International Journal of Recent Advances in Multidisciplinary Topics
Volume 2, Issue 2, February 2021
https://www.ijramt.com | ISSN (Online): 2582-7839

Quality Checking System in Automobile Engines using Data Mining Algorithms

S. Devanand*
SRM Institute of Science and Technology, Ramapuram, India

***Abstract*: The Automobile industry is a billion-dollar industry. New vehicles are continuously designed while the existing ones are constantly being made better. But at its core, the entire industry runs, essentially, on engines. Engines are some of the most used pieces of technology in the world. They move the earth and the economy, thus driving towards a better tomorrow. In order to tap the maximum potential of an engine, the importance of the parameters that contribute to the working, running and maintaining the life and the efficiency of the said engine is paramount. Maintaining this efficiency of work is done by the mechanic who repairs and restores the shape of the car back to its prime condition. But, due to the nature of humans, the task of assessing the problem usually takes away precious time which can be used to enrich productivity. Thus comes the true purpose of this system. Keeping in mind the "Quality-Checking" approach where accuracy is critical, this system will compare the current conditions of the engine with the factors of the same parameters using an innovative method of assessing the quality of engine. Thus using a data mining algorithm, the factors contributing to inefficiencies of an engine and the problems lying within can be easily brought to light. This research paper, however proposes only a model for the above mentioned system. This paper highlights the usage of three data mining algorithms and compares the predictions generated by them.**

***Keywords*: Efficiency, Quality, Random Forest, Decision Tree, Four stroke engine.**

1. Introduction

Quality is an important factor when it comes to any product or service. With the high market competition, quality has become the market differentiator for nearly all products and services. Quality control is important in putting together a successful business that delivers products that meet or exceed customers' expectations A quality system supported a recognized standard, like ISO 9001 published by the world organization for Standardization, provides a robust foundation for achieving a good range of marketing and operational benefits. Therefore, all manufacturers and repair providers out there constantly search for enhancing their product or the service quality. In order to take care of or enhance the standard of the offerings, manufacturers use two techniques, internal control and quality assurance. These two practices confirm that the top product or the service meets the standard requirements and standards defined for the merchandise or the service. ISO (International Standards Organization) is one among the prominent bodies for outlining quality standards for various industries. Therefore, many organizations attempt to adhere to the standard requirements of ISO. In addition, thereto, there are many other standards that are specific to varied industries.

Quality Assurance is a broad practice performed in lieu of assuring the standard of products or services delivered. There are many differences between internal control and quality assurance. In quality assurance, a continuing effort is always exercised to strengthen the standards of the organization. Therefore, continuous improvements are expected in quality functions within the company.

For assessing the condition of any given engine, it can be done with the help of 3 quality checking ways. One Condition will be "PRIME" which states the perfect conditions that ultimately constitute a well-rounded maintained engine. The other will be "PASSIVE" which rates the engine, in a nutshell, acceptable. Finally, the Last condition will be "REJECTED", where, due to abysmal and/or unsatisfactory performance of parts, the said engine is rated UNFIT for any use. This system is designed in Rapidminer TM along with the necessary packages. This, at its peak can be incorporated into an expert system that will assess the engine within mere seconds and suggest the best course of action.

2. Literature Work

This research paper was done by various other scientists purely for the sake of improving the efficiency figures of the engine and its overall operation mechanics. Eminent experts have researched and published research papers about the same. [1] HAO GANG in 2016 published a paper argues that the figures can be analysed and the predictions can be made by the method of BP Neural Network Based on artificial fish swarm algorithm. The merits of the paper include usage of Fuzzy logic, object-oriented technology and neural networks are all used together. This results in Razor sharp predictions. But the results are highly formulaic and arithmetic based. Results may be slower and harder to achieve. [2] S.N. DANDARE et al in 2014 explains the tools used in fault detection of engines by the method of Signal conditioning, Signal processing, statistical analysis and Artificial neural network. Now, since this paper uses 4 stroke bike engine and a recently made car engine, the Methods portrayed and conclusions drawn are very relevant.

*Corresponding author: devanand071@gmail.com

But this paper detects faults by using sound as a core tool. For accurate predictions, a huge repository is needed. [3] DEULGAONKAR.V. R et al expands the idea of using sound produced to diagnose and analyses predictions. The methods portrayed here in the ways of using sound are very diverse and rich. The accuracy obtained will be high. However, this paper instutionalises the ideology of using external sound as the critical tool. It needs a major repository of datasets only for one make this can be tedious. [4] DANIEL BARNICLE et al, proposes the usage of a specially constructed apparatus. Merits include usage of many nodes, which uses models interfacing wirelessly fixing OBD-II which easier access. But, it might possess delay between the engine control unit, ECU interface and the PDA module.

