Decision Support System for Load Cell Selection

W.M.R.Thamel

179477D

Faculty of Information Technology

University of Moratuwa

2020

Decision Support System for Load Cell Selection

Thamel WMR
179477D

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the fulfillment of the requirements of Degree of Master of Science in Information Technology.

April 2020

Declaration

We declare that this is our own work and has not been submitted to another degree or diploma at a university or other higher education institution. Information obtained from published or unpublished work by third parties is acknowledged in the text and a list of references is provided.

Name of Student	Signature of Student
W.M.R.Thamel	………………………
	Date: ………………..

Supervised by

Name of Supervisor	Signature of Supervisor
Chaman Wijesiriwardana	……………………….
	Date: ………………...

Acknowledgements

First of all, I would like to thank Mr.Chaman Wijesiriwardana, a lecturer at the Faculty of Information Technology at Moratuwa University, for his valuable time in teaching, monitoring, consulting and research projects.

In addition, I would like to thank Dr. Mohamed Fildos for teaching me research methodology and literature review, as well as my thesis in which the subject was written on which this study was based.

Thanks also to all the colleagues in IT degree program and my family who gave us valuable comments in order to improve the results of our research throughout my life, especially the support that I provided. I must acknowledge the support and editorial assistance from my wife, I would not finish this thesis.

ABSTRACT

Loadcell is a transducer that converts force into an electrical signal, used in diverse industrial applications. Selecting a proper Loadcell is a key success point of a weighing application. The Loadcell selection process requires expert knowledge. Usually, the customer support team of the manufacturer shares expertise knowledge with the customer. Automation of the customer support process can be done using data mining techniques. Previous customer support records can be utilized for this process and the Naïve Bayes theory guides this automation process. The predictions can be used to determine important relationships which help to make critical decisions.

Contents

List of figures

List of Tables

CHAPTER 1

1.INTRODUCTION

1.1 Prolegomena

The weighing applications are taking an important part in the modern world. The need for accurate reading of the weigh is essential for many fields e.g. the Food sector, medical sector, agricultural sector, medical sector. Because of the government's legal regulations these accuracy measurements take a big role and that trend brings benefits to the customer and the producer.

With the high demand for the production high effective and well efficient weighing applications are required. Because of this requirement various types of weighing applications and loadcells are needed.

The selection of a correct loadcell is the key point of a weighing application and the success of the weighing application is highly depending on the loadcell. To select a correct load cell, it needs expert knowledge and without having that expert knowledge it will be a time-consuming task. Most of the loadcell vendors unable to provide that expert knowledge because of the high demand of customers.

In this research, the hypothesis is based on the above problem and the main objective is to automate the loadcell prediction process for given customer requirements. The automation process is based on the classification concept. Naïve Bayes the theory is well support for machine learning and here we used the Naïve Bayes to fulfill the basic objectives.

1.2 Background

Load cell

The load cell is a transducer and that can convert load into an electrical signal. Also, there are various types of loadcells and out of this, strain gauges load cells are the common type of loadcells. Dynamic loading and static loading methods are capable to handle through this strain gage load cells. The strain gage is based on the Wheatstone bridge and resistance change is measure after giving a certain voltage.

The captured signal from the load cells is in a few millivolts and before it uses the signals are sent to the amplifier board to amplify the existing signal. Finally, amplified signals are sent to the analog to digital converter and the signals are converted to the digital form.

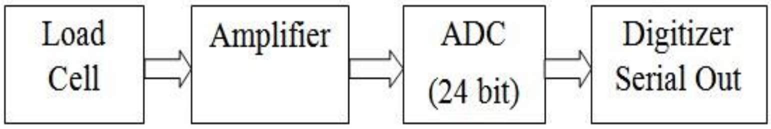


Figure 1**:** Interfacing of the load cell with an amplifier

Data mining

Because of the rapid growth of technology, data become a more valuable and useful asset. The amount of the data is increasing day by day and many partied capture data for their future usage. Most of the data stored in electronic media and the access of the data are very easy. Data mining techniques are used to convert data into understandable formats. AS a result of the data mining process the recognizable patterns can be found from large amounts of data. Also, it provides an analyzer to conduct accurate and responsive analysis. The pattern or the results which are found in the data mining process provides many benefits such as economic benefit, new finding, Solution to the existing problem, etc.

1.3 Problem

Selecting a suitable load cell for the application is a very complex process and it directly affects the success of the weighing application. There are many types of load cells in the load cell market and various attributes are consists of them. Matching the load cell attributes with the required attribute in the application is difficult for a person who does not have expert knowledge of the load cells. Because of that most of the customer request help from the customer support team or sales team. With the growth of the customers, both the customer support team and the sales team difficult to manage their workload.

1.4 Aim and Objectives

Aim

Provide an automated solution to propose suitable load cells for the customer's requirement in a short time and give a loadcell comparison option.
Find important relationships between loadcell selection parameters.

Objectives

1. Predict and propose the proper load cell family to the customer according to his Requirement.
2. Minimize the working load of the sales team.
3. Minimize the load cell selection time of the customer.
4. Find important data patterns which are important to customer support and sales team.
5.provide a loadcell comparison option.

1.5 Proposed solution

In this research, we propose a web-based tool to predict the correct load cell for the customer requirement. The customer only has to select the application type and its capacity, then the system will propose multiple choices and the customer can choose the best load cell among them. The system is provided the comparing facility and customers can compare three loadcells once.

1.6 Main parameters in load cell family selection

Application type

There are many industries use force and weight measurements without having any limitations for their requirements and, there are different applications across various industries. For the easiness, applications are categorized into some application types.
Agricultural Vehicles, Bagging machinery, Bench scales, Bottle filling machinery, Bulk tanker vehicles, Calibration of assembly machinery, Coil weighing, Crane scales, End-of-line test equipment, Gas container scales, Grading machines, High capacity scales, High speed checkweighers, Hoppers weighing, IBC bagging machinery, Joining applications, Ladle weighing, Lifting systems, Low profile scales, Machine monitoring and control, Marine scales, Medical equipment and process control, Multi-head weighers, On-board vehicle weighing, Packaging machines, Pallet truck scales, Patient

weighing systems, Platform Scales, Press-fit applications, Railroad scales, Retail scales, Riveting applications, Small platform scales, Suspended tanks and hoppers, Tank and silo weighing, Trailers and feed mixers, Truck and Train weighing, Waste collection vehicles, Weighbridges

Loadcell type

Loadcell types are categories which are classified Loadcells based on measured load, the working principle and the construction of the load cell. Beam Load Cells, Compression Load Cells, Miniature Force Sensors, Planar Beam Load Cell, Single Point Load Cells, Small platform scales, Tension Load Cells

Capacity

0.5t, 1000kN, 1000lb, 100kg, 100kN, 100lb, 100t, 10kg, 10kN, 10lb, 10N, 10t, 150kg, 150kN, 150t, 15kg, 15kN, 15lb, 15t, 18kg, 1kg, 1t, 2.5t, 2000kN, 2000lb, 200kg, 200kN, 200N, 200t, 20kg, 20kN, 20lb, 20t, 22.5t, 22kg, 250kg, 250lb, 25kg, 25lb, 25t, 2kg, 2kN, 2lb, 2t, 3000lb, 300g, 300kg, 300t, 30kg, 30lb, 30t, 375kg, 37kg, 3kg, 3t, 400kg, 400t, 40kg, 40t, 45t, 5000kN, 5000lb, 500g, 500kg, 500kN, 500lb, 500N, 50kg, 50kN, 50lb, 50t, 5kg, 5kN, 5lb, 5t, 6kg, 7.5t, 750kg, 75kg, 75t, 7kg, 800lb, 8kg

Material

Aluminum, Stainless steel, Tool steel

Environment condition

According to the application, load cells should be work in an environment which are differ to each other. As an example, some load cells are placed in a clean environment such as research laboratory and some are place in open, dusty environment such as garbage trucks. Because of this reason's protection of the loadcell is important and it is effect to the performance, durability, accuracy.

Most of the load cells will be in pots. This means that the cavities are filled with an epoxy or resin material that completely covers the electronic components. This helps protect circuits from external damage and moisture and helps dissipate heat.

