THE IDENTIFICATION OF RESULTS OF INSPECTION IN LOGISTICS USING DECISION TREE CLASSIFIER AND DTREEVIZ

Dr. M. Rajeswari,
Assistant Professor, Department of B.Com (Business Analytics and Accounting & Finance),
PSGR Krishnammal College for Women, Coimbatore, India.
rajeshwarim@psgrkcw.ac.in

V. Divya dharsini,
UG Scholar, Department of B.Com (Business Analytics),
PSGR Krishnammal College for Women, Coimbatore, India.
divyadharsiniv21@gmail.com

ABSTRACT:

Due to inadvertent mistakes in the production process, the manufacturer processes the step called inspection for checking and assuring the quality of the products. The inspection helps to rectify the mistakes or damage in the products and also increases customer satisfaction by delivering good quality products to the customers. The damage rate of the product and inspected quantity are used to classify in the decision tree for decision making. The main aim of the study is to identify the inspection results that "how many products were passed; how many went to rework and how many products were failed in the Inspection Process" using Decision tree classifier and dtreeviz in Jupyter Notebook. The inspected product with pass result products will move to the next process called loading.

Keywords: inadvertent, customer satisfaction, inspection, Decision tree classifier, dtreeviz

I. INTRODUCTION

Quality is the main thing that the customer prefers, so by inspecting the product quality in the inspection process, the customer will get satisfied with the product quality and will also prefer buying the product again. Inspection plays a massive role in every manufacturing company where the quality of the product will be checked and examined by the inspectors and proceed the inspected products to the next process. The next process is nothing but the loading process where the passed quantity of inspected results will move to the loading with effective load planning. The decision tree using damage rate and inspected quantity will be implemented for the identification of the inspection results. If the Damage rate of the product is zero, then the product will come under the passed products that mean the product doesn't have any damage. Thus, the product seems to be a good product to be loaded for the next process.

II. OBJECTIVE

The main objective of this study is to find out the inspection result using the Decision tree classifier in Jupyter notebook (Python) and shows the graphical visualization of the decision tree with the help of a library called dtreeviz for the easy understandability of the users. After analyzing the inspection results, finding out how many pass products are moved to the loading Process.

III. REVIEW OF LITERATURE

Logistics is an integral part of our lifestyle in the modern world. It is the whole operations of management and of the organization of the physical flows and of the knowledge inside the corporate, also as between the corporate and its partners. The main aim is to coordinate the sourcing of production and distribution activities. Transport is an essential component of it. Still, it also includes the demand planification, orders' processing, procurement, planification, production, relations with customers and suppliers, storage, handling, assembling, packaging, products packaging, and support functions associated with these activities [5].

Product quality is a crucial factor that affects consumers' purchase intention. Nowadays, more and more enterprises are outsourcing product design and manufacturing activities to manufacturers, and therefore the price war among enterprises is increasingly fierce. The huge competitive pressure requires manufacturers to further reduce cost, which aggravates the motivation for manufacturers to conduct production with adulterated quality. The very most used measure of quality management and control in enterprises is quality inspection. But quality inspection has two major drawbacks: (a) they're costly, and (b) they're susceptible to cause quality inspection evasion motivation [3]. A multi-level logistics supply network optimization model with constraints of distribution capacity, inventory capacity and customer's best delivery time is made whose optimization objective is that the sum of service costs, inventory costs, and transportation costs of the whole logistics supply chain network [4].

Introducing inspection in a production process, is expected to be a profitable course of action since at some point the associated costs will be recovered from the benefits realized via the detection of defective items and isolation of defect-causing variation sources. If one only inspects, reworks, or scraps a finished product, the inspection cost is low. However, scrapping a finished product is usually expensive; its influence on the final productquality is direct but its contribution to the variation-source diagnosis is often limited. If, however, one adopts a strategy of inspecting, reworking, or scrapping upstream intermediate products then the inspection costs could increase considerably but the scrapping of an unfinished product may be relatively inexpensive. Also, whereas its influence on the overall final product quality is indirect, its ability to facilitate variation-source diagnosis is often good. Any sound inspection strategy will have to consider the needs of multiple stakeholders and make reasonable trade-offs between their objectives. Therefore, when and where in a production process an inspection should be performed or sensing devices be distributed is an important and challenging decision in quality control. The various cost and constraint factors as well as operational alternatives interact in an intricate fashion and make the solution far from trivial [8].

