QUALITY ENHANCEMENT THROUGH APPLICATION OF CLASSIFICATION TREES[1]

Subrata Nath, Director, Darjeeling Himalayan Railways, Darjeeling, India
Purnendu Kishore Banerjee, Indian Statistical Service, New Delhi, India
Vishnuprasad Nagadevara, Indian Institute of Management, Bangalore, India

ABSTRACT

Significant efforts are being made to enhance the quality of products and services in all sectors of economy across the world. Organizations, which fail to provide quality goods and/ or services, are likely to fail. The cost of poor quality is estimated to be about 30 percent to 40 percent if the organization is operating at what is called 2 Sigma level. Classification tree method is one of the effective techniques for identifying the source of poor quality. This paper applies classification tree method to identify the major sources of poor quality in the manufacture of rail wheels at the Rail Wheel Factory (RWF) at Bangalore, India.

Keywords: *Classification trees, Quality Enhancement, Cost of Poor Quality, Rail Wheel Factory*

1. INTRODUCTION

Significant efforts are being made to enhance the quality of products and services in all sectors of economy across the world. Organizations, which fail to provide quality goods and/ or services, are likely to fail. The cost of poor quality (COPQ) is estimated to be about 30 percent to 40 percent if the organization is operating at what is called 2 Sigma level (Brue and Howes, 2006). The COPQ represents visible and invisible costs that exist in all processes. The warranty claims, maintenance costs incurred in fixing failures in the field, cost of scrap and rejects, wastage of material, labour, machine time, utilities etc. It is necessary ton identify the sources of poor quality before attempts are made to reduce the cost of poor quality.

Classification tree method is one of the effective techniques for identifying the sources of poor quality. A classification tree is an algorithm usually presented as a tree diagram that classifies an object into predefined classes and uses the classification rules to predict the class membership of a new object. While other techniques such as discriminant analysis, logistic regression, artificial neural networks etc. could also be used for classification and prediction purpose, a classification tree differs in the way it models relationships between class membership and the explanatory variables (Khoshgoftaar, 2000). A classification tree has an additional advantage that can be readily interpreted. Various algorithms have been developed to build the classification trees. ID3 algorithm is used to partition recursively the tuples of training data set using an entropy-based criterion (Quinlan, 1986). Other methods are CHAID which determines the criteria for splitting based on χ^2 tests and a predetermined value of statistical significance (Kass, 1980). Classification and Regression Trees (CART) methodology builds a parsimonious tree from continuous ordinal predictors by first building a maximal tree and then pruning it to the desired level (Breiman, Friedman, Olshen and Stone, 1984). While the classification tree technique is effective in using categorical variables (predictors) for classifying and predicting the class membership of the dependent variable, unbalanced data sets create problems in training the model (Kitchenham, 1998). The technique can be made effective even for unbalanced datasets through various methods such as under-sampling and over-sampling as well as error weighting (Anuj kumar and Nagadevara, 2006)

This paper applies classification tree method to identify the major sources of poor quality in the manufacture of rail wheels at the Rail Wheel Factory (RWF) at Bangalore, India. The wheel production at the RWF is sequential and every single wheel passes through a series of work centers where various production activities are carried out. Quality checks are also carried out at various production stages. While the wheels are either “passed” or “rejected” based on various quality parameters, the actual causes for such rejections were not easily identified.

[1] Published in the European Journal of Management, Vol 8, No. 1, 2008, pp104-109

The main objective of this study is to identify the causes for poor quality or rejection of wheels. The second objective is to estimate the gains in terms of increased revenue by reducing the cost of poor quality.

The RWF, established in 1984, produces cast wheels for Freight Car Stocks (BOXN Wheels) and for Passenger Car Stocks (ICF Wheels). It is one of the very few wheel production plants in the world that employs pressure-casting technology. Between 2000 and 2006, BOXN wheels accounted form more than 82 percent of the total production while ICF wheels accounted for only 12 percent and remaining are produced on made-to-order basis. The annual production of good wheels is about 1, 15,000 and it is expected to grow at 6-8% per annum in next five years. The production process is sequential and quality control checks are carried out at different stages of the process. The plant employs pressure-casting method for pouring of molten metal into specially designed metal moulds. The upper part of the mould is known as '*Cope'* whereas the lower part is called '*Drag'*. The wheels are tested using Ultrasound (UT) and Magna-glow process. Once a wheel passes these two stages, it is tested for its plate thickness and hardness (measured as BHN). After the hardness test, the passed wheels are despatched for use at different wagon and coach manufacturing factories. The failed wheels go back to the initial stage of production for re-use. The Unit captures all process data on-line and the data is processed to generate various Management Information Reports (MIS Reports).