Some load cells have hermetic seals for the most complete environmental protection. A fully welded joint provides a robust, watertight enclosure, resistant to the harshest environments.

Environmentally sealed by potting, Hermetically sealed, IP40, IP60, IP64, IP65, IP66, IP67, IP68, IP69, Very rugged construction

Certification

The load cells used in commercial and legal applications must be legally certified by the authorized body as "legal for trade" to ensure that they comply with the required standards. This can apply to both load cells and the corresponding electronics used in weighing systems. Certification guarantees that they operate with the precision that allows them to be used in specific applications.

These certifications guarantee that the load cell and the electronic components comply with a set of common and agreed requirements. To obtain approval, the load cells and other components undergo a complex series of tests and checks.

The Organization for International Metrological Law (OIML) is the common body which approves load cells and is recognized in most countries of the world. National Type Assessment Program (NTEP).

ATEX, FM, OIML, NTEP

Surface

Bead-blasted, Electroless nickel plated, Electro-polished, Painted

Flatform size

1000mm X 1000mm, 1400mm X 1400mm, 250mm X 250mm, 350mm X 350mm, 400mm X 400mm, 600mm X 600mm

1.7 Load cell Families

Load cell family is created for the easiness of identification and which is a calling name for a loadcell.

AP5, AP8, BK2, DSB7, ISA S-Beam Force Sensor, JF1 Press Force Sensor, MBA Button Force Sensor, MBA-TW Button Force Sensor, MBC, MBD2, MHT1 Compression Force Sensor, MHT1 Compression Force Transducer, MHT2 Compression Force Sensor, MK Button Force Sensor, PA1, PA2, PA3, PB, PBW, PC1, PC12, PC2, PC22, PC2H, PC30, PC3H, PC42, PC46, PC52, PC5H, PC6, PC60, PC6D, PC6H, PC7 PC7H, PC81, PCB, Q1 Smart Low Profile Force Sensor, Q50, RC1, RC3, RC3D, SB14, SB2, SB4, SB5, SB6, SB61, SB8, SB8L, SB9, SBT, SLB, UB1, UB6, ULB, UXT, ZLB, ZLS

1.8 Structure of the Thesis

Chapter 1 describe about the introduction for the “Decision support system for the load cell selection” and chapter 2 present the literature review for the following topic. Chapter 3 explain about the technology used in the research. The Approach used to solve the research question is explained in chapter 4. Chapter 5 explained about the analysis and designing phase and chapter 6 described about the implementation of the research. In finally we present the discussion.

CHAPTER 2

2.REVIEW OF LITERATURE

2.1 Introduction

This chapter focus to review the existing researches in the same subject domain. Basically, the research problem relates to load cells and data mining. According to the above fields, the number of researches reviewed and found the concepts and theories to conduct this research.

2.2 Related works in same topic

Related to the above problem there is no research found which used data mining techniques. But there is some research which use the data mining technique to solve the similar type of problems, but the subject domain is different.

2.3 Datamining

Classification

S. Agrawal [7] expressed that Classification is a supervised machine learning technique which is a dominant technique of all the machine learning techniques. How this operates is that it uses examples to train the machine by using data sets as examples. Finally, he mentioned this can be used to find, segregate categories, to realize free define classes.

Naïve Bayes

Naïve Bayes algorithm is a simple classification algorithm in machine learning, and it is well supported with textual data. This algorithm is faster and only needs the minimum space, a small number of training data set for the evaluation. Naïve Bayes algorithm can consider the data point with n number of features, features can be any text data. The main condition is features should be conditionally independent of each other. Because of this independency each attribute can do the learning separately either data is large.

In this context of text classification, the probability that a text data Y belongs to a Class X is calculated by using Bayes Theorem as follows:

P(X | Y) = P(Y | X) * P(X)/P(Y)

Here, P(X |Y) denotes the Posterior; P(Y | X) the likelihood; P(X) the prior and P(Y) is the Evidence. The condition posed here when the probability is considered is that the probability of P(Y) should never be zero.[9]

F. Harahap [3] this study proposed a Naïve Bayes assumption for Predicting Purchases. The researcher highlights the benifits of the Naïve Bayes and he clearly mentioned it is capable to do the classification with small data set. The implementation of the research was tested against the WEKA tool. This algorithm allows each independent attribute to the final decision.

H. Somwanshi [4] predicts real-time Dengue conditions using Naïve Bayes Predictor in the IoT. In their study, Naïve Bayes has been discussed as the technique for classification of collected and predefined data, which meets literature results. It also calculates the posterior probability for predefined data which gives the maximum probability of given dataset in a prediction process. The result depicts that Naïve Bayes is the best algorithm that predicts the disease and the accuracy rate of the algorithm higher than other algorithms.

R. Reeta [6] this research suggests the early diagnosis of autism by Naïve Bayesian classification approach. In this research proves Naïve Bayer's approach as an efficient classifier. Naïve Bayes approach is easy to build its structure has its priority. The features are not strongly correlated in Naive Bayes classifier which makes its classification process efficient. Thus, the Naïve Bayes approach was successfully used as an efficient method for early autism prediction.

Decision Tree

The decision tree algorithm is a classification algorithm to build a tree structure with leaves. The leaves are representing the class labels and branches are indicate the functionalities. The classification tree is the result of the algorithm and each division indicates the If-then rule and each leaf represents the value of the relevant variable. For the predicting purposes the sample data can be used and the selection of the best variable is done by comparing the variables with the sample variables. This process is done in each step and the success of the selection is dependent on the success of the sample. This algorithm is a simple model, and this is not supported for the complex scenarios.

J-48

There are many algorithms used in tree learning methodology and among them C4.5 is the popular algorithem.C4.5 algorithm is built by extending the ID3 algorithm concepts and the main improvement is continuous, discrete attribute handling based on a threshold-basis split. In calculating the phase of the gain and the entropy, the missing values are not considered and remove useless branches by swapping them by the leaf node. The C4.5 algorithm named as J48 algorithm in the WEKA tool.

Algorithm J-48

1: Check for any base cases
2: for each attribute a do
3: Find the normalized information gain from splitting on a
4: end for
5: Let a_{best} be the attribute with the highest normalized information gain
6: Create a decision node that splits on a_{best}
7: Recur on the sub lists obtained by splitting on a_{best}, and add those nodes as children of node.[8]

K.R Prageep [5] this study main aims to predict diabetes using blood glucose levels. The research work is done by using the J48 Decision tree algorithm. As a result of this work, he implemented the online web application and remote patient monitoring system. By using the J48 algorithm he implemented a very accurate and less costly system.

Decision Table

Decision table is a set of well-organized If-then conditions and, which is a good for numeric prediction. Because of the compactness, it is more understandable and because of the simplicity it required less computation power than decision tree.
The WEKA tool provides a decision table algorithm under the rules and the output of a data set provide simply with similar format to the input. The main purpose of this algorithm is to summarize the data with same amount of the attribute like the original data set. The proper category of the new item is selecting by comparing the data value with the non-class data items in the decision table. The decision tables use best-first search for the searching and use cross-validation for evaluation.

D. Catherine Rexy[2]predict women survival rate during pregnancy using decision table classification approach. In his research consider fourteen parameters. They found the impact of every single parameter on the survival status of the women and how much percentage is the survival rate obtained. He used WEKA tool in his research work. Finally, they found main three factors that directly impact to the survival rate of a child.

2.4 Load cell

A. Qandil [1] In his research classified Loadcells based on measured load, working principle and the construction of the load cell. Based on this classification there are some different loadcell types such as Compression, Tension, shear-based type, single-ended, double-ended shear beam, load button, S shape, pancake. This research clearly expresses the relationship between the loadcell type and the application type. It also explains the key importance to protect the gauge and its wiring system from the environmental factors. There are many circumstances that damage loadcells such as mechanical abuse, moisture, oil, dust, dirt and so on. The research expressed that the performance of the loadcell depends on the capacity of the loadcell and the building material of a load cell.

