

**“Find out a new product value quality prediction in market using Data mining”**

**BY  
MEHEDI HASAN  
ID: 153-15-6615**

**MOSHAROF HSSAIN  
ID: 153-15-6652**

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering.

Supervised By  
**Masud Rabbani**  
Lecturer  
Department of CSE  
Daffodil International University

Co-Supervised By  
**Ahmed Al Marouf**  
Lecturer  
Department of CSE  
Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY  
DHAKA, BANGLADESH  
9 JULY 2020**

## APPROVAL

This project title is “**Find out a new product value quality prediction in market using Data mining**”, submitted by Name: Mehedi Hasan, ID No: 153-15-6615, Name: Mosharof Hossain, ID No: 153-15-6652 to the Department of Computer Science of Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents.

### BOARD OF EXAMINERS



**Dr. Syed Akhter Hossain**  
**Professor and Head**

**Chairman**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University



**Subhenur Latif**  
**Assistant Professor**

**Internal Examiner**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University



**Raja Tariqul Hasan Tusher**  
**Senior Lecturer**

**Internal Examiner**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University



**Dr. Md. Saddam Hossain**  
**Assistant Professor**

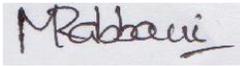
**External Examiner**

Department of Computer Science and Engineering  
United International University

## DECLARATION

We here by declared, this project has been done by us under the supervision of **Masud Rabbani**, Lecturer, Department of CSE in Daffodil International University. We also declare that neither this project or any part of this project has been submitted elsewhere for awards of degree.

### Supervised by:



---

**Masud Rabbani**  
Lecturer  
Department of CSE  
Daffodil International University

### Co-Supervised by:



---

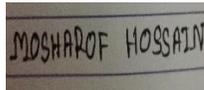
**Ahmed Al Marouf**  
Lecturer  
Department of CSE  
Daffodil International University

### Submitted by:



---

**Mehedi Hasan**  
ID: 153-15-6696  
Department of CSE  
Daffodil International  
University



---

**Mosharof Hossain**  
ID: 153-15-6652  
Department of CSE  
Daffodil International  
University

## ACKNOWLEDGEMENT

First of all, we express our sincere gratitude and thank the Almighty Lord for his divine blessing and we have been able to carry out the project since last year.

We are very grateful and sincerely thanks to **Masud Rabbani, Lecturer** in the Department of CSE Daffodil International University , Dhaka. The deep knowledge and intense interest of our supervisor in the field of "**Data Mining**" for the realization of this project. Thanks to his infinite patience, his wise advice, his constant encouragement, his energetic and persistent supervision, his constructive criticism, his valuable advice, the reading of many small projects and their correction at each stage, the project can be completed.

We would like to express our sincere thanks to **Prof. DR .Syed Akhter Hossain, Professor and Head**, Department of CSE, for his kind assistance in carrying out our project, as well as other faculty members and staff of the CSE Department at Daffodil International University.

We would like to thank all our course mate in Daffodil International University, who participated in this discussion during the end of our courses.

Finally, we must recognize and respect the constant support and patience of our parents.



## **ABSTRACT**

The objective of this paper is to inform the information system manager and business analyst about the role of machine learning techniques in business data mining. Data mining is a fast growing application area in business. Machine learning techniques are used for data analysis and pattern discovery and thus can play a key role in the development of data mining application. Understanding the strengths and weaknesses of these techniques in the context of business is useful in selecting an appropriate method for a specific application. The paper, therefore, provides an overview of machine learning techniques and discusses their strengths weaknesses in the context of mining business data. A survey of data mining application in business is provided to investigate the use of learning techniques. Rule induction was found to be most popular, followed by neural networks and case – based reasoning. Most application were found to be most popular.