Analysis of frequency of wheels finally passed indicates that over the years, the percentage of passed BOXN wheels has increased marginally. However, level of rejection in terms of percentage is about 8%. For every one percentage point reduction in rejection of wheels, annual savings that can be accrued is about Rs. 85 Lakhs (8.5 Million Rupees).

Table 1 presents the percentage of wheels passed for both BOXN and ICF wheels.

	Percentage of Wheels Passed				Number of all
Year	BOXN	ICF	Others	Total	wheels produced
2000	87.6	62.6	65.9	84.6	77,300
2001	88.9	70.6	67.7	85.5	107,092
2002	86.5	79.2	77.7	84.4	115,998
2004	83.8	78.3	76	82.6	132,987
2005	83.7	78.8	78.8	83.3	114,720
2006	84.2	80.7	61.9	83.5	117,240

Production figures for 2003 were not available because the RWF was modernized during the year

The cost of rejection computed on the basis of input cost of energy, manpower and machine hours, is about Rs. 8000 per wheel which aggregates to about Rs. 68 million per annum. A reduction of one percentage in rejection level shall not only result in a savings of about Rs. 8.5 million per annum, but also it will increase percentage of good wheels and subsequently improve total turn-over of the plant as shown in Table 2. Therefore, by reducing the rejection level, the RWF can cut costs and improve productivity and can pass on the advantage of better quality to its customers. Since it operates on transfer pricing system (not for profit), the costs and benefits thus obtained can be passed on to the customers. Therefore, it is extremely important to enhance the quality in the production of wheels at the Rail Wheel Factory, Bangalore.

Table 2. Cost of poor Quality

In Rs. Crores/Annum	8% Rejection	6% Rejection	4% Rejection
Cost of Rejection	6.8	5.1	3.4
Total Revenue	195	200	205

2. METHODOLOGY

Data on a total of 650,000 wheels which were produced between 2000 and 2006 was obtained from the MIS reports of the factory. Data was obtained on the following variables:

1. Status: The final status of the wheel showing whether the particular wheel has been passed on not. This variable is used as the dependent variable in building the classification tree.
2. Jump Value: This is the level of sulphur after chemicals are added to the molten metal in order to achieve the desired level of sulphur at the final stage.
3. Type of wheel: There are mainly two types of wheels, namely BOXN and ICF wheels. Even though the data set contained a few records of "Other" wheels, the analysis is limited to BOXN and ICF wheels only.
4. Plate thickness: the data with respect to the plate thickness indicates whether the wheel has the desired thickness or not.
5. Pour Order: This is one of the most important variables as far as the wheel data is considered. "Pour Order" depicts the order in which the molten metal is actually poured into moulds. The temperature of the molten metal is maximum at the initial stage of pouring and then it reduces gradually. According to the domain knowledge, if the temperature drops below a certain lower limit, the output quality deteriorates. However, there is no system to record the temperature of the molten metal continuously. Hence the Pour Order can be used as a proxy for the temperature.
6. Cope Number: Cope number is the serial number given to the upper part of the mould into which the molten metal is poured. There were a total of 1306 copes used in the production wheel over the past 7 years. Of these, 1302 were for BOXN wheels and the remaining for ICF wheels. The serial number of the cope provides an idea of the age of the cope. Based on the age, these copes are grouped into 5 different groups with respect to BOXN wheels and 4 different with respect to ICF wheels for the purpose of the analysis.
7. Drag Number: Drag number is the serial number given to the lower part of the mould into which the molten metal is poured. The serial number of the drag provides an idea of the age of the drag. Based on the age, these drags are grouped into 5 different groups with respect to BOXN wheels and 4 different with respect to ICF wheels for the purpose of the analysis.

After cleaning up the data and removing the incomplete and inconsistent records from the database, 546,492 records were available with respect to BOXN wheels and 77,705 records were available with respect to ICF wheels. Classifications trees were trained based on this data using the variables mentioned above. These variables are binned into appropriate groups to facilitate construction of classification trees.

3. RESULTS AND DISCUSSION

Initially, classification trees were built using the entire data set without resorting to either under-sampling or over-sampling. As mentioned earlier, the dataset was unbalanced and the minority class (rejected wheels in this case) was overwhelmed by the majority class (passed wheels). Hence, over-sampling technique was resorted to by tripling the number of records of the minority class, i.e., the rejected wheels. In addition, a misclassification penalty of 5 was imposed on the model. The classification trees were built separately for BOXN wheels and ICF wheels. The classification tree for the BOXN wheels is presented in Figure 1. Similarly, the classification tree for ICF wheels is presented in Figure 2.