In the data processing community, decision tree algorithms are very fashionable since they're relatively fast to coach and In data processing, decision tree algorithms are very fashionable thanks to their characteristics like fast to coach and produce transparent models. The machine learning community has produced an outsized number of programs to make decision trees for classification. These classifiers provide support for several health care areas in deciding. Out of those CART has been proved to the simplest classifier for medical data [2]. Classification is that the process of generating an outline or a model for every class of a given dataset. The dataset consists of a variety of records, each consisting of sew eral fields called attributes which will either be categorical or numeric. Each record belongs to at least one of the given (categorical) classes. The set of records available for developing classification methods are generally decomposed into two disjoint subsets, training set and test set. The former is employed for deriving the classifier, while the latter is employed to live the accuracy of the classifier. Decision-tree classifiers SLIQ and SPRINT are shown to realize good accuracy, compactness and efficiency for very large datasets [1].

Classification trees offer an efficient implementation of such hierarchical classifiers. Indeed, classification trees became increasingly important thanks to their conceptual simplicity and computational efficiency. A decision tree classifier features an easy form that can be compactly stored which efficiently classifies new data. Decision tree classifiers also can perform automatic feature selection and complexity reduction, and their tree structure provides easily understandable and interpretable information regarding the predictive or generalization ability of the classification. To construct a classification tree by heuristic approach, it's assumed that a knowledge set consisting of feature vectors and their corresponding class labels is available. The features are identified supported problem-specific knowledge. The decision tree is then constructed by recursively partitioning a knowledge set into purer, more homogenous subsets on the idea of a group of tests applied to at least one or more attribute values at each branch or node within the tree. This procedure involves three steps: splitting nodes, determining which nodes are terminal nodes, and assigning a class label to terminal nodes. The assignment of sophistication labels to terminal nodes is straightforward: labels are assigned supported a majority vote or a weighted vote when it's assumed that certain classes are more likely than others [6].

The decision tree classifier by Quinlan is one among the foremost well-known machine learning techniques. A decision tree is formed of decision nodes and leaf nodes. Each decision node corresponds to a test X over one attribute of the input data and features a variety of branches, each of which handles an outcome of test X. Each leaf node represents a category that's the results of a decision for a case. The process of constructing a decision tree is basically a dividend-conquer process [7].

IV. METHODOLOGY

4.1 THE INITIAL STEP IN INSPECTION

The inspected results are of three levels in inspection process:

1) PASS
2) REWORK
3) FAIL

The product with the results of the pass will move on to a subsequent process called Loading, the product of rework will move to the repair section and therefore the product of the fail is going to be the failed product.

4.2 SUPERVISED LEARNING

Supervised learning, also referred to as supervised machine learning, may be a subcategory of machine learning and AI. It is defined by its use of labeled datasets to coach algorithms that classify data or predict outcomes accurately. The input data is fed, which adjusts its weights through a reinforcement learning process, then the model has been fitted appropriately. Supervised learning helps organizations solve a spread of real-world problems at scale, like classifying spam during a separate folder from your inbox. Supervised learning is divided into two parts: classification and regression:

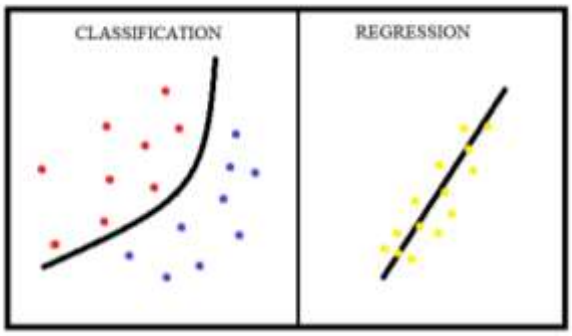

Fig 4.1 classification and regression

a) Classification are used to accurately assign test data into specific categories. It shows specific entities within the dataset and draws some conclusions on how those entities should be labelled or defined. The most popular classification algorithms are linear classifiers, support vector machines (SVM), decision trees, k-nearest neighbor, and random forest.