[5] WANG WEIJIE et al, proposed analyzing predictions using wavelet frame theory and neural networks technology. Highlights of this paper includes the characteristics values of automobile engine vibrational signal are meticulously obtained using wavelet packets. This yields high accuracy. Still, the methods used for obtaining result are highly formulaic and arithmetic based. Results may be slower and harder to achieve. [6] SADAO AKISHITA et al elucidates the theory of analyzing and processing sound produced by the engine for failure diagnosis. Here, Diagnostics is designed and mapped to an OBD-II port. This can ensure faster real time diagnostics. Accuracy will be very high. However, this paper portrays a method where a whole of sensors are connected to a drive shaft connected via the crankshaft. This means that the system can work only when the engine is running. [7] ZHENG XIAOJUN et al in 1988 proposed a crude but theoretically possible knowledge based system for diagnosis of automobile based engines. It uses the method of Hierarchical diagnosis principle. The greatest merit of this system is that this system can be eventually modelled into an expert system which has a wide variety of applications. But, extensive and diversified repository and datasets are required for greater accuracy which is laborious.

The purpose of conducting literature work has been well realized. The parts of engine and the importance of each component's role has been highlighted. The importance of precision in the generation of predictions and the ways to obtain it have been elucidated. Also, the usage of various algorithms in the research papers have been noted. The same ideologies have also been imported, reassessed and incorporated appropriately.

3. Proposed Work

In an engine [8], there are Innumerable parts that contribute greatly to the smooth running of an engine. The working of a four stroke engine and its most critical parts are as follows.

1) Stages of a four stroke engine [9]

Intake stroke: The intake stroke lets air and fuel into the combustion chamber. The piston descends within the cylinder bore to evacuate the combustion chamber. When the inlet valve opens, air pressure forces the air-fuel charge into the evacuated chamber. As such, the combustible mixture fills the chamber. Compression stroke: After the intake stroke, both inlet and exhaust valves are made to close. The inertial action of the crankshaft successively lifts the piston which compresses the mixture. The ratio of the combustion chamber volume before and after compression is named the compression ratio. Typically, the worth is approximately 9:1 in spark ignition engines and 15:1 in diesel engines.

- *Power stroke:* When the piston ascends and reaches top dead centre, an electric current ignites the spark plug and as the mixed gas burns, it expands and builds pressure in the combustion chamber. The resulting pressure pushes the piston down with several plenty of force.
- *Exhaust stroke:* During the exhaust stroke, the inlet valve remains closed whilst the exhaust valve opens. The moving piston pushes the exhumed fumes through the now open exhaust valve and another thus another cycle starts again.
- *Parts of an engine: Engine:* An engine or a motor is a machine designed to convert one form of energy into mechanical energy. More specifically, it is a machine that converts linear motion to a rotational motion by using heat energy. As you can probably imagine, it does a tremendous amount of work for us so that we can extract energy from it. Naturally it will demand periodic maintenance which must be carried out to prolong the life and efficiency of the engine.
  Though there are lots of parts that contribute towards the proper running of the engine, parts and/or entities that are important are as follows [10] [11];
- *Spark Plug:* A sparking plug may be a device used for delivering current from an ignition to the combustion chamber of an engine to ignite the compressed fuel/air mixture by an electric spark. Spark plugs are what supply the spark that ignites the air/fuel mixture, creating the Controlled Explosion which makes the engine to produce power. These small but simple plugs create an arc of electricity across two electrodes which are not touching, but close enough together that electricity can jump the gap between them.
- *Loss of Compression:* Compression loss is a result of a leak in one or more of the cylinders caused by normal engine wear and tear. Compression loss in the cylinder of the engine can cause misfiring and poor vehicle performance
- *Ignition Timing:* In an engine, Ignition timing refers to the release of a spark in the combustion chamber near the end of the compression stroke. Setting the right ignition timing is crucial within the performance of an engine.
- *White Smoke:* White smoke from the exhaust is when the engine is running hotter than usual. One of the main causes of white exhaust smoke is a cracked cylinder head, a cracked engine block, or head gasket failure caused by overheating. A cracked head may allow coolant to leak into one or more cylinders or into the combustion chamber of the engine, which when

burned, causes white smoke.

