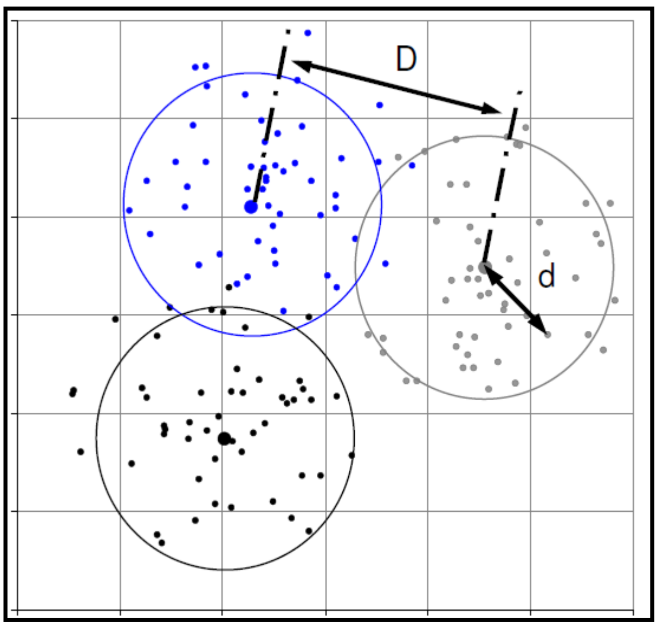

Figure 5 – Elements within a Group

- Distance d should be small.
- Similarities between the groups should be minimal.
- Distance D should be large

4.4.2. Features

Because all information is used for grouping, cluster analysis is more objective than just a subjective impression. There is no optical illusion.

4.4.3. Concepts of Cluster Analysis

Key steps in using a cluster analysis

(i) Measure of distance or similarity between objects (also called proximity measure)
   - Depends on type of data: interval, counts, binary
   - Distance: geometrical measure. Similarity: content-related measure

(ii) Formation of Clusters
   - Calculation of proximity matrix
   - Many different procedures: Hierarchical / non-hierarchical, agglomerative / divisive etc.

(iii) Tools / Criteria for Determining the Number of Clusters
   - Criteria (on basis of maintenance cost, depreciation cost etc).

(iv) Display and Save Cluster Membership
   - Show the results in the cluster graphical form.

(v) Interpretation of Clusters
   - Taking into account means (possibly variances) of cluster members

Security Warning Certain content in the database has been disabled Options...

Veh_Brand	Veh_Model	Reg_Ser_Kil	Reg_Ser_Co	Manu_Year	Manu_Ro
TATA	VISTA	2000	3000	2011	52
HYUNDAI	SANTRO	3000	1500	2008	32
MARUTI	ALTO	2500	1000	2009	27
TATA	VISTA	8000	4000	2010	45
TATA	VISTA	8500	3200	2010	25
TATA	INDICA	2800	1800	2007	35
TATA	VISTA	9200	4200	2009	15
HYUNDAI	I-10	2200	2200	2010	42
TATA	VISTA	9600	6000	2012	22
TATA	VISTA	10000	8000	2012	2

Figure 6 – Interpretation from Database

4.5. Result Set

4.5.1. K-Mean Maintenance Analysis

By analyzing kilometers run, costs and other data for every vehicle, we can determine efficiency and report up the chain of command if things start to slip or if we see where improvements can be made. Every penny counts and we'll help make sure you get the most from every one of them.

Career Tasks

- Collect and maintain maintenance performance data for use in analysis reports, studies and in identification of optimum critical replacement point.
- Assemble information by extracting and tabulating maintenance data in logical presentation sequence.
- Use statistical techniques; measure significance of correlations and trend analyses; and compare the maintenance performance to plans, schedules, and in calculating the optimum replacement point.
- Identify and assist in analysis and study of materiel deficiencies, maintenance cost and trends and deviations from standards

So, by analyzing the vehicle performance historical data by the help of the K-Mean algorithm it provides us the results in the form of cluster. It provides the results in the clustered colored form by dividing the whole of the vehicle performance in the three different colour clusters and these clusters represent the different maintenance stages as Low Maintenance, Medium Maintenance, High Maintenance so that the end user can take an effective quality decision without any comparison or typical manual value analysis.

4.5.2. Depreciation Analysis

One of the prevailing wisdoms floating around the internet is that your best value when buying a car is to get one that is only 2 or 3 years old. The idea is that you avoid some of the heaviest depreciation years in the beginning but the car should still have a number of good years left with minimal repairs. I'm not going to try to come up with a list of things you have to do to save money on your next car – there are many different situations out there and the fact is that a lot of people don't necessarily want to save as much as possible on cars – they just want to get the best value for the money they are willing to spend – not everybody is flat broke. The important thing is to consider these factors and include them in your calculations if you think they apply to you.

Car Depreciation Amounts per Year

I did a bit of research on this and came up with a general rule that a car will depreciate about 20% per year. Now, I've seen some ridiculous depreciation estimates like 30-40% as soon as you drive it off the lot. All I can say is that if you know someone who is willing to sell their car for a huge discount after driving it off the lot then buy it.

