ISSN 2321 3361 © 2021 IJESC

Research Article Volume 11 Issue No.08

Rejection Analysis of Cast wheel by CRISP-DM and Machine Learning

S. Vijay Shandilya[1], Akshay Kulkarni[2]
MBA Business Analytics, REVA University, Bangalore, Karnataka, India[1]
Mentor, Artificial Intelligence, RACE, REVA University, India[2]

Abstract:
The pandemic, global competition, customer demand for high-quality products, a wide variety of products, shortened delivery times, and declining profit margins have all had a huge impact on the manufacturing industry. In response to these needs, various industrial engineering and quality management strategies have been formulated, such as ISO 9000, Enterprise resource planning, Business process reengineering, lean management, etc. A new paradigm in this area of manufacturing strategies is CRISP-DM. The project's work focuses on improving the quality and productivity of manufacturing company through CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology and address the Rejection analysis, which provides a framework for identifying, quantifying, and eliminating sources of variation during manufacturing and provide planned control measures to reduce wheel waste in the casting wheel production plant, improve and maintain the production performance of the wheel workshop.

Keywords: CRISP-DM, Manufacturing Industry, Product Quality, Machine learning Model. Rejection analysis.

1. INTRODUCTION:

CRISP-DM, which stands for Cross-Industry Standard Process for Data Mining, is an industry-proven way to guide data mining efforts. As a methodology, it includes descriptions of the typical phases of a project, the tasks involved with each phase, and an explanation of the relationships between these tasks. As a process model, CRISP-DM provides an overview of the data mining life cycle. The life cycle model consists of six phases with arrows indicating the most important and frequent dependencies between phases. The sequence of the phases is not strict. In fact, most projects move back and forth between phases as necessary.(Pete et al., 2000)

2. LITERATURE SURVEY

1. With the ANN model of HPDC machines, Imad ran conventional dies. Their goal is to better understand defect formation in HPDC machines. Accordingly, it was noted that most of the results corroborated existing knowledge, while others contradicted it. (Imad, 2003).
2. A model of knowledge discovery has been developed by Polczynski and Kochanski. KDAM in manufacturing has been successfully applied to optimize sand-cast moulds for gas porosity by these researchers. Because databases are so large, to get the most out of them, we must determine the patterns and structures that can be found there. This is a feature of next-generation quality technologies that can effectively utilize highly consistent, noisy, and corrupt technologies. data. A technique based on KDAM is proposed to produce cast metal parts in foundries as well. Researchers at Warsaw University are applying KDAM for:
1. A method for detecting gas porosity in steel castings.
2. Heat treatment parameters for cast iron optimization.
3. Molding sand formulation for green construction.
4. Casting properties such as strength, elongation, and hardness are predicted and improved.
3. In his explanation of the solidification process, M. Chaudhari noted that one can view the progress of freezing within the foundry and identify the last frozen zone or hot spot. By eliminating expensive and time-consuming trial runs, we were able to place and design feeders effectively and optimize yield while ensuring casting soundness. By reducing the cost of development and increasing the speed of product improvement, these temperature variations greatly improve casting quality. The significance of the gating system was observed during the experiments. Through software, an optimal riser location was determined, resulting in fewer solidification-related casting defects (no internal cavity). (M.Choudhari et al., 2013).
4. During his presentation, Dr B.Ravi listed some of the methods that foundry engineers have used to tackle the problem, such as increasing the fillet radius and thickening thin walls. He argued that such measures result in extra costs and reduced productivity due to machining. According to the author, product engineers ought to consider design for manufacturability (DFM) early instead of later, as casting suppliers do at present. According to him, castings should be made to be manufactured, not manufactured. Parts should have features that prevent defects. It results in many castings being rejected. (Ravi, 2011).
5. The classifier is built to evaluate the research we describe. We select the classifier based on two criteria based on the software engineering literature. The classifier must represent the eight most common machine learning series. The standard is designed to ensure the universality and applicability of our results. Second, the chosen classifiers should have a classifier-specific approach. (Ghotra et al., 2015)
6. The Six Sigma approach has been increasingly adopted worldwide in the manufacturing sector in order to enhance productivity and quality performance and to make the process robust to quality variations. As a result of the project, the rejection level of Ingates and Cracks after the six sigma methodology has been reduced. Several statistical tools and techniques were effectively utilized to make inferences during the project. The Six Sigma method is a project-driven

management approach based on the theories and procedures to reduce the defects for a specified process.(Manohar & Balakrishna, 2015)