Wilmar Hernandez [11] also explain that there are a large number of load cell types and mentioned the common loadcell types which is using in industrial applications such as bending beam, shear beam round, shear beam rectangular, miniature, low profile, ‘s’ or ‘z’ beam, canister, compression/tension, compression, and platform and single point load cells.

2.5 Summary

After critically reviewing the similar type of previous researches we found that the best data mining rule for this research is classification rule. Also, the WEKA tool is a good data mining tool for testing and many people used this tool for their research works. By analyzing the previous research about the load cells, we found some important attributes for this research. The next chapter will discuss the technologies adapted for solving our problem.

CHAPTER 3

3.TECHNOLOGY ADAPTED

3.1 Introduction

Chapter 2 discussed the previous findings related to this research and found the important theories, concepts, and tools that can use in this research. this chapter will clearly present the technologies used to solve this problem.

3.2 The .NET Framework 4.5

The .net framework is a platform which is develop for software development by Microsoft. Using this framework developers can develop web and the windows applications. Most of the programming languages and the libraries capable to run on this framework smoothly. Because of that developers are free to develop their applications using many languages such as VB, C#, C++ etc.

3.3 Microsoft SQL Server 2014

SQL Server is developed to manage relational databases and it is a product of Microsoft. The SQL server allows to handle multiple databases and provide facility to create, managing features such as modify, update, delete, backup, etc. This runs on windows over twenty years and since 2016, it works on Linux as well. The several programming structures are included in the new SQL versions for the easy development.

3.4 Weka tool

Weka is a data mining tool and it consists pool of algorithm which can use in machine learning tasks.it is capable to apply directly to the given data set. The WEKA tool consists many areas such as classification rule, data preprocessing, regression analysis, clustering analysis, data visualization, and assignment rules. In this research, The Weka is used to test the result by comparing the system with the Weka result.

3.5 MVC Architecture

MVC (Model View Controller) is a popular software designing pattern and it is used to develop web applications. This architecture divides web application into three main components (model, view, controller). All this component is used to manage the application development. This architecture is suitable for large scalable projects.

Model

This is the component which is responsible for manage data related logics. By transferring data into the models this can pass to the view and controller. ex: take a client object from database, modify and update.

View

The main usage of the view component is to provide interface for all logics in the application. As an example, text boxes, check boxes, dropdown lists are used in web application.

Controller

The controller unit acts as an intermediate component between the model and the view, manages all the requied logics and incoming requirements, uses the model components to operate the data and related with the views to create the counter client all interactions and inputs from the customer view and met to the database with the customer model. Also, controller helps to display customer data.

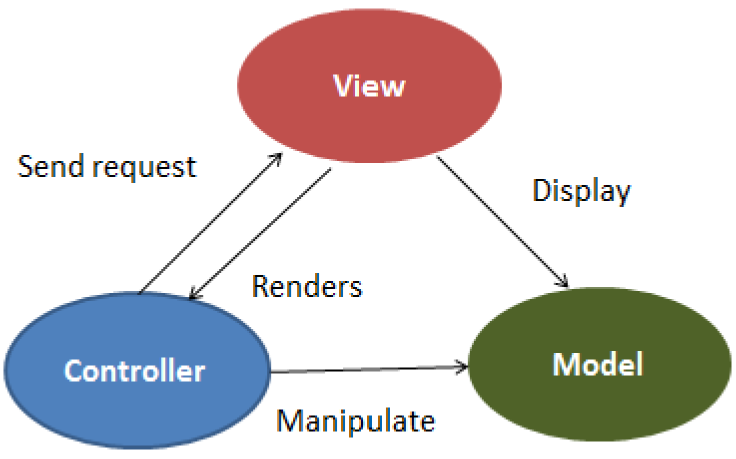


Figure 2: MVC architecture

3.6 Summary

In this chapter, we described the technologies which are used in research. Under this chapter, we present the .NET Framework, SQL Server, WEKA tool, MVC Architecture and in next chapter will explain the approach deeply which use to solve this research question.

CHAPTER 4

4. My Approach

4.1 Introduction

This chapter presented the technology to be used to solve the research problem and described our approach to address the problem of load cell prediction of different customer requirements about the weighing applications. Load cell prediction is the main module of this research and the second module is to find the important relationship between the collected data. We present our approach by highlighting hypothesis, input, output, process, users, and features of the load cell prediction solution.

4.2 Hypothesis

The main hypothesis is to build the automated loadcell prediction tool for the loadcell manufactures and for the sellers to serve their customers who want to buy the loadcell for their weighing requirement. Apart from the loadcell prediction, this developed tool provides a facility to find important relationships among the properties.

4.3 Data Modeling

According to the review of literature the classification technique is good for prediction and it's well supported for machine learning. There are many approaches in classification and in this research mainly we considered three approaches. After doing the review of the literature we found Naïve Bayes, R48, Decision Tree are suitable for this research domain. In data modeling, we found that naïve Bayes is the best approach among these three approaches. Also, naïve Bayes is simple and well understand for the development phase.

4.4 Input

Initially, this record set starts with 9500 records and it covers 2018-01-01 to 2019-01-01 time period. Row data was in Excel, e-mail, note format and data captured form the sales team, customer support officers, etc...

4.5 Attribute selection

A. Qandil [7] and Wilmar Hernandez [11] clearly expresses the importance of the following attributes to the weighing application.

- Loadcell type
- Application type.
- Capacity.
- Environmental conditions.
- Material

4.6 Data Pre-Processing Data

Some collected data is not irrelevant and redundant information as well as noisy and unreliable data. Therefore, data cleaned steps such as filling missing values, smoothing noisy data and removing inconsistencies.

4.7 Input through GUI

User needs to select the application type and the required capacity according to his requirements.

4.8 Output

There are two outputs of this proposed solution and the main one is load cell prediction for the given user requirement. The second output is mainly focused on the sales team and to the top management. This module helps users to analyses past data to build a relationship among the selected attributes and make strategic decisions. Both outputs are implemented as a web solution and that solution is capable to join existing loadcell manufacturer’s web sites.

4.9 Process

At the initial stage, the system will use the preprocessed data and the tool generates the proposed outputs. The outputs are present to the user as a percentage and the data mining concept is based on Naïve Bayes theory. After implementation on a live environment, this tool will generate the outputs using real-time data.

4.10 Users

Mainly there are two types of users, internal users, and external users. Internal users are sales team, customer support team, decision-makers and external users are customers.

4.11 Features

Except for the functional requirement, there are some non-functional requirements and the basic one is this developed tool should capably merge existing web site. And the other requirements are followed.

- Low cost
- User independent
- Browser independent
- Easy to use
- Expandable
- High performance

4.12 Summary

This chapter expressed how data mining use in the machine learning approach to solve practical problems. Here this chapter discussed the research hypothesis, input, output, process, users, and features. The next chapter will show the analysis and design.

CHAPTER 5

5.ANALYSIS AND DESIGN

5.1 Introduction

The Last chapter we discussed the best approach which can use to solve the research problem. Chapter 5 present the analysis and design and this chapter mainly discuss deeply the two phases.

1. loadcell prediction
2. Find important relationships for top level decision making.

5.2 Loadcell prediction

In my approach, here it is going to implement a decision support system for the loadcell selection process. The customer has to be select his application type and the capacity (max load), then the system suggests the best loadcell families to the customer. Customers can compare three load cell families in the suggested list.Fig.3 shows the design diagram of and Fig.4 explain the flow chart of the loadcell prediction.