b) Regression is used to understand the relationship between dependent and independent variables. It is commonly used to make projections in the business fields. the most popular regression algorithm are Linear regression, logistical regression, and polynomial regression.

4.3 DECISION TREE CLASSIFIER

Decision tree is used to visually represent the decisions with tree-like structure for easy understanding. It also helps to solve the classification problems. They are non-parametric supervised learning methods used for classification and regression. It learns from data to approximate a sinusoid with a group of if-then-else decision rules. The bigger the tree, the bigger the decision rules. It builds classification or regression models within the sort of a tree structure. It breaks down a knowledge set into smaller and smaller subsets while at an equivalent time an associated decision tree is incrementally developed. A decision node has two or more branches. Leaf node represents a classification or decision. The topmost decision node corresponds to the simplest predictor called the root node. Decision trees can handle numerical data.

In the inspection process, a decision tree classifier is used to identify the inspection results by analyzing the damage rate. And then it makes decisions with the damage rate like, If the damage rate is less than 0.5, then the product will be Failed, it shows that there are 8 samples of product that are failed and if the condition was false, then it goes to another node where 9 products were passed and 6 products went to Rework

4.3.1 WORK FLOW AND IMPLEMENTATION OF DECISION TREE

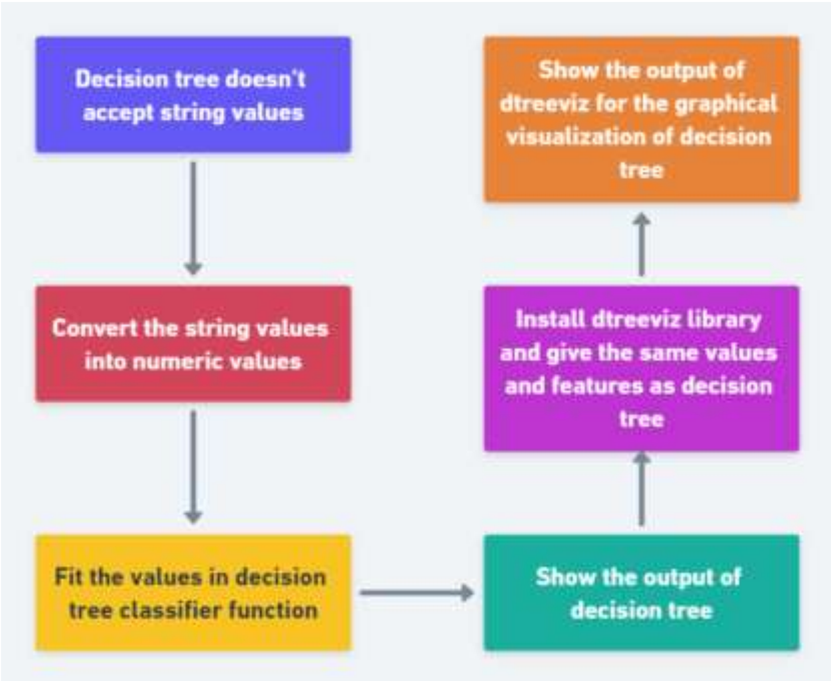

Fig 4.2: Work flow

The Decision Tree doesn't accept the string values, so the string values will be converted to numeric values for fitting the values in decision tree classifier function and then get the features for the decision tree and finally plot and show the output of decision tree. Using dtreeviz library, the decision tree will graphically be visualized for the easy understandability of the tree

4.3.3: GINI METHOD IN DECISION TREE

The Gini Index or Gini Impurity is one of the attribute measures in the decision tree. It helps the larger partitions and very easy to implement. In simple terms, it calculates the probability of a particular randomly selected feature that was classified incorrectly. It varies between 0 and 1, where 0 represents the purity of the classification and 1 denotes random distribution of elements among various classes. A Gini Index of 0.5 shows that there's equal distribution of elements across some classes.