# TABLE OF CONTENTS

## CONTENTS PAGES

Board of Examiners

DECLARATION ii

ACKNOWLEDGEMENT iii

ABSTRACT v

LIST OF FIGURES viii

## CHAPTER 1: INTRODUCTION 9-3

1.1 Introduction 9

1.2 Motivation 1

1.3 Rational of the Study 1

1.4 Reserch question 1

1.5 Expected Outcome 1

1.6 Report Layout 2

## CHAPTER 2: LITERATURE REVIEW

4-7

2.1 Introduction 4

2.2 Related work 4

2.3 Research summary 6

2.4 Scoop of the project 7

2.5 Challenges 7

## CHAPTER 3 : REQUIREMENT SPECIFICATION 8-12

3.1 BPM 8

3.2 Requirement analysis 9

3.3 Functional Requirement 10

3.4 Non-Functional Requirement 10

3.5 Use case 9

- 3.6 Logical Data model12
- 3.7 Design Requirements11

**CHAPTER 4 : DESIGN SPECIFICATION12-14**

- 4.1 Front- End design13
- 4.2 Data mining application tool 14

**CHAPTER 5:IMPLEMENTATIONAND TESTING15-24**

- 5.1Algorithm analysis15
- 5.2Machine learning algorithmprocess15-16
- 5.3Formula analysis16-17
- 5.4Regression Technique 18
- 5.5Lasso regression19
- 5.6Predicting data with formula19
- 5.7Calculation with table 20
- 5.8 Data calculation 21
- 5.9Statical View 22
- 5.10 Result23
- 5.11 Discussion of result 24

**CHAPTER 6: CONCLUSION25-28**

- 6.1 Conclusion25

**APPENDIX 25-26**

**REFERENCES26-28**

**PLAGIARISM REPORT40**

## LIST OF FIGURES

### FIGURES PAGES

Figure 3.1: Data Flow Diagram	8
Figure 3.2: Use-case modeling of machine learning	10
Figure 4.1: A Screenshot of process of data mining	13
Figure 5.2: Machine learning techniques	16
Figure 5.7: Regression statistics view	21
Figure 5.8: Price range graph	22
Figure 5.9: A screenshot of the price range	23
Figure 4.6: A Screenshot of the product percentage	24

# CHAPTER 1

## INTRODUCTION

### 1.1 Introduction

The tech market includes many other small markets. Technology includes everything from a computer, a projector, and other smart devices to the Internet and virtual reality. Technology has significant effects on business operations. The technological infrastructure affects the culture, efficiency and relationships of a company. It also affects the security of confidential information and business benefits. [1] (E.W.T. Ngai et.al.2010) in this article, he is a representative of academic literature in this national method of approval, place and world. Apply FFDT by examining the article browsing step. One of the greatest benefits that technology offers to companies is increased productivity.

[2] (Moro et.al.2016) described in his article the approach of using data processing to predict the performance parameters of the publications published on the Facebook pages of the brands.

Machine learning methods are powerful tools for extracting data from large and noisy databases, and provide researchers with the opportunity to better understand consumer behavior and improve the performance of marketing operations. The problem must be defined and the appropriate machine learning analysis method selected.

In this study, our main contribution, (1) Mainly focused on data mining and machine learning. (2) Product value case prediction. (3) Analysis of the cost of the product value. (4) Combination table of various products using data mining. (5)

Formulation of information and analysis of algorithms. (7) Give a visualization and a precision of the result.

## **1.2 Motivation**

Mainly focusing on data mining and machine learning. Table of different combinations of products using data extraction. The marketing data extraction application includes analysis of retail sales, analysis of the shopping cart, analysis of product performance and analysis of market segmentation. British Telecom uses to distinguish between users and users of telecommunications products. These new information structures meet customers closely. These types of products are more valuable than the other product. Calculating the data estimates that a new product can buy from the customer at an average price. The last step is to assess the performance of the algorithm. As we have seen, the linear regression model achieves practically the best intercept and slope value, which gives a line that best suits the data

## **1.3 Rationale of the Study**

We have studied linear regression algorithms. We implemented multiple linear regression using the Scikit-Learn machine learning library. This document focused on data mining and machine learning, forecasting product value, analyzing the cost of product value. Provide visibility and accuracy of the result. Therefore, accurate quality forecasts play an important role in providing high quality products to further increase competitiveness.