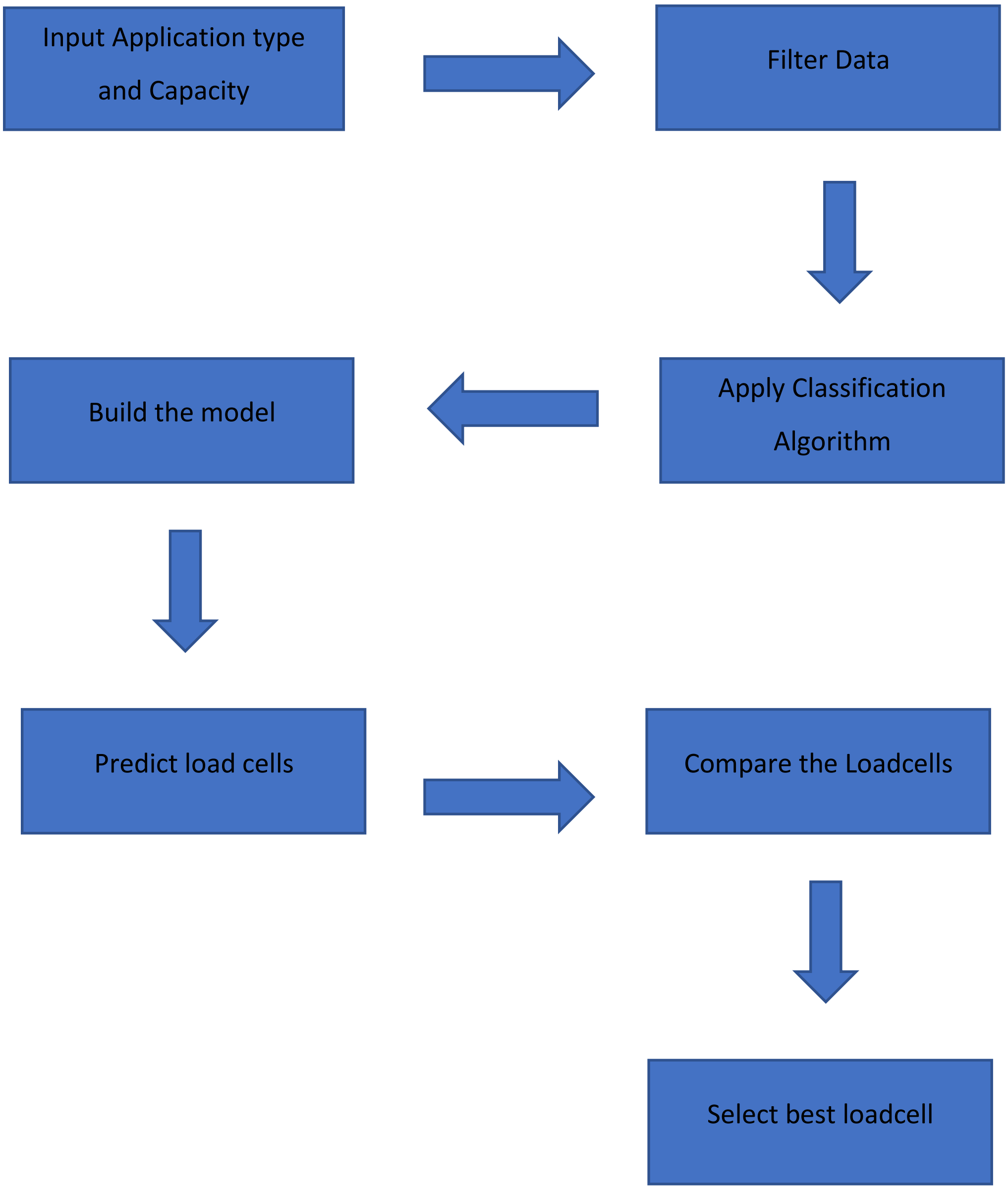


Figure 3: Design of Loadcell prediction

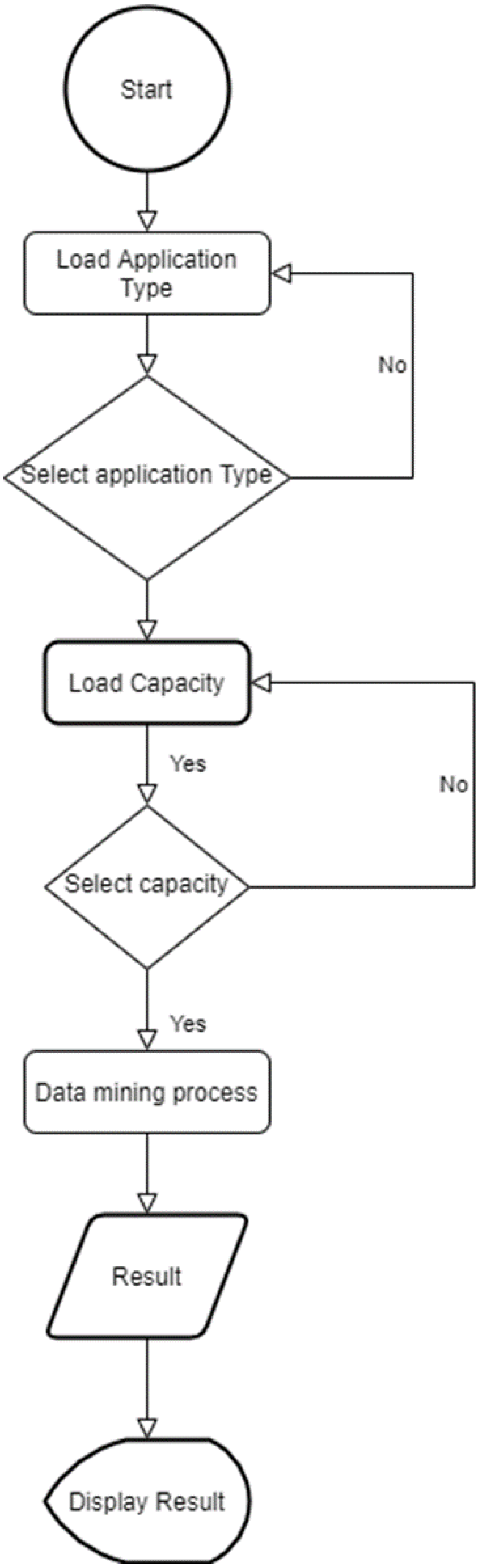


Figure 4: Flow chart of decision support web system

5.3 Find important information for top level decision making

The second phase of the web system is to find important information for top-level decision making by finding the relationships between the loadcell family and the other properties of the loadcell. The sales team or customer supporting team can select the required attributes to find the required relationship.

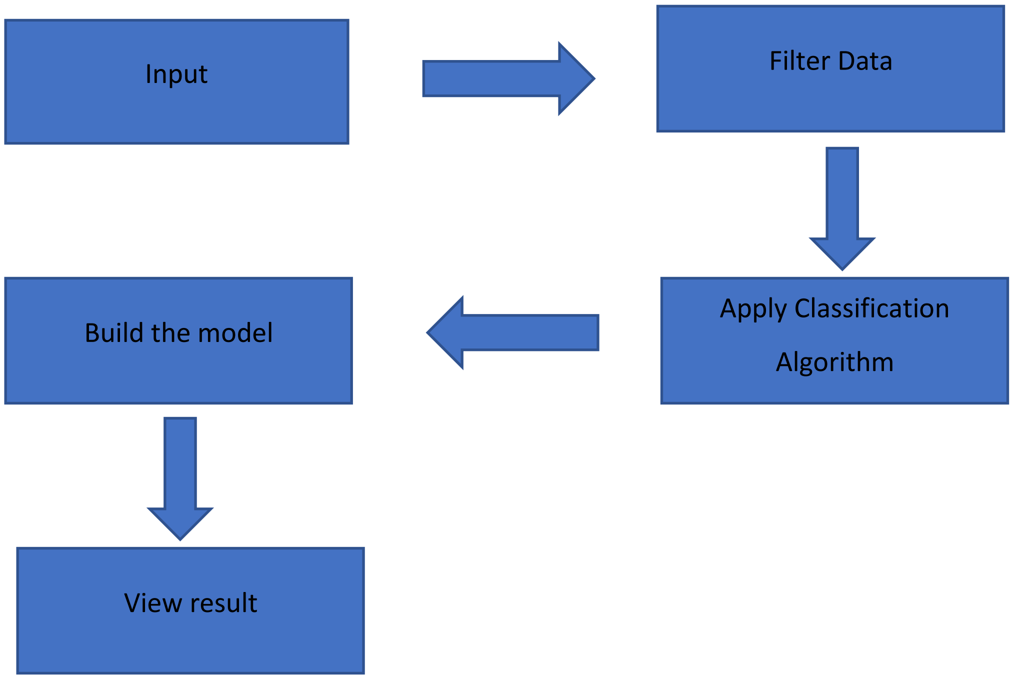


Figure 5: Design of find important information for top level decision making

5.4 Back end Database

Initially, this record set starts with 9500 records and it covers 2018-01-01 to 2019-01-01 time period. This database is updated on time after joining this module to the loadcell manufacturer web site.