- *Dark Smoke:* The black smoke is composed primarily of elemental carbon from incomplete combustion of diesel fuel and traces of engine lubricant. Incomplete combustion occurs when there is an overabundance of fuel and insufficient amount of oxygen. Over-fuelling is the primary cause of black smoke from the exhaust of a heavy-duty diesel engine.
- *Catalytic Converter:* A catalytic converter is an exhaust emission control device that reduces toxic gases and pollutants in exhaust gas from an engine. Catalytic converters are usually used with combustion engines fuelled by either gasoline or diesel
- *Cooling system:* A vehicle's engine-cooling system serves not just to stay the engine cool, but to also keep its temperature warm enough to make sure efficient, clean operation. System components include a radiator to dissipate heat, a lover or fans to make sure adequate airflow for radiator cooling, a thermostat valve that opens when the specified operating temperature is reached and a pump (or coolant pump) to circulate coolant through the engine, hoses and other components.

2) Grading system

This system is designed to analyse the quality of an engine. It does this by identifying the condition of parts. On complete analysis, it will compare the conditions of the said engine with 3 quality checking ways.

- *Prime:* PRIME is a condition that belongs to one of the three possible conditions. An engine will only be labelled as PRIME if and only if the performances of all specified parts are perfect in terms of life, efficiency and produce which ultimately constitutes a well-rounded maintained engine
- *Passive:* PASSIVE is a condition that belongs to one of the three possible conditions. An engine will only be labelled as PASSIVE if the performances of all specified parts have acceptable tolerance levels in terms of life, efficiency and produce.
- *Rejected:* REJECTED is a condition that belongs to one of the three possible conditions. An engine will only be labelled as REJECTED if the performances of all specified parts' tolerance levels in terms of life, efficiency and produce are unacceptable

4. Analysis Based on Metrics

In measurement of a group, [12] accuracy is closeness of the measurements to a selected value, while precision is that the closeness of the measurements to every other. Given a group of knowledge points from repeated measurements of an equivalent quantity, the set are often said to be accurate if their average is on the brink of truth value of the number being measured, while the set are often said to be precise if the values are on the brink of one another. within the first, more common definition of "accuracy" above, the 2 concepts are independent of every other, so a specific set of knowledge are often said to be either accurate, or precise, or both, or neither.

[13] A measurement system are often accurate but not precise, precise but not accurate, neither, or both. For instance, if an experiment contains a scientific error, then increasing the sample size generally increases precision but doesn't improve accuracy. The result would be a uniform yet inaccurate string of results from the flawed experiment. Eliminating the systematic error improves accuracy but doesn't change precision. A measurement system is taken into account valid if it's both accurate and precise. Related terms include bias (non-random or directed effects caused by an element or factors unrelated to the independent variable) and error (random variability). In order to see the standard of an engine, an accurate dataset is of paramount importance. Hence the dataset must be set in accordance to the tolerances specified.

5. Results and Discussion

A dataset may be a collection of knowledge. Within the case of tabular data, a knowledge set corresponds to at least one or more database tables, where every column of a table represents a specific variable, and every row corresponds to a given record of the info set in question. The info set lists values for every of the variables, like height and weight of an object, for every member of the info set. Each value is understood as a datum. Data sets also can contain a set of documents or files. It's an excel sheet which has all the specified columns needed to perform the required operations. This dataset is critical because it has holds the info without which no action/process/identification/verification/determination/calculation and analysis, amongst others, can happen. This dataset named "Base Repository" is an excel sheet containing 9 columns and 100 rows. The rows display the info with reference to the columns. For the primary 3 columns, only PRIME or REJECTED will apply as PASSIVE has no scope there. These columns indicate the condition of the required a part of the engine. The columns of the dataset are as follows;

- *Condition:* This column states one among the three quality conditions, PRIME, PASSIVE and REJECTED additionally, this may also function the "Label" attribute inside Rapidminer.
- An algorithm can be defined as finite sequence of well-defined, computer-implementable instructions, typically to unravel a category of problems or to perform a computation. Algorithms are always unambiguous and are used as specifications for performing calculations, processing, automated reasoning, and other tasks. Algorithm, in its most elementary definition, are often described because the step-by-step method towards solving a drag. Our system here uses 4 algorithms for two different purposes. The algorithms are as follows.
- *Naïve Bayes:* Naïve Bayes classifiers are a family of straightforward "probabilistic classifiers" supported applying Bayes' theorem with strong (naïve) independence assumptions between the features. They're among the only Bayesian network models, but

including Kernel density estimation, they will achieve higher accuracy levels.