I did a few calculations and I have to admit the results surprised me – I thought that a new car buyer who kept the car a long time would pay less (or the same) depreciation than someone who bought lightly used cars but didn't keep them very long. In fact, I was wrong.

A. Scenario I

- Depreciation rate is 20% per year.
- New car is Rs. 5,40,000.
- New car buyer keeps the car for 9 years.
- Used car buyer buys the car when 3 years old and sells at 6 years old – he does this 3 times.
- New car buyer pays Rs. 62748 more in depreciation (16% more).

This scenario doesn't include potential repair costs which should be higher for the new car buyer – but it's a moot point since the new car buyer has already lost the competition in the depreciation category.

B. Scenario II

- Depreciation rate is 20% per year.
- New car is Rs 5,40,000.
- New car buyer keeps the car for 10 years.
- Used car buyer buys the car when 2 years old and sells at 4 years old – he does this 5 times.
- Used car buyer pays Rs. 144058 more in depreciation (29% more).

That's the result I was looking for – of course this scenario is a bit silly. I doubt there are very many people who buy 2 year old cars and only keep them for 2 years. This scenario would need an estimate for repair costs as well since the new car buyer would have more of them in the end.

C. Scenario III

- Depreciation rate is 20% per year.
- New car is Rs. 5,40,000.
- "Buy and hold" new car buyer keeps the car for 9 years.
- "Market timer" new car buyer keeps the car for 3 years and then buys another new car. He does this 3 times in total.
- "Market timer" pays Rs. 3,24,000 more in depreciation (70% more).

I'd be remiss if I didn't point out the extremely obvious comparison of two new car buyers – one keeps the car for a long time whereas the other keeps trading it in every 3 years. Well guess what? The guy who keeps the car for a long time pays way less depreciation. Of course his repair bills will be higher but it's hard to believe that they will exceed that amount of depreciation difference.

This is mainly depending on our quality decision making – how much qualitative decision we take and whole of the cost-efficiency and reliability depends on our decision making. You can research which cars are more reliable and buy one of those. You will probably pay more for that car than one that is less reliable so it could be a trade off. I don't have any good advice here except to say before buying any car a few steps towards researching the background details by system can save cost as well as can give better performance. I hate paying for car repairs to the point where I might actually be better off just buying new cars and only keeping for 5-6 years just to avoid the aggravation of it all.

4.5.3. Optimum Replacement Policy

In the calculation of optimum replacement point we have to consider and analyze both the factors as in terms of the maintenance and in terms of the depreciation also. As the maintenance factor in the starting of the vehicle is less but on the other side if we draw our attention towards the depreciation that costs falls much rapidly in the starting but as the vehicle goes older it starts to fall slowly so it costs lesser.

So for the qualitative decision we have to find the optimum point as in the starting of the vehicle its costs downfall in terms of depreciation cost and in terms of the maintenance its goes higher and increase as much rapidly as it goes older. So our main purpose is to find and calculate by the means of our designed tool that optimum point at which we can replace the vehicle at that stage where we don't cost much higher for the depreciation as well as also for the maintenance. So, here now we analyze for the depreciation and maintenance for the clear explanation and understanding of the concept simultaneously.

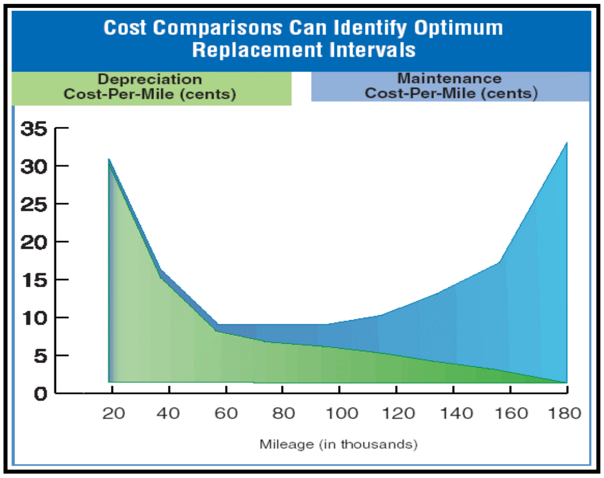

Figure 7 – Cost Comparison-Optimum Replacement Policy

This graph represents cost-per-mile in cents for depreciation and maintenance of a typical fleet sedan averaging 20,000 miles per year. Use this chart to help figure out the optimum replacement interval. For example, replacing vehicles at 60,000 miles means four replacement vehicles in a 12-year period. Fleets replaced on this schedule will incur the higher first-year costs four times and will not take advantage of the lower fourth-year costs (60,000-80,000 miles) at all. Replacing vehicles at 80,000 miles means three replacement vehicles in the same 12-year period. Fleets replaced on this schedule will incur the higher first-year costs only three times, and will take advantage of the lower fourth-year costs three times.

V. Conclusion

The research presented in this research work sought to contribute to the field of real-time fleet management systems in a way that can provide tangible results as well as practical assistance to those who will implement such systems for the analysis of maintenance and replacement policies of different vehicles. Decision for purchase of new vehicle, analysis of performance of vehicles, evaluation and comparing the past record of vehicle depend upon for a comprehensive and coherent theoretical and practical understanding of such problems.