5.5 Data Preprocessing

Initially, row data was in Excel, e-mail, note format and the data cleaning process was done by applying the following methods.

1. Remove Duplicate
2. Removing unwanted data.
3. Filling missing data.
4. Correct incorrect data

5.6 Decision making

Decision making is based on the Naïve Bayes theory and the result is ordered according to the probability.

5.7 Software development life cycle (iterative model)

The iterative model is a special implementation of a software development lifecycle (SDLC) that focuses on a simplified initial implementation, which then gradually increases in complexity and functionality until the final system is ready. When considering the iterative method, the concept of incremental development is often also used generously and interchangeably, describing the incremental changes that were made during the design and implementation of each new iteration.

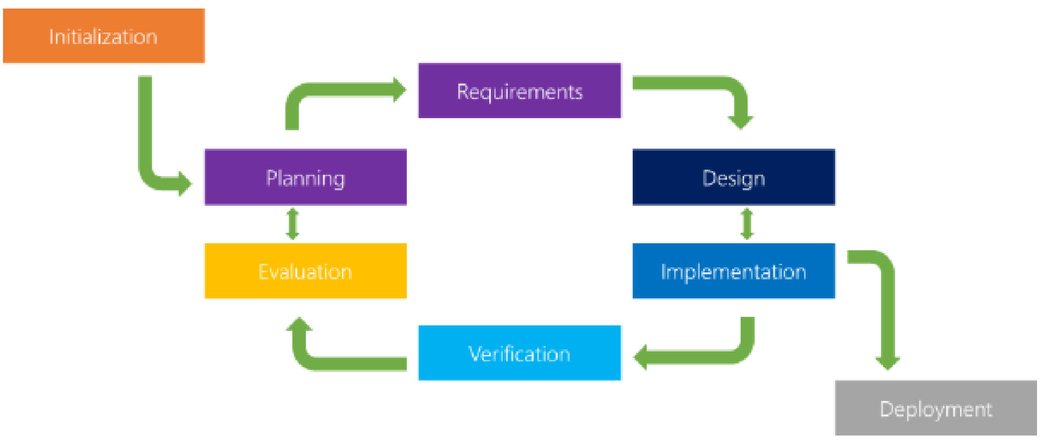


Figure 6: Iterative Modal

5.8 Summary

In this chapter, we discussed the Loadcell prediction, Find important information for top-level decision making, back end database, Data processing, decision making, software development life cycle thoroughly. Next chapter we will discuss the implementation of the research.

CHAPTER 6

6. IMPLEMENTATION

6.1 Introduction

In chapter 5 we discussed the analysis and design and Chapter 6 presents the implementation of the research solution. The total solution is divided into two phases and under this chapter, each phase implementation will clearly express.

6.2 Load cell prediction Phase

Mainly this interface is designed for customers and the results are clearly presented as a percentage. The layout of this page is the focus to provide a facility to compare three load cell families once in single interface. By comparing the customer can get a better idea about the loadcell selection properties and the customer can select any load cell among them for his requirement. Also, present the percentages of each property which experienced previous customers.

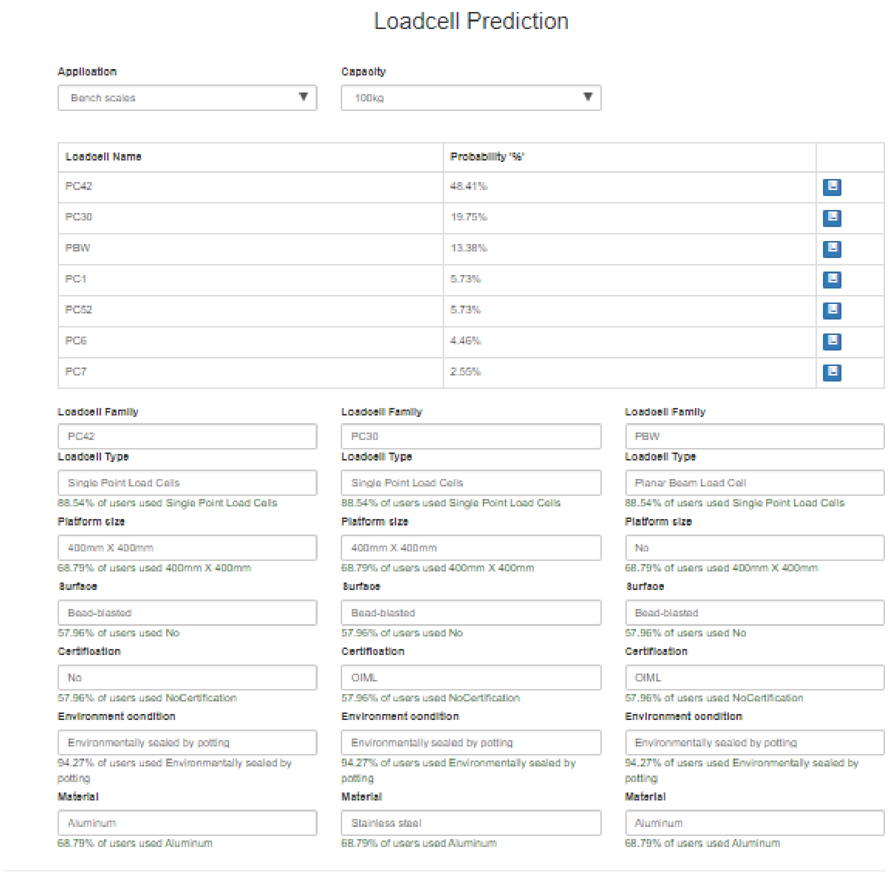


Figure 7: loadcell selection interface for customer

6.3 Find important information for management decision

This phase mainly aims to users the sales team and the engineering team and they can collect important information from the data. This data analyzes based on the load cell family. The users can get the output of the following criteria showing in the Fig.7

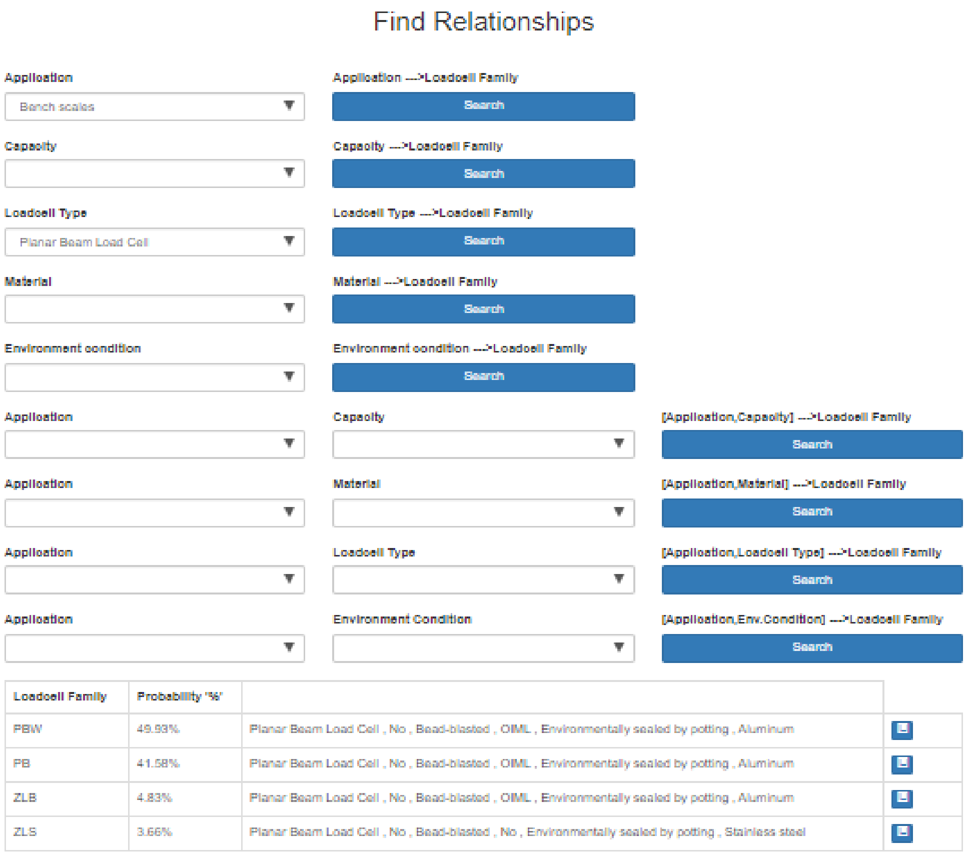


Figure 8: Interface to find relationships

6.4 User Interfaces

Basically, the user interface is developed using ASP.Net MVC technology, which is the latest web technology provided by Microsoft. All web interface development is done using the latest technology and the use of jQuery Ajax for all client-side scripts. In addition, all data is formatted in. JSON to improve interaction with the user interface. In application development, user interfaces design simply and user friendly.