- Naïve Bayes classifiers are highly scalable, requiring variety of parameters linear within the number of variables (features/predictors) during a learning problem. Maximum-likelihood training are often done by evaluating a closed-form expression, which takes linear time, instead of by expensive iterative approximation as used for several other sorts of classifiers.
- *Decision tree [14]:* a choice tree may be a flowchart like tree structure, where each internal node denotes a test on an attribute [meaning it specifies an "if or else" situation], each branch represents an outcome of the test, and every leaf node (terminal node) holds a category label. A choice tree is drawn the wrong way up with its root at the highest. The top of the branch that does not split anymore is that the decision leaf. Decision trees are commonly utilized in research and operations management. If, in practice, decisions need to be taken online with no recall under incomplete knowledge, a choice tree should be paralleled by a probability model as a most suitable option model or online selection model algorithm. Another use of decision trees is as a descriptive means for calculating conditional probabilities.
- *Gradient Boosted Trees [15]:* Gradient boosting may be a sort of machine learning boosting. It relies on the intuition that the simplest possible next model, when combined with previous models, minimizes the general prediction error. The key idea is to line the target outcomes for this next model so as to attenuate the error. If a little change within the prediction for a case causes no change in error, then next target outcome of the case is zero. Changing this prediction doesn't decrease the error.
- *Random Forests [16]:* Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a mess of decision trees at training time and outputting the individual trees on all possible combinations. Random forests generally outperform decision trees, but their accuracy is less than gradient boosted trees. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is less than gradient boosted trees. However, data characteristics can affect their performance.

Table 1
Comprehension of readings obtained

Component	Prime Range	Passive Range	Rejected range
Spark plug	1		0
Loss of Compression	0		1
Ignition Timing	0		1
White Smoke[ppm]	0-10	10-500	>500
Dark Smoke [ppm]	0-10	10-500	>500
Catalytic converter [ppm]	<50	50-100	>100 ppm
Cooling System [^{o}c]	<=199	200-210	>210

Table 2
Tabulation of all algorithm's performance

S. No.	Algorithm used	Accuracy obtained
1	Random Forest	90.0%
2	Gradient Boosted Trees	86.0%
3	Decision Trees	83.3%

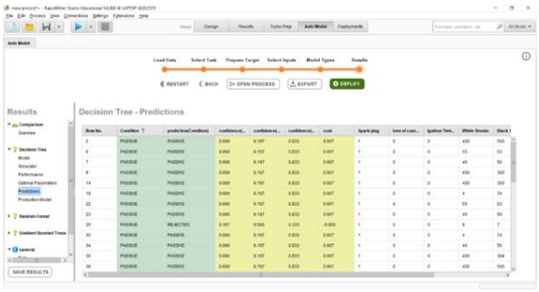

Fig. 1. Decision Tree Predictions

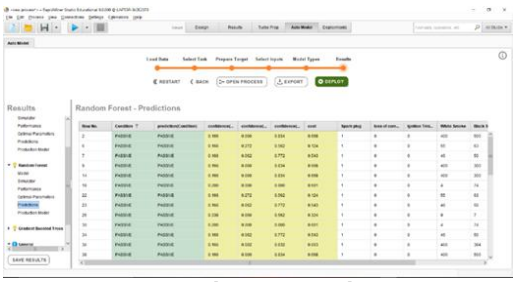

Fig. 2. Random Forest Predictions

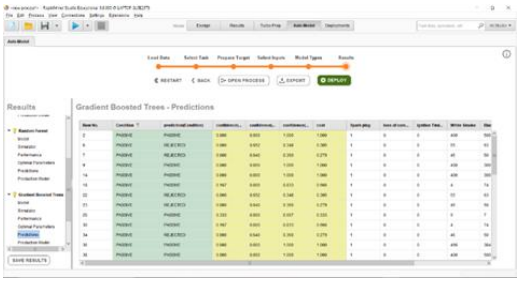

Fig. 3. Gradient Boosted Tree Predictions

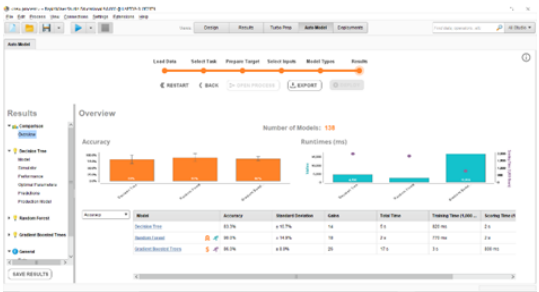

Fig. 4. Performance of each algorithm

- *Implementation:* The software "Rapidminer" is employed for creating predictions. Rapidminer here possesses a dual purpose; to calculate the accuracy of the dataset "Base Repository" with the assistance of the algorithms aforementioned. For Predictions, anybody algorithm is enough. 2 more algorithms are present to assess and compare which algorithm suits