7. Various parameters affecting the casting defects: Using Taguchi analysis, he determined the optimal setting for various parameters by analyzing the effect of various process parameters on casting quality. Lakshman Singaram analyzed the various levels of processes. Using a novel neural network model, he relates process conditions to quality characteristics. An optimized green sand-casting process reduced casting defects, reduced variance, and improved process performance. ANN models are used to model moisture content, green strength, and mould hardness relationships between total casting defects and corresponding casting processes. Our analysis of an ANN with varying process parameters led to close alignment of the targeted outcome with the predicted value. (Mech Journal - 1.Pdf, n.d.).

3. PROJECT METHODOLOGY

CRISP-DM Approach: The CRISP-DM model is flexible and can be customized easily. For example, if an organization aims to detect money laundering, it is likely that it will sift through large amounts of data without a specific modeling goal. Instead of modeling, work will focus on data exploration and visualization to uncover suspicious patterns in financial data. CRISP-DM allows to create a data mining model that fits any particular needs. (Shafique & Qaiser, 2014).

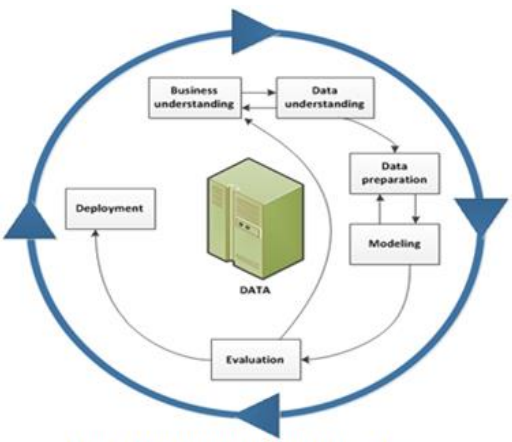


Fig 1 The data mining lifecycle.

In such a situation, the modeling, evaluation, and deployment phases might be less relevant than the data understanding and preparation phases. Cross-Industry Standard Process for Data Mining is a proven way to guide data-mining efforts. As a method, it includes descriptions of the different phases of a project, the tasks involved with each phase, and an explanation of the relationships between these tasks. The CRISP-DM architecture is shown below. With any given business problem one has to first try to understand the business process on how the wheel production is carried out. After understanding the manufacturing process there is a requirement of data collection relating to the business problem. After understanding the given data, it is required to prepare the data for further data cleaning and the other process of data preparation needs to be done for the modeling purpose. data preparation is one of the most tedious tasks which was comfortable to handle with the Microsoft Power BI tool. The raw data was imported directly into the power BI as a raw data source and the visualizations were generated using Power BI data visualizations. Variables were identified as input variables and one target variable. The rejection column becomes the target variable, and eight variables were identified as input variables like poor order, M, CR, V, ladle temperature, tube immersion temperature, S, and furnace code. After the data cleansing and the data preparation, there is a fine data set generated from the power BI and is exported for further modeling purposes. In the modeling section, the Implementation of the logistic regression model and the XG boost feature importance model is implemented. Logistic regression is a machine learning algorithm that is used for classification problems.The output of the logistic regression model is specified below. The output of the XGB feature importance classifier is also mentioned below. The accuracy of both the models is more than 90% hence with the accuracy of 90% It is concluded that the effecting variables are the basic reason for the rejection.

4. DATA UNDERSTANDING.

Adding to the foundation of Business Understanding, it drives the focus to identify, collect, and analyze the datasets that can help to accomplish the project goals. This phase also has four tasks.

Collect initial data: Acquire the necessary data and (ifnecessary) load it into analysis tool.

Table.1. Summary of the dataset

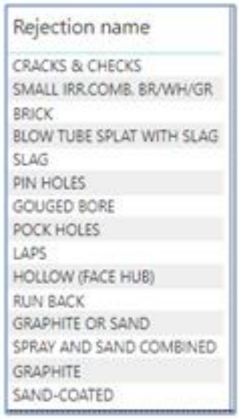
Period	Variables	From
2015-2019	All production data from wheel shop	Melting Shop

Describe data: Examine the data and document its surface properties like data format, number of records, or fieldidentities.
Explore data: Dig deeper into the data. Query it, visualize it, and identify relationships among the data.
Verify data quality: How clean/dirty is the data? Document any quality issues. Collect production data, including wheel number with heat number, rejection code, date of extraction, wheel type, and other information. The list of sample dataset columns is below.