The starting point of this research was the need for effective decision management at the time of purchasing of new vehicle as at the time of purchase of new vehicle there are a number of options are available in the market so it becomes too much confusing for the purchaser to choose the most appropriate and effective vehicle so that effective decision can be taken the system had been designed in this research work was evaluated and verified through the opinion of experts, conformity opinions of existing fleet owners and through two different freight carriers

5.1. System Evaluation through Real-Time Testing

System's evaluation through real-time testing is done by results as provided by system by comparing the Manual real time testing for the purpose of Cross-Verification. All the hypotheses were validated and the results showed that the performance of the system is statistically significant when compared with the current methods and helps the end-user in the quality decision making. Consequently, it can be claimed that the proposed system improves customer service and decreases the direct as well as indirect costs when implemented in replacement policies as well as at the time of purchasing of new vehicles.

5.2. Future Research Issues

At the same time, researchers in this area may use the proposed analysis of maintenance and replacement policies in fleet management system to assess existing systems. Although the field of analysis of maintenance and replacement policies fleet management is still immature (most systems that are being used cannot analyze in a systemic fashion of unforeseen events), we would like to advocate that this system may be employed as a tool by fleet owners to revisit their design specifications, and improve them based on the suggested prescriptions. Unfortunately, no other research, in the best knowledge of the author, deals with the generic issue of analyzing the factors if the vehicles is used in unfamiliar environment as like: on rural rough roads, on hilly areas, in rush areas or any other specific attributes, in order to evaluate whether the proposed system may be utilized as a benchmarking tool. This prospect is highlighted as an opportunity for future research. A second area of research is along with the generic attributes as used in this research another attribute Mileage analysis of the fleet can also be included for the better analyzing of the vehicles attributes and more effective decision can be achieved.

References

[1] Michael J. Berry & Gordon S. Linoff (1988), "E-Book Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management", Jain AK, Dubes RC (1988) "Algorithms for Clustering Data". *Prentice-Hall, Englewood Cliffs*.

[2] RM. Gray & DL. Neuhoff (1998), "Quantization", *IEEE Transactions on Inform Theory*, Vol. 44, No. 6, Pp. 2325–2384.

[3] D. Pyle (1999), "Data Preparation for Data Mining", *Morgan Kaufmann Publishers*, Los Altos, California.

[4] J.W. Creswell (2002), "Research Design – Qualitative, Quantitative and Mixed Method Approaches", Second Edition, *Sage Publications*, Thousands Oaks CA, Dawson C. W., The Essence of Computing Projects – A Student's Guide, *Prentice Hall, Harlow UK*, 2000.

[5] N.K. Malhotra, D.F. Birks & Peter Wills (2003), "Marketing Research: An Applied Approach", *Pearson, Harlow, Prentice Hall*.

[6] Y. Elovici, A. Kandel, M. Last, B. Shapira & O. Zaafrany (2004), "Using Data Mining Techniques for Detecting Terror-Related Activities on the Web", *Journal of Information Warfare*, Vol. 3, No. 1, Pp. 17–29.

[7] Jiawei Han & Micheline Kamber (2006), "Data Mining: Concepts and Techniques", 2nd ed., *The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor Morgan Kaufmann Publishers*, ISBN 1-55860-901-6.

[8] J.A. Harding, M. Shahbaz, Srinivas & A. Kusiak (2006), "Data Mining in Manufacturing: A Review", *Journal of Manufacturing Science and Engineering*, Technical Papers, Vol. 128, No. 4, Pp. 969–976.

[9] R. Jin, A. Goswami & G. Agrawal (2006), "Fast and Exact Out-of-Core and Distributed K-Means Clustering", *Knowledge and Information Systems*, Vol. 10, No. 1, Pp. 17–40.

[10] A. Kusiak (2006), "Data Mining: Manufacturing and Service Applications", *International Journal of Production Research*, Vol. 44, No. 18–19, Pp. 4175–4191.

[11] Nassar Khaled (2007), "Application of Data-Mining to State Transportation Agencies' Projects Databases", *ITcon*, Vol. 12, Pp. 139–149.

[12] Chiheb Saidane, Man-Tak Shing & Arijit Das (2007), "Decision Support System for the Management of an Army's Tracked and Wheeled Vehicle Fleet", *Naval Postgraduate School Monterey*, CA 93943-5000, Performing Organization report number 9. Sponsoring / Monitoring

[13] Aurangzeb Khan, Khairullah Khan & Baharum B. Baharudin (2009), "Frequent Patterns Minning of Stock Data using Hybrid Clustering Association Algorithm", *International Conference on Information Management and Engineering*, Pp. 667–672.

[14] J. Gardner & L. Xiong (2009), "An Integrated Framework for De-Identifying Unstructured Medical Data", *Data & Knowledge Engineering*, Vol. 68, No. 12, Pp. 1441–1451.

[15] http://www.worldometers.info/cars/

[16] http://oica.net/