The Gini method uses this formula:

Gini = 1 - (x/n)2 - (y/n)2

Where x is the number of positive answers, n is the number of samples, and y is the number of negative answers, which gives us this calculation.

4.4 DTREEVIZ

Dtreeviz is the graphical visualization of the decision tree. It is used for better understanding of decision tree and with the graphical view, it will be more attractive to the users to know about the decision tree accurately. It is one of the nice libraries, which brings much more tables and create graphical visualizations that are not only prettier but also very informative about the decision process. Another handy feature of dtreeviz which improves the model's interpretability is path highlighting of a specific observation on the plot. This way clearly see the features contributed to the class prediction.

V. RESULT AND DISCUSSION

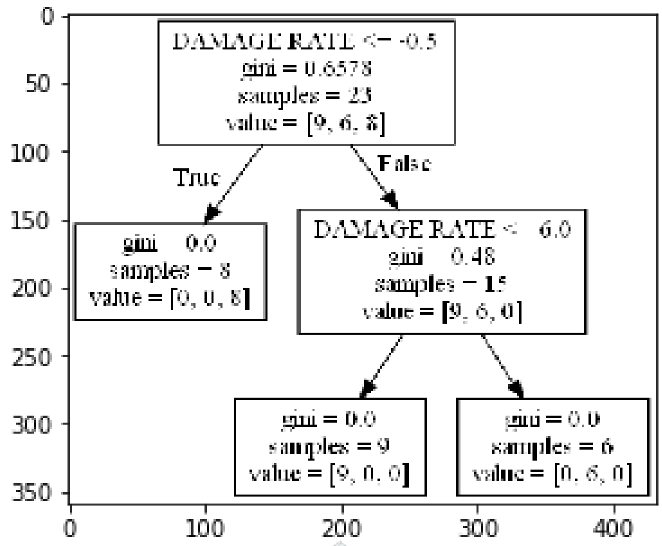

Fig 5.1 Decision Tree of Inspected Results on Damage rate

EXPLAINATION:

DAMAGE RATE< 0.5 means that every quantity with a rank of 0.5 or lower will follow the True arrow (to the left), and the rest will follow the False arrow (to the right).
gini = 0.6578 refers to the quality of the split, and is always a number between 0.0 and 0.5, where 0.0 would mean all of the samples got the same result, and 0.5 would mean that the split is completed exactly within the middle.
samples = 23 means that there are 23 quantities left at this point in the decision, which is all of them since this is the first step.
value = [9, 6, 7] means that of these 23 quantities, 9 will "PASS", 6 will go for "REWORK", and 8 will be "FAIL".