The different variables identified in the dataset are below

Table.2. Variables identified

Column Name	Details	Data type
pour_order	Pour Number	Number (20)
description	Wheel type	varchar2 (64)
status	Status of the wheel produced	varchar2 (64)
Rejection	Type of rejection	varchar2 (64)
furnace_code	Furnace code	varchar2 (64)
shift_supervisor	shift_supervisor	varchar2 (64)
tube1_immersion1_temperature	The tube immersion temperature	Number (20)
ladle_temp	ladle_temp	Number (20)
C	Carbon	Number (20)
Mn	Manganese	Number (20)
Si	Silicon	Number (20)
Ph	Phosphorus	Number (20)
S	Sulphur	Number (20)
Cr	Chromium	Number (20)
Cu	Copper	Number (20)
V	Vanadium	Number (20)
Al	Aluminum	Number (20)
Mo	Molybdenum	Number (20)
TappedDate	Produced date	Date



The list of rejection types is as shown in the image. These are the top ten rejection types identified based on the input variable and target. The rejection names are given based on their location, situation and as well as based on the vision eg, Cracks and Slag.

5. DATA PREPARATION:

This stage is often called "data fitting" and it prepares the final data set for modeling. It has five tasks:

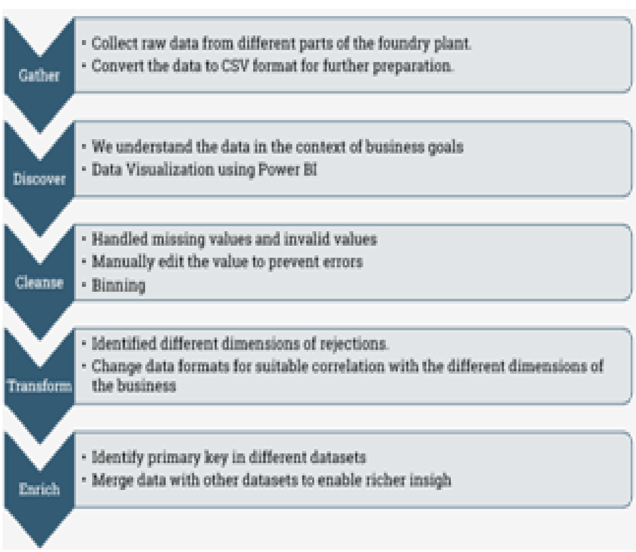


Figure.2. Data preparation process chart

5.1. Gather data: Determine which data sets will be usedand why they were included or excluded.
5.1.1. Collect raw data from different parts of the foundry plant.
5.1.2. Convert the data to CSV format for further preparation.

5.2. Discover data: Re-format data as necessary. For example, convert string values that store numbers to numeric values so that mathematical operations can be performed. Data Visualization using Power BI and EDA is performed for key insights

5.3. Clean data: This is frequently the most time-consuming task. It'll be prone to garbage-in, garbage-out if it doesn't have it.
5.3.1 Correcting, imputing, or removing erroneous values is a regular procedure during this work.
5.3.2 Handled missing values and invalid values Manually edit the value to prevent errors.

5.3.3. Binning

5.4. Transform data: Derived new attributes and dimensions that will be helpful for identifying defects. For example, derive someone's body mass index from height and weight fields.
5.4.1. Identified different dimensions of rejections.
5.4.2. Change data formats for suitable correlation with the different dimensions of the business

5.5. Enrich data: Combine data from multiple sources to create new data sets.
5.5.1. Merge data with other datasets to enable richer insight.
5.5.2. Identify primary key in different datasets

Waterfall chart shows the different wheel rejection types over % of impact on the total percentage of rejections.

Figure.3. Waterfall chart showing the impact of different rejection

6. MODELING

Build and assess various models based on several different modeling techniques. This phase has four tasks.

6.1. Select modeling techniques: Logistic regression and XG Boost feature importance technique is used for analyzing the dataset as there are one or more independent variables that determine the outcome. The intention behind using logistic regression is to find the best fitting model to describe the relationship between the dependent and the independent variable.

6.2. Generate test design: Depending on the modeling approach, there is a need to split the data into training, test, and validation sets.