the dataset better i.e. which algorithm paired with the dataset makes more accurate predictions.

- *Accuracy Calculation:* First the dataset is imported. Then the label attribute is assigned to the column label. Dataset and validation tools are imported.
- *Obtaining Predictions:* Auto Model is an extension to Rapid Miner Studio that accelerates the method of building and validating models. Auto Model addresses three large classes of problems: Prediction Clustering Outliers within the Prediction category, it can solve both classification and regression problems. Auto Model helps to gauge data, provides relevant models for the answer and helps to match the results for these models, once the calculations are completed.

Auto Model is employed here to form predictions simultaneously with Decision Tree, Random Forest and Gradient Boosted Trees.

6. Conclusion

The idea of using data mining algorithms to develop a model on identifying the quality of an engine have been made. An innovative grading system has been constructed so as to provide understandable but deep information of the condition of the engine. This paper highlights the usage of datamining algorithms. The predictions delivered by exercising the brute forcing of the algorithm have been documented and their results have been comparatively tabulated. Quality checking plays a huge role in customer retention, building a good brand image and delivering what is paid for. Using a repository/dataset and building upon it using algorithms to make predictions makes this system robust and clear cut in its procedure. This system can be easily modelled into a framework using API and ultimately be made into an expert system. Additional features such as taking into consideration of many more parts and developing a separate algorithm will make the system analysis indubitably better. Thus with the usage of data mining algorithms, namely decision tree, gradient boosted trees and random forest as shown here can be diversified into many more applications much beyond the scope of the model portrayed here.

References

[1] H. Gang "Automobile Fault Diagnosis System Based on Improved Neural Network," 2016 *International Conference on Smart City and Systems Engineering* (ICSCSE), Hunan, China, pp. 494-497, 2016,
[2] S. N. Dandare, S. V. Dudul "Multiple fault detection in typical automobile engines: a soft computing approach", WSEAS transactions on signal processing, Maharashtra, India, (2014),
[3] Deulgaonkar. V. R, Kamurukar S. P, Matlani A. G, "Review and diagnostics of noise and vibration automobiles", *International Journal of Modern Engineering Research* (IJMER), 2011.
[4] https://patents.google.com/patent/US7363129B1/en
[5] Wang Weijie, Kang Yuanfu, Zhao Xuezheng and Huang Wentao, "Study of automobile engine fault diagnosis based on wavelet neural networks," Fifth World Congress on Intelligent Control and Automation (IEEE Cat. No.04EX788), Hangzhou, China, vol. 2, pp. 1766-1770, 2004,
[6] Zheng Xiaojun, Yang Shuzi, Zhou Anfa and Shi Hanmin, "A Knowledge-Based Diagnosis System for Automobile Engines," Proceedings of the 1988 IEEE International Conference on Systems, Man, and Cybernetics, Beijing, China, pp. 1042-1047, 1988.
[7] https://en.wikipedia.org/wiki/Engine
[8] https://en.wikipedia.org/wiki/Four-stroke_engine
[9] https://www.discountrebuiltengines.com/blog/what-are-main-parts-of-an-engine.php
[10] https://www.sunautoservice.com/car-engine-parts/
[11] https://www.investopedia.com/terms/d/data-analytics.asp
[12] https://en.wikipedia.org/wiki/Data_set#:~:text=A%20data%20set%20(or%20dataset,the%20data%20set%20in%20question.
[13] https://en.wikipedia.org/wiki/Decision_tree#:~:text=A%20decision%20tree%20is%20a,only%20contains%20conditional%20control%20statements.
[14] https://en.wikipedia.org/wiki/Gradient_boosting
[15] https://en.wikipedia.org/wiki/Random_forest#:~:text=Random%20forests%20or%20random%20decision,average%20prediction%20(regression)%20of%20the