6.5 Summery

This chapter discussed about the implementation of this research. Basically implementation divide in to two parts and first part is for the customers to predict correct load cell for his or her requirement. Second part is for the people who want important information to take necessary decisions. Chapter 7 present the evaluation of this research work.

CHAPTER 7

7. EVALUATION

7.1 Introduction

This chapter focuses on how to test strategies that are implemented to sub-interrogate in terms of evaluative measures for the selected data mining technique, such as percentage of accuracy, TP and ROC for classification.

7.2 Evaluation of Classification techniques

In this research, we used three classification algorithms J48, Naïve Bayes, Decision Table and trained the collected data. Here we used the WEKA tool for data mining purposes, and we compared different algorithm's results that are generated using the WEKA tools in the data modeling stage.

Measurement	**Formula**	**Description**
Precision	TP/(TP + FP)	The percentage of positive predictions those are correct.
Recall / Sensitivity	TP / (TP + FN)	The percentage of positive labeled instances that were predicted as positive.
Specificity	TN / (TN + FP)	The percentage of negative labeled instances that were predicted as negative.
Accuracy	(TP + TN) / (TP + TN + FP + FN)	The percentage of predictions those are correct.

Table1: Classification Evaluation Measurements

Using the measures mentioned above, we can obtain the TP speed, the FP speed, the F measurement, and the ROC area. The TP is equal to the sensitivity, and the FP is equal to the specificity 1 of the F measurement, calculated by the precision and the feedback. The area under the ROC curve determines the overall ability of the test to distinguish the usefulness and futility of an unnecessary test as an area of 0.5. An ideal test has an area of 1.00. In general, the best models have a higher TP speed, a lower FP speed and a ROC space close to 1.00.

Technique	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC
J48	0.981	0.016	0.964	0.981	0.972	0.991
Naïve Bayes	0.981	0.000	0.972	0.981	0.975	1
Decision Table	0981	0.016	0.964	0.981	0.972	0.981

Table 2: Comparison of classification techniques

Table 2 shows that all the techniques give similar results and very few value differences, among those techniques, Naïve Bayes show higher accuracy value among all techniques. The naïve Bayes selected as best technique for most of the data sets which was selected according to the user requirement.

7.3 Evaluation of load cell prediction tool

Fig. 9 and Fig. 10 shows the result for given user requirement (application type= bench Scales ,Capacity =100kg). The results generated using WEKA tool and developed tool. Both results are similar and accurate for all user requirement criteria.

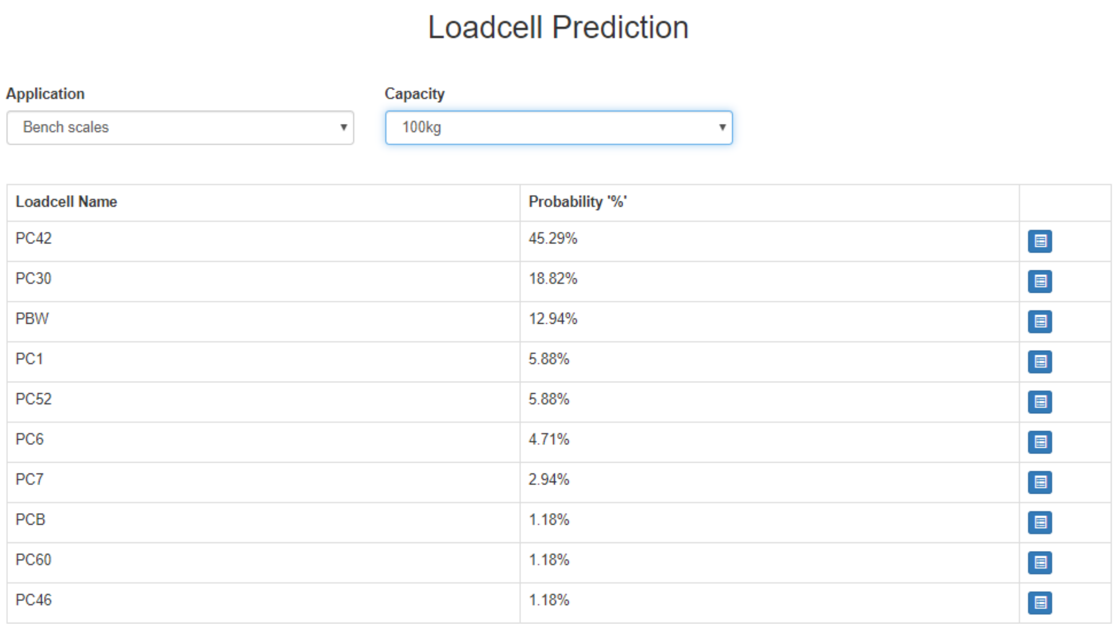


Figure 9: Data mining result scenario

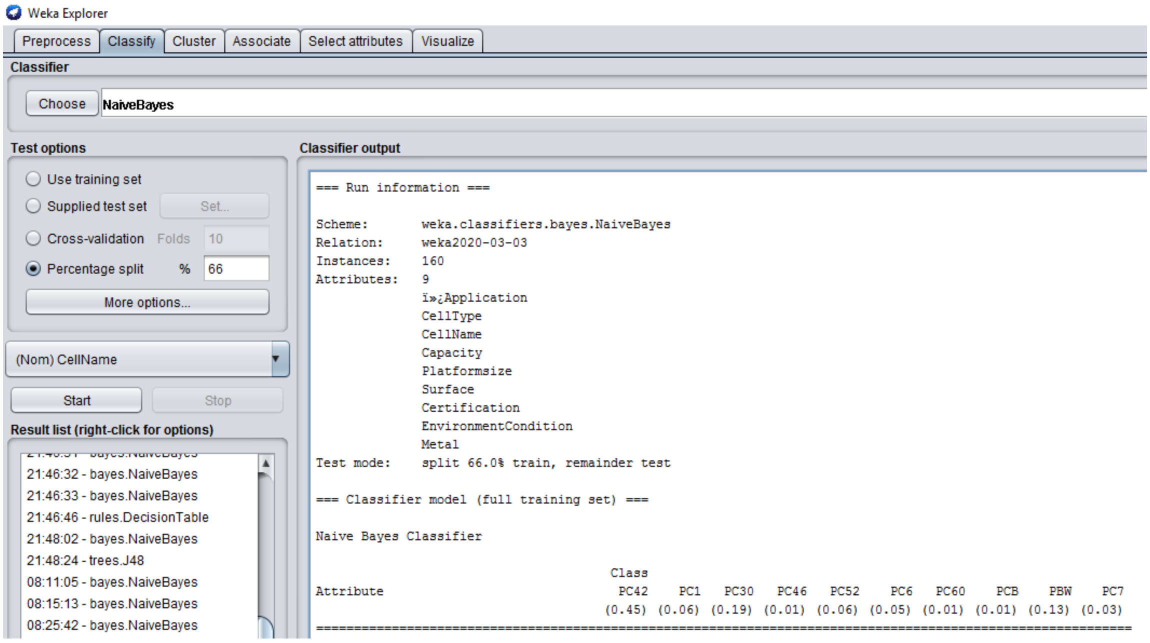


Figure 10: WEKA result for given scenario

7.4 Summary of Evaluation

This chapter presented the evaluation technique used in this research domain. Mainly there are two evaluations considered. In the first part, we compared the three classification techniques namely J48, Naïve Bayes, Decision Tables and found the Naïve Bayes as the Best classification technique for this work. In this stage, we compared the data mining result of the given data set according to the user requirement. In the second part, we compared the WEKA result with the developed tool and fond the results are very similar.