## **1.4 Research Questions**

Research questions are what we face from the start of the project to the end of the project. This is a very important steps during the period of our project. Basically this is a questionnaire of what we faced during this period.

1. First question was what is the kind of project we are doing? Is it a project based on research or a development project?
2. How do we build the project?
3. Which product looks more standard to you?
4. Which product will be most attractive to you?
5. In the price of a good product is high, how much do you want to buy?
6. Is the local product or brand is most acceptable to you?

7. If the price and value of the brand product is worse than the local, will you buy the product?
8. Which of the following products do you like most? (TV,MOBILE,PROJECTOR)
9. Why do you like the product more? (1.cause of prices 2. Lasting of product 3.Brand 4.Good looking )
10. Do you trust a new company to purchase your product?
11. Which quality are you want for buying a new product?

### **1.5 Expected Outcome**

1. Easy and nice business planning. .
2. Better expressions of learning things.
3. Product value case prediction
4. Cost analysis of product value.
5. Giving a visualization and accuracy of result.
6. These features will benefit industry, a new company, new customer, new product if it were to be implemented.
7. Data mining web analysis
8. Product quality measurement.
  
9. One of the biggest advantages technology offers businesses is gains in productivity.

### **1.6 Report Layout**

Chapter 1, Introduction; Here we talked about the introduction of our application. Also the motivation and the rationale behind the project with expected outcomes.

Chapter 2, Literature review; Here we discussed about Related works in this field, research summary and the scope of the project and also the problems we faced doing this work.

Chapter3, Requirement Specifications; Here we talked about the requirement specifications for the application such as requirement analysis and functional needs in this application.

Chapter 4, Design Specification; Here, the necessary frontend and backend designs were developed.

Chapter 5, Implementation and Testing; How we implemented different objects and the testing process of our application.

Chapter 6, Conclusion and Future Scope; We talked about the limitations of our application and the future scope of this application.

## CHAPTER 2

### LITERATURE REVIEW

#### 2.1 Introduction

Technology covers everything from computers, projectors, and other smart devices, to the Internet and virtual reality. Technology has significant effects on business operations. Technological infrastructure influences culture, efficiency, and business relationships. It also affects the security of confidential information and business benefits. [1] (E.W.T. Ngai et.al.2010) in this document, is a representative of the academic literature on this method of national, global and global approval. Apply FFDT to the Review Step section. One of the greatest benefits that technology offers to companies is increased productivity.

[2] (Moro et.al.2016) described his document, the approach of using computers to predict the performance parameters of the publications published on the Facebook pages of the brands. The problem should be defined and the appropriate method for analyzing machine learning analyzed.

#### 2.2 Related Work

Data mining is a process used by companies to convert raw data into useful information. By using software to find patterns in large amounts of data, companies can learn more about their customers to develop more effective marketing strategies, increase sales, and reduce costs.

Philip S. Yu and. Al. 1999 [3] We examine new developments in data mining and their application of personalization in electronic commerce. Personalization is what marketers and editors try to tailor to the site or advertise and promote a customer's products based on the past behavior and conclusions of each individual mind.

Michael J. Shaw et al. 2001 [4] A systematic methodology is proposed that uses data mining and knowledge management techniques to manage marketing information and support marketing decisions. This methodology can be used as a basis to improve customer relationship management. q 2001 Elsevier Science B.V. All rights reserved.

Indranil Bose et al. 2001 [5] Machine learning techniques are used for data analysis and model discovery and therefore can play a key role in the development of the data mining application.

Dong-Hyeon Kim et al. 2018 [6] A breakthrough in the field of artificial intelligence was discovered primarily thanks to advances in machine learning, which allow machines to learn, improve, and perform a specific task using data without being explicitly programmed

Pal Sundsoy et al. 2014 [7] We show that this data-driven approach using machine learning and social media analytics leads to higher conversation rates than the best practice marketing approach. We also show that historical natural adoption data can be used to model when campaign response data is not available.