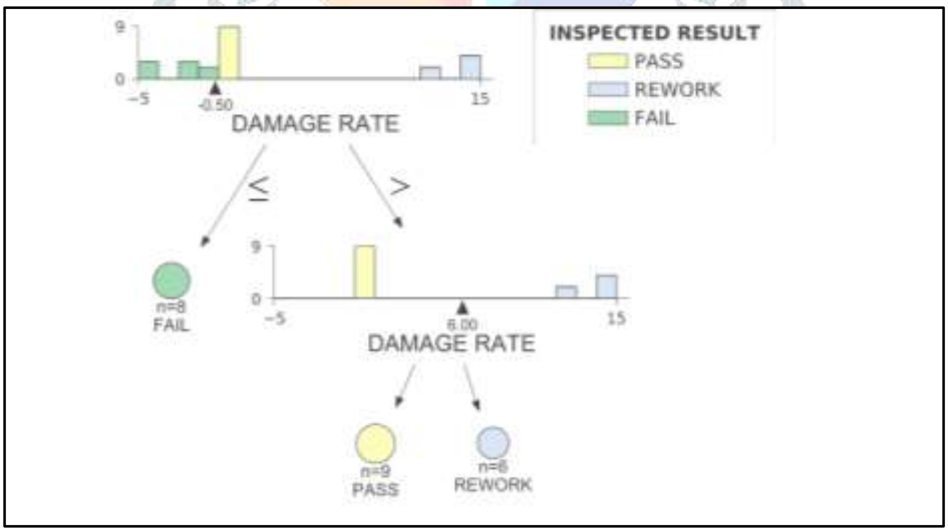

Fig 5.2: Decision Tree Visualization (dtreeviz) of Inspected Results on Damage rate.

The above Fig 5.2 represents the graphical visualization of fig 5.1. Decision Tree using dtreeviz. It shows "9 products were passed; 6 products went to rework and 8 products were failed in the Inspection Process". Let's see the plot step by step. At each node, stacked histogram of the feature that's used for splitting the observations, colored by class has been shown. By this way, the classes are segregated by each split will be clearly shown. The small triangle at the x-axis is the splitting point. In the first histogram, all observations of the Inspected results class have damage rate smaller than 0.5 has been shown.

The right branches of the tree indicate selecting the values larger or adequate to the splitting value, while the left one lesser than the splitting value. The leaf nodes are represented as circles, which tells the fraction of the observations within the leaf that belongs to which class. This way, we easily see which class is that the majority one, so also the model's prediction. One thing we don't see at this plot is that the value of the Gini coefficient at each node. In my opinion, the histogram provides more intuition about the split and therefore the value of the coefficient won't be that relevant just in case of a presentation for the stakeholders either way. Thus, 9 products has pass results which will move to the loading process, then transports to various business units and finally to the customers.

V. CONCLUSION AND FURTHER WORK

In this paper, Jupyter notebook is used to identify the Results of inspected products in inspection process using the Decision tree algorithm. By dtreeviz library, Decision tree easily visualizes in graphical format for the better understanding. The products which have successful results that is pass results in inspection will moves to the next process. The main intention of this paper is to identify the products which goes to the loading process in the logistics.

VI. REFERENCE

1) Alsabti, Khaled; Ranka, Sanjay; Singh, Vineet, CLOUDS: A Decision Tree Classifier for Large Datasets. In Proceedings of the Fourth Knowledge Discovery and Data Mining Conference, eds. Rakesh Agrawal and Paul Stolorz, 2-8. Menlo Park, Published by The AAAI Press, Menlo Park, California. Calif.: AAAI Press.

2) D.Lavanya, Dr.K.Usha Rani, “Performance Evaluation of Decision Tree Classifiers on Medical Datasets”. International Journal of Computer Applications 26(4):1-4, July 2011

3) Hu, H.a,* , Wu, Q.a , Zhang, Z.a , Han, S.a, Effect of the manufacturer quality inspection policy on the supply chain decision-making and profits, Advances in Production Engineering & Management, Volume 14, Number 4,December 2019.

4) LIU Yan-Qiu, Wang Hao – Optimization for Logistics Network Based on the Demand Analysis of customer, Chinese Control and Decisions, 2016.

5) Mamdouh Tlaty & Mohamed Moutmihi, “From the logistics function to the logistics service: A Literature review.” Global Journal of Management and Business Research: A Administration and Management Volume 15 Issue 6 Version 1.0 Year 2015 Type: Double Blind Peer Reviewed International Research Journal Publisher: Global Journals Inc. (USA).

6) Mahesh Pal, PM Mather - 22nd Asian conference on remote Sensing,5 -9 November 2001, Singapore.

7) Quinlan, J. R. (1993). C4.5, Programs for Machine Learning. Morgan Kaufmann San Mateo Ca, 1993

8) Sampatraj S. Mandroli, Abhishek K. Shrivastava And Yu Ding, A survey of inspection strategy and sensor distribution studies in discrete-part manufacturing processes, IIE Transactions (2006)