Some of the rules that emerged from the classification tree for the BOXN wheels are:

Rule 1: a combination of Pour Order bin "2" and cope bin "10" will result in large percentage of defective wheels and hence such combination should be avoided.

Rule 2: a combination of Pour Order bin "1" and Cope bin other than 10 results in a large percentage of passed wheels and hence such combination should be encouraged.

It can be easily seen that Rule no. 2 above can be easily implemented by pre-programming the Pour Order and the Cope number. On the other hand, the strategy of not using Cope no. 10 is not very desirable because the factory has already invested a large amount of money in these moulds. A separate analysis is carried out to isolate the possible combinations of Pour Order (bin 2), different Cope bins (Cope bin 10 vs. other bins) and Drag bins. It is known that certain combinations of Cope and Drag work better than others. The analysis resulted in Table 3.

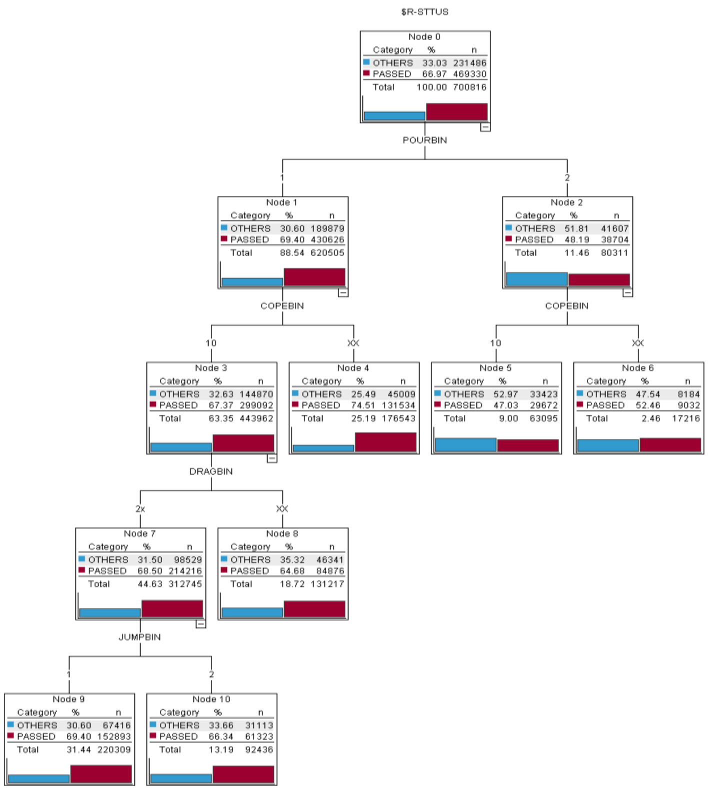

Figure 1. Classification Tree for BOXN wheels with oversampling and penalty for misclassification

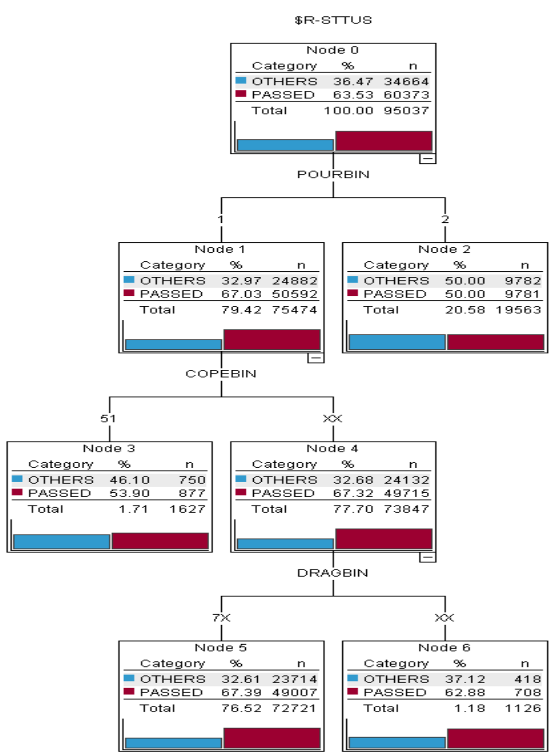

Figure 2. Classification Tree for BOXN wheels with oversampling and penalty for misclassification

It can be seen from the above table that the combination of Cope bin “other than 10” and Drag bin “2X” gives significantly higher number of passed wheels as compared to other combinations (where Pour Order Bin is 2). Hence Rule no. 1 is modified accordingly.