### **2.3 Research Summary**

The following product is Television , Mobile , Projector . Many brand projector is market available just like Hitachi, Sony, Panasonic, View sonic, Optoma , Benq , Ricoh etc .Here is Many local projector is available here. But in this projector most popular is Brand Projector like Sony. Here is many Television and mobile is market available , But in time a new company goods on the market people trust level in Trust level(0-25%) is 29.7% , Trust level(26-50%) is 40.1% , Trust level(51-75%) is 19.2% , Trust level(76-100%) is 5.2% .

Here is most acceptable product is Brand product .Brand product Television , Mobile, Projector compare in Sony ,Walton ,Mi , Lenevo . Sony product is most acceptable in people, Because of price, lasting of product ,brand product, Good looking. R value = 0.461299 that mean its moderate value in reggration .

So, if the any brand price rate is high and this brand is trusted also in customer this product value rate increase .

This types of product valuable more than the other product .Data calculation view a new product can buy in customer in approximately mid-level price .

## **2.4 Scope of the Project**

Scope of the project means what we are going to cover in this project. They are

1. Importance of business reality development
2. Using machine learning algorithm .
3. Using Data mining
4. Using linear regression algorithm .
5. Ordinary Least Squares (OLS)
6. Ridge Regression

## **2.5 Challenges**

We learnt many challenges while completeing this paper. They are

1. Finding better product quality ensure techniques
2. Using lonear regression analysis
3. Building new formula.
4. Downloading predicting many formula.
5. Writing market segmant analysis.
6. Moving the equation.
7. New product analysis and create predicted output.

## CHAPTER3

### REQUIREMENT SPECIFICATIONS

#### 3.1 Business process model

BPM or Business Process Modeling is the mobility of industrial production systems. This is done in business process management and systems engineering. Business process modeling allows you to analyze, improve and automate existing or ongoing activities in an industry. Business analysts generally work to model business processes in an industry. Sometimes this work is done by a team that is not someone. They are extremely sophisticated and competent people. This modeling is carried out by retrieving the event logs from the events. This is done using a data extraction tool called data extraction processing. Marketing is often aimed at increasing the speed of the business or reducing the cycle; Increased efficiency; or to reduce the price, such as labor, materials, scrap or capital costs. In practice, management's decision to invest in business process modeling often motivates the need to document the requirements of the IT project [11]. Software developers use the enterprise word processing model in the same way. One of the objectives of this model is to illustrate key questions that include objectives, structures, operational activities and policies. This processing is created by creating a diagram. This diagram is called an organization chart or data flow diagram. Let's create this graph to show the data flow of our applications.

A data flow diagram or DFD shows the data you are browsing. In a single application, data transfers from several sources to several branches. Software developers use software diagrams to show the flow of information from branch to branch in their application. The main purpose of this diagram is to show how data or information progresses in the application system.. These features will benefit the industry, the new business, the new customer, the new product, if implemented.

In Figure 3.1, we can see our applications Data Flow Diagram. It explicitly describes our systems flowing of information.