6.3. Build model: As glamorous as this might sound, this might just be executing a few lines of code like.

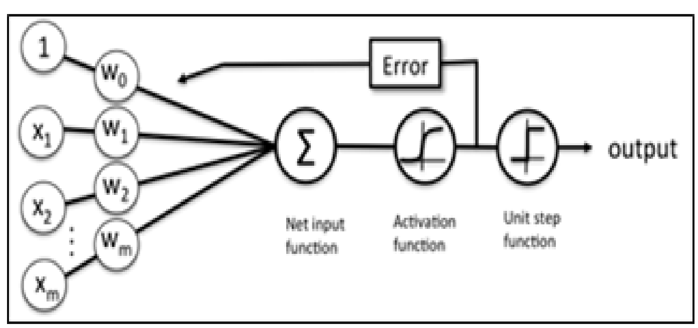


Figure.4. Logistic regression model operation

6.4. Assess model: Generally, multiple models are competing against each other, and the interpreted the model results based on domain knowledge, and the pre-defined success criteria, and the test design. In the fig3 x consisting of different variables or features (eg, pour order, temperature, chemical composition) and w is the weight vector. Imputing the values at the right place provides the accuracy of the model and probability value of the wheel status.

$$z = w_0x_0 + w_1x_1 + \ldots + w_mx_m = \sum_{j=1}^{m} w_jx_j = \mathbf{w}^T\mathbf{x}.$$

XG Boost feature importance ranking operation is performed to find the highest-ranking feature which is impacting the output. The typical output of the XGB Classifier after the model fitting is below.

```
XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
              colsample_bynode=1, colsample_bytree=1, gamma=0, gpu_id=-1,
              importance_type='gain', interaction_constraints='',
              learning_rate=0.300000012, max_delta_step=0, max_depth=6,
              min_child_weight=1, missing=nan, monotone_constraints='()',
              n_estimators=100, n_jobs=8, num_parallel_tree=1, random_state=0,
              reg_alpha=0, reg_lambda=1, scale_pos_weight=1, subsample=1,
              tree_method='exact', validate_parameters=1, verbosity=None)
```

7. RESULT AND CONCLUSION

A CRISP-DM method reduces the defects for a process by using the data and procedures. The work describes the step-by-step application of the CRISP-DM methodology to reduce the rejection rate of cast wheel production. A number of statistical tools and techniques were used to make inferences during the project. Using this methodology, the rejection levels of lap or Ingates and Cracks were reduced from 0.75% for Ingates to 0.32% and from 0.69% for Cracks to 0.23%. Sulphur when crosses more than 0.02 is more likely to be causing cracks and checks which we can say with an accuracy of 90% that this is the most impacting variable for this rejection.

8. REFFERNCE

[1]. Arasu, M. (2010). Development of quality cost system in cast iron foundries. Metal Casting Technologies - Foundry.

[2]. Krimpenis, P.G. Benardos, G.-C. V. & A. K. (n.d.). Simulation-based selection of optimum pressure die-casting process parameters using neural nets and genetic algorithms. The International Journal of Advanced Manufacturing Technology. https://link.springer.com/article/10.1007% 2Fs001 70-004-2218-0

[3].Asplund, M., Palo, M., Famurewa, S., &Rantatalo, M. (2016). A study of railway wheel profile parameters used as indicators of an increased risk of wheel defects. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 230(2), 323–334. https://doi.org/ 10.1177/ 0954409714541953

[4]. Wheels, R., & Manufacturing, A. (n.d.). RAILWAY WHEELS AND AXEL MANUFACTURING.

[5]. metallurgical aspects of rail wheel.pdf. (n.d.).

[6]. Kendall, M. G. (1948). Rank correlation methods. In Rank correlation methods. Griffin.

[7]. Tech, B., & Semester, T. H. (2010). Summer internship project. Engineering.

[8]. Rail Wheel Factory. (2016). Indian Railways. https:// rwf. indianrailways.gov.in/

[9]. Portal, I. R. (2021). Genesis of RWF. 8–11.

[10]. Krimpenis, A., Benardos, P. G., Vosniakos, G.-C., & Koukouvitaki, A. (2006). Simulation-based selection of optimum pressure die-casting process parameters using neural nets and genetic algorithms. The International Journal of Advanced Manufacturing Technology, 27(5), 509–517. https://doi.org/10.1007/s00170-004-2218-0

[11]. Sata, A., &Khire, M. . (2020). New Approach to Casting Defects Classification and Analysis Supported by New Approach to Casting Defects Classification and Analysis Supported by Simulation. October 2014.