CHAPTER 8

8. Conclusion and Further Work

8.1 Conclusion

Data mining techniques are used in diverse industries for various purposes yet in the loadcell industry data mining techniques are not frequently used. In this research, the Naïve Bayes classification method is used for decision supporting to select proper loadcells for the customer's application. Past customer handling data is used for decision making and some important relationships between cell families and other properties are captured.

As a result of this research, a web module is developed to amend the existing web site of the customer. That module will help the customer to select the most appropriate loadcell family for his weighing application without having any help for the customer support team or the sales team. Also, customers can compare three loadcell families in a single interface. As a result of implementing this module, the sales team and the customer support team can save their time and workload. As well as, the customer support team and sales team can identify the important relationships for decision making.

8.2 Limitations

- The developed system must be joined to the loadcell manufacturer web site and the data insert part should be implemented.

- Before starting the loadcell selection process customer should know the application type and the maximum capacity of his application.

- This system can use only for strain gauge loadcells.

8.3 Future work

According to the research, the developed web modules can generate an accurate result for the user's input by using previous data. But in best use, this module needs to amend to the loadcell manufacturer's web site and real-time data should be inserted to the system database.

REFERENCES

[1] Ahmad Qandil and Adnan I. O. Zaid(2015), “Consideration in the Design and Manufacturing of a Loadcell for Measuring Dynamic Comprehensive Load.”

[2] D. Catherine Rexy(2017), S. Rita Samikann, M.K.Nallakaruppan,” PREDICTION OF WOMEN SURVIVAL RATE DURING PREGNANCY USING CLASSIFICATION AND CLUSTERING”, Research Scholar VIT, University, Vellore, India.

[3] Fitriana Harahap(2018),” Implementation of Naïve Bayes Classification Method for Predicting Purchase”

[4] Harshada Somwanshi and Pramod Ganjewar(2018),” Real-Time Dengue Prediction using Naive Bayes Predictor in the IoT”

[5]Pradeep K.R, Dr Naveen N.C(2016), ”Predictive Analysis of Diabetes using J48 Algorithm of Classification Techniques ”,Dept. CSE,JSS of Academy of Technical Education, Bengaluru.

[6] Reeta R, Pavithra G, Priyanka V, and Raghul J S(2018), ” Predicting Autism using Naive Bayesian Classification Approach”

[7] S. Agrawal, S. Vishwakarma, K, and A. Sharma(2017), “Using Data Mining Classifier for Predicting Student’ s Performance in UG Level,” Int. J. Comput. Appl., vol. 172, no. 8, pp. 39–44.

[8] Smita Jhajharia, Seema Verma, Rajesh Kumar(2016),” A Cross-Platform Evaluation of Various Decision Tree Algorithms for Prognostic Analysis of Breast Cancer Data” Department of Computer Engineering Banasthali University.

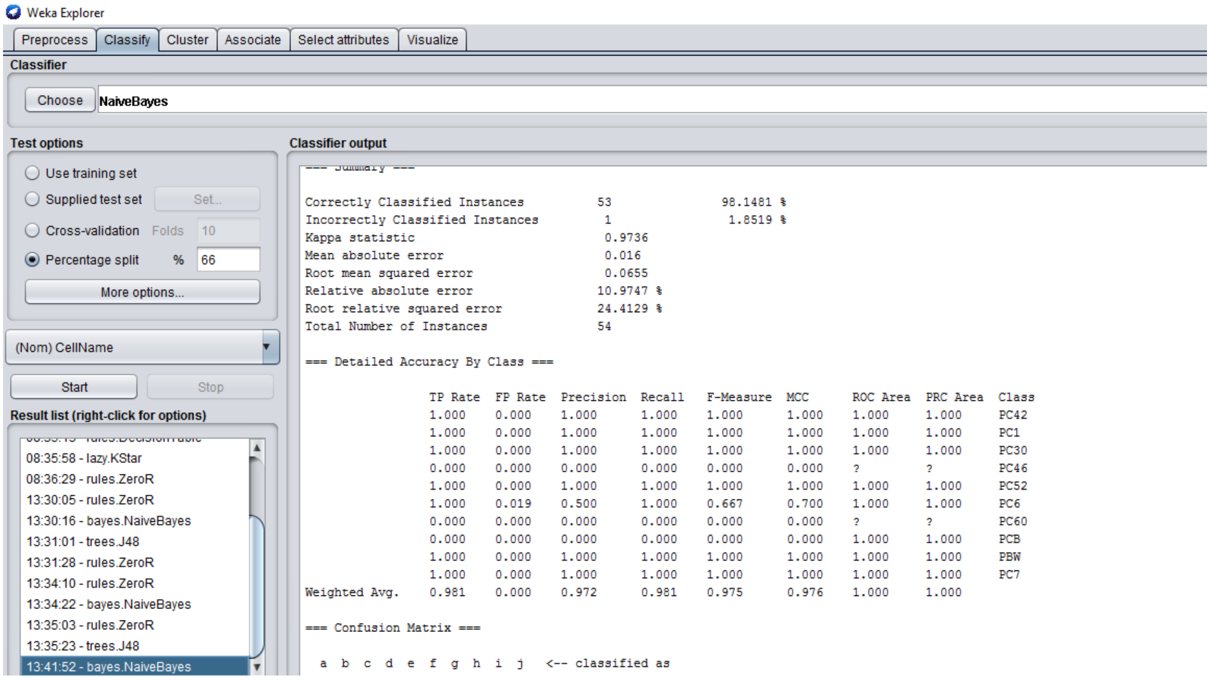
[9]Venkatesh(2018), Classification and Optimization Scheme for Text Data using Machine Learning Naïve Bayes Classifier, University Visvesvaraya College of Engineering ,Department of Computer Science and Engineering ,Bangalore, India.

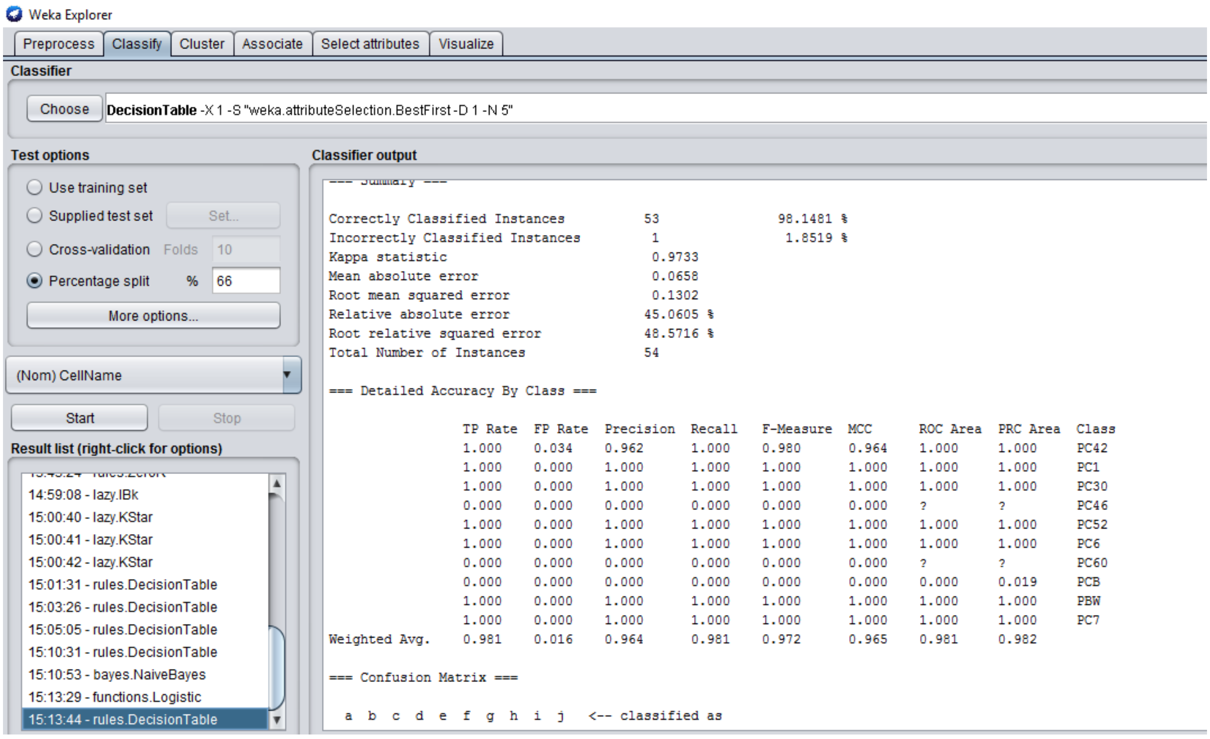
[10] Wei Zhang and Feng Gao(2013), “Performance Analysis and Improvement of Naïve Bayes in Text Classification Application”, in proceedings of IEEE Conference.