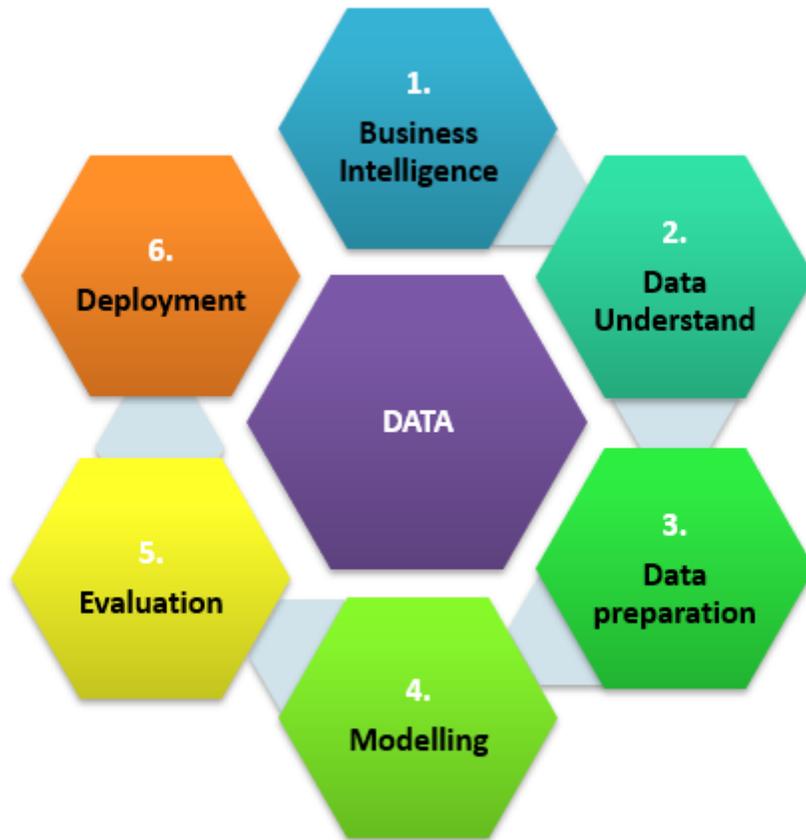


Figure 3.1: Data Flow Diagram of business model

People can take lessons, English and words. Data mining is a complex process that involves several iterative steps. Here is an overview of this process. An engineering process involved in the development of the basic edge system. Searching for a model in the training dataset is a commonly used strategy.

### 3.2 Requirement Analysis

In systems and software engineering, needs analysis focuses on the tasks that determine the requirements or conditions to satisfy the new or modified product or project, taking into account the contradictory requirements of the different stakeholders, the analysis of software, document it, validate it and manage it. or requirements. of the system

Needs analysis is essential for the success or failure of a software system or project. The requirements must be documented, usable, measurable, verifiable, traceable, linked to the business opportunities or needs identified and defined in detail for the design of the system [12].

In application or software development, the collections of requirements are very important. Furthermore, analyzing these sets of requirements is a very crucial part of development. The requirements are divided into two parts. These two parts are functional requirements and non-functional requirements. Software or an application can perform actions on the card called functional requirements. Sometimes an application behaves in a certain way or the application has a personality, such as its efficiency, the performance problems of the application and many others, what is called the non-functional requirement.

### **3.3 Functional Requirement**

The functional requirement defines a function of a system in software engineering, where a function is described as a specification of behavior between outputs and inputs. Functional requirements may include calculations, technical details, data manipulation and processing, as well as other specific functionalities that define what a system must meet [13]. The software can perform or has some actions that it can perform that are related to the technical functions of an application system. In our document, there are some necessary requirements. These requirements are like maintaining the database. These requirements also apply to functional requirements.

### **3.4 Non-functional Requirement**

There are some requirements that specifies certain criterias in an applications system which could be used in certain times, rather than certain treatments such as how much efficient our application is, user friendliness of our application, execution issue of the software etc. In our paper , it has many functional requirements such as more efficiency, relevanance, optimizing performance, consumption of memory and smooth operations, load quickly. It also has the Its goal is to predict or analyze a set of dependent variables from a set of independent variables or predictors. These features will benefits industry , a new company , new customer, new product if it were to be implemented .

### **3.5 Use Case**

A use case is a possible list of certain actions that can be done in an applications system in Software and Systems Engineering. Use case also defines the relationship between an actor and the processes of an application. It is also known as UML or Unified Modeling Diagram. Users, external systems, humans are known as actors here. Use case is often portraid as the collection of missions or goals of stakholders in software and systems engineering[14]. Software developers uses use case to better understand the level and the interactions of the application.