[12]. Shafique, U., &Qaiser, H. (2014). A Comparative Study of Data Mining Process Models (KDD , CRISP-DM and SEMMA). International Journal of Innovation and Scientific Research, 12(1), 217–222. http://www.ijisr.issr-journals.org/

[13]. Pete, C., Julian, C., Randy, K., Thomas, K., Thomas, R., Colin, S., & Wirth, R. (2000). Crisp-Dm 1.0. CRISP-DM Consortium, 76.

[14]. Tiwari, R. K., & CHANDRATREY, A. (2015). An application of Taguchi's method for optimization of process parameters: a case study of cast iron sand casting. International Journal of Industrial Engineering & Technology, 5(3), 29–51. http://ijiet.com/issues/volume-5-issue-3-june-2015/

[15]. Hansen, M. R. P., & Haj-Bolouri, A. (2020). Design Principles Exposition: A Framework for Problematizing Knowledge and Practice in DSR. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12388 LNCS(June), 171–182. https://doi.org/10.1007/978-3-030-64823-7_16

[16]. Manohar, C., & Balakrishna, A. (2015). Defect Analysis on Cast Wheel By Six Sigma Methodology To Reduce Defects and Improve the Productivity in Wheel Production Plant. International Research Journal of Engineering and Technology, 2(3), 1659–1663.

[17]. Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. Procedia Computer Science, 181(2019), 526–534. https://doi.org/10.1016/j.procs.2021.01.199

[18]. Ravi, B. (2011). A Holistic Approach to Zero Defect Casti ngs. February, 1–9.

[19]. M.Choudhari, C., E. Narkhede, B., & K. Mahajan, S. (2013). Modeling and Simulation with Experimental Validation of Temperature Distribution during Solidification Process in Sand Casting. International Journal of Computer Applications, 78(16), 23–29. https://doi.org/10.5120/13607-1399

[20]. THAKARE, A. G., & TIDKE, D. J. (2013). Data Mining for Casting Defects Analysis. International Journal of Engineering Research & Technology, 2(12), 2153–2157.

[21]. Imad, M. (2003). Modelling of Porosity Defects in High Pressure Die. 1–6.

[22]. Tiwari, R. K., & CHANDRATREY, A. (2015). An application of Taguchi's method for optimization of process parameters: a case study of cast iron sand casting. International Journal of Industrial Engineering & Technology, 5(3), 29–51. http://ijiet.com/issues/volume-5-issue-3-june-2015.

[23]. Polczynski, M., & Kochanski, A. (2010). Knowledge discovery and analysis in manufacturing. Quality Engineering, 22(3), 169–181. https://doi.org/10.1080/08982111003742855

[24]. Ghotra, B., McIntosh, S., & Hassan, A. E. (2015). Revisiting the impact of classification techniques on the performance of defect prediction models. Proceedings - International Conference on Software Engineering, 1, 789–800. https://doi.org/10.1109/ICSE.2015.91

[25]. Mech Journal - 1.pdf. (n.d.).

[26]. Fulfillment, I. N., Internship, O. F., For, P., & Devel opment, C. (2014). RAIL WHEEL FACTORY (MINI STRY OF RAILWAYS), YELAHANKA , BANGALORE. August, 1–75.

[27]. Queen, P. (2016). T owards G eneralizing D efect P redic tion M odels. January.

[28]. Rajbahadur, G. K., Wang, S., Ansaldi, G., Kamei, Y., & Hassan, A. E. (2021). The impact of feature importance methods on the interpretation of defect classifiers. IEEE Transactions on Software Engineering, January. https:// doi .org/ 10. 11 09/ TSE.2021.3056941

[29]. Zheng, J., Wang, Q., Zhao, P., & Wu, C. (2009). Optimization of high-pressure die-casting process parameters using artificial neural network. International Journal of Advanced Manufacturing Technology, 44(7–8), 667–674. https://doi.org/10.1007/s00170-008-1886-6

[30]. Mane, V. V, Sata, A., &Khire, M. Y. (2010). New Approach to Casting Defects Classification and Analysis Supported by Simulation. 59th Indian Foundry Congress, October 2014, 87–104.

[31]. Hiremath, S. S., &Dulange, S. R. (2015). Advanced Techniques In Casting Defects And Rejection Analysis : A Study In An Industry. International Journal of Innovations in Engineering Research and Technology (IJIERT), 2(9), 1–9.

[32]. Junyent-ferre, A. (2015). Summer Internship Project Report. January, 0–10