Table 3. Percentage of passed wheels under different combinations of Cope and Drag bins where Pour Order Bin is 2 – BOXN wheels

Sl. No.	Cope Bin	Drag Bin	No. of Wheels		Percentage Passed
			Produced	Passed	
1	10	2X	28809	21035	73.02%
2	Other than 10	2X	11714	9009	76.91%
3	10	Other than 2X	12004	8637	71.95%
4	Other than 10	Other than 2X	46	23	50.00%

In case of ICF wheels, the over-sampling was done by doubling the number of records of minority class (rejected wheels). The misclassification penalty was kept at the same level as that of the BOXN wheels. Some of the rules that emerged from the classification tree constructed are:

Rule 1: do not proceed with manufacturing of the wheel if the Pour Order bin is 2.
Rule 2: As far as possible, use the combination of Pour Order Bin 1 and Cope number other than 51.

As in the case of BOXN wheels, it may not be desirable to implement Rule no. 1. Consequently, further analysis was done to isolate the possible combinations of Cope bins and Drag bins while maintaining the Pour Order bin at 2.

Table 4. Percentage of passed wheels under different combinations of Cope and Drag bins where Pour Order Bin is 2 – ICF wheels

Sl. No.	Cope Bin	Drag Bin	No. of Wheels		Percentage Passed
			Produced	Passed	
1	51	7X	380	211	55.53%
2	Other than 51	7X	14171	9513	67.13%
3	Other than 51	Other than 2X	121	57	47.11%

As in the case of BOXN wheels, Rule no. 1 could be refined to encouraging use of Cope number “other than 51” and Drag number “7X” whenever the Pour Order bin happens to be 2.

Since, BOXN Wheels comprise major proportion of the production further analysis is carried out as per Rules obtained from the classification tree. Each of the Pour Order Bins is combined with different Cope numbers for detailed analysis. Since, Cope Series 10 comprises 70% of total population of Copes, it may not be possible to completely avoid the same. However, if Cope Series 10 is avoided in Pouring Order 1-3 and beyond 24, sufficient savings can be achieved. Total Annual Savings under such condition are expected to be about Rs. 11.2 million. Moreover, this rule can be made operational by a minor process change in Cope Scheduling and no substantial investment will be necessary.

Table 5. Percentage of Passed Wheels under different combinations of Pour Bin and Cope Number

		Pour Order Category (Pour Order)				
Pour Order Bin		1	2	3	4	No Restriction
Pour Order No.		Other than 4-24	(1-3)	(25-28)	(>28)	
Cope:	No Restriction	80.60%	78.20%	83.90%	77.30%	85.90%
	Other than 10	84.50%	79.50%	87.90%	85.20%	88.80%
Overall %		29.40%	9.90%	13.20%	6.30%	100%
Annual Production		41160	13860	18480	8820	140000
Extra Good Wheels		1605	180	739	697	
Savings/Annum (in Rs.)		11,236,680	1,261,260	5,174,400	4,877,460	

4. CONCLUSIONS

This paper applied classification tree technique to identify the major source of poor quality in the manufacture of rail wheels at the Rail Wheel Factory (RWF) at Bangalore, India. The recommendations made on the basis of the rules developed through the classification tree provide valuable insights in terms of possible combinations of Pour Order with Cope and Drag numbers. It is shown that RWF can achieve considerable savings by implementing the recommendations arising out of the rules developed through the classification tree.

5. REFERENCES

Anuj kumar and V. Nagadevara, “Development of Hybrid Classification Methodology for Mining Skewed Data Sets – A Case Study of Indian Customs Data”, Proceedings of the 4th ACS/IEEE International Conference on Computer Systems and Applications, Mar 8-10, 2006, Sharjah, UAE.

Brue, Greg and Rod Howes, Six Sigma, Tata McGraw-Hill Publishing Company, 2006

Breiman, L., J. H. Friedman, R. A. Olshen and C. J. Stone, Classification and Regression Trees”, Chapman and Hall, 1984

Kass, G. V., “An Exploratory Technique for Investigating Large Quantities of Categorical Data”, Applied Statistics, Vol 29, pp 119-127, 1980

Khoshgoftaar, T. M., Classification-tree Models of Software Quality over Multiple Releases, IEEE transactions on Reliability, Vol 49, No. 1, March 2000, pp4-11

Kitchenham, B. A., A procedure for Analyzing unbalanced Datasets”, IEEE Trans. Software Engineering, Vol 24, pp 278-301, April 1998

Quinlan, J. R., “Introduction to Decision Trees”, Machine Learning, Vol 1, pp 81-106, 1986