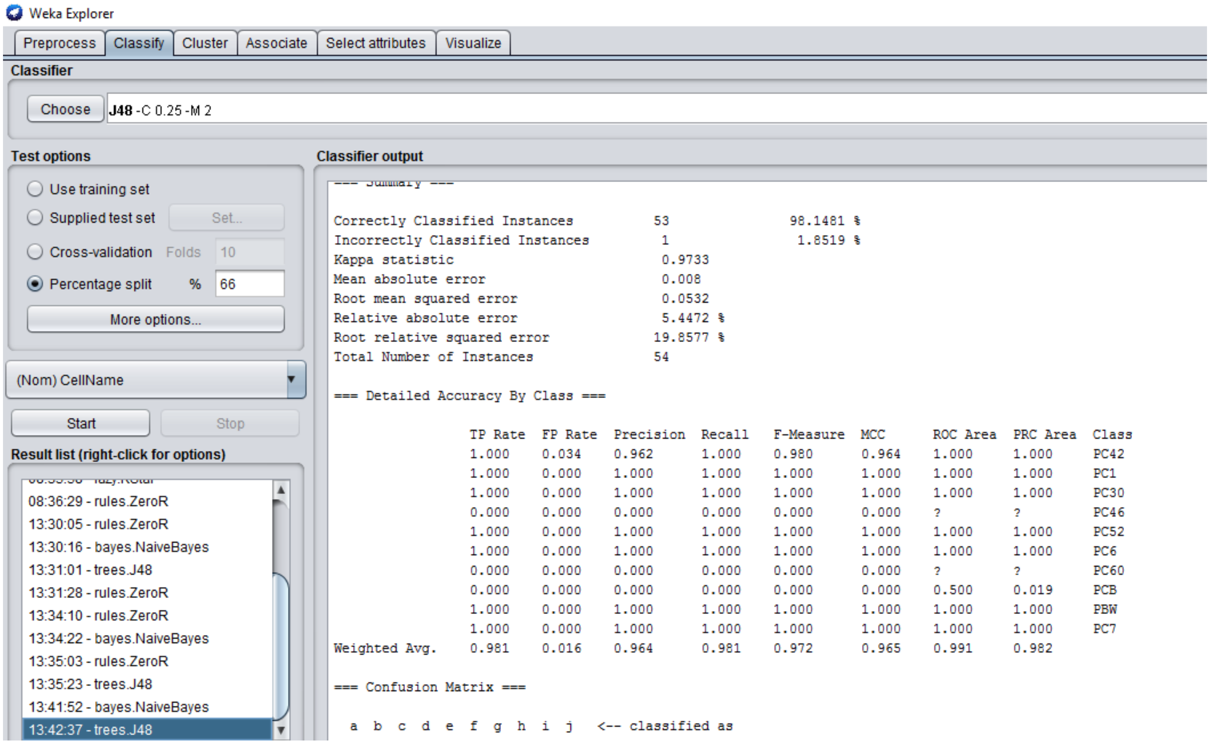
[11] Wilmar Hernandez(2006),” Improving the Response of a Load Cell by Using Optimal Filtering”, Department of Circuits and Systems in the EUIT de Telecomunicacion at the Universidad Politecnica de Madrid (UPM), Campus Sur UPM, Ctra. Valencia km 7, Madrid 28031, Spain.

Appendix A

Algorithm comparison using data set






Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) CellName
Start
Stop
Result list (right-click for options)
08:36:29 - rules.ZeroR
13:30:05 - rules.ZeroR
13:30:16 - bayes.NaiveBayes
13:31:01 - trees.J48
13:31:28 - rules.ZeroR
13:34:10 - rules.ZeroR
13:34:22 - bayes.NaiveBayes
13:35:03 - rules.ZeroR
13:35:23 - trees.J48
13:41:52 - bayes.NaiveBayes
13:42:37 - trees.J48
Classifier output
Correctly Classified Instances 53 98.1481 %
Incorrectly Classified Instances 1 1.8519 %
Kappa statistic 0.9733
Mean absolute error 0.008
Root mean squared error 0.0532
Relative absolute error 5.4472 %
Root relative squared error 19.8577 %
Total Number of Instances 54
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
1.000 0.034 0.962 1.000 0.980 0.964 1.000 1.000 PC42
1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 PC1
1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 PC30
0.000 0.000 0.000 0.000 0.000 0.000 ? ? PC46
1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 PC52
1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 PC6
0.000 0.000 0.000 0.000 0.000 0.000 ? ? PC60
0.000 0.000 0.000 0.000 0.000 0.000 0.500 0.019 PCB
1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 PBW
1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 PC7
Weighted Avg. 0.981 0.016 0.964 0.981 0.972 0.965 0.991 0.982
=== Confusion Matrix ===
a b c d e f g h i j <-- classified as

Appendix B

Result for given criteria

Application type: Bench scales

Capacity: 100kg

13:04:38 - bayes.NaiveBayes

```
=== Summary ===

Correctly Classified Instances          53               98.1481 %
Incorrectly Classified Instances         1                1.8519 %
Kappa statistic                          0.9736
Mean absolute error                      0.016
Root mean squared error                  0.0655
Relative absolute error                 10.9747 %
Root relative squared error             24.4129 %
Total Number of Instances               54

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     PC42
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     PC1
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     PC30
                 0.000    0.000    0.000      0.000    0.000      0.000    ?         ?         PC46
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     PC52
                 1.000    0.019    0.500      1.000    0.667      0.700    1.000     1.000     PC6
                 0.000    0.000    0.000      0.000    0.000      0.000    ?         ?         PC60
                 0.000    0.000    0.000      0.000    0.000      0.000    1.000     1.000     PCB
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     PBW
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     PC7
Weighted Avg.    0.981    0.000    0.972      0.981    0.975      0.976    1.000     1.000

=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i  j   <-- classified as
 25  0  0  0  0  0  0  0  0  0 |  a = PC42
  0  4  0  0  0  0  0  0  0  0 |  b = PC1
  0  0 14  0  0  0  0  0  0  0 |  c = PC30
  0  0  0  0  0  0  0  0  0  0 |  d = PC46
  0  0  0  0  2  0  0  0  0  0 |  e = PC52
  0  0  0  0  0  1  0  0  0  0 |  f = PC6
  0  0  0  0  0  0  0  0  0  0 |  g = PC60
  0  0  0  0  0  1  0  0  0  0 |  h = PCB
  0  0  0  0  0  0  0  0  5  0 |  i = PBW
  0  0  0  0  0  0  0  0  0  2 |  j = PC7
```

Loadcell Prediction

Application: Bench scales

Capacity: 100kg

Loadcell Name	Probability '%'
PC42	45.29%
PC30	18.82%
PBW	12.94%
PC1	5.88%
PC52	5.88%
PC6	4.71%
PC7	2.94%
PCB	1.18%
PC60	1.18%
PC46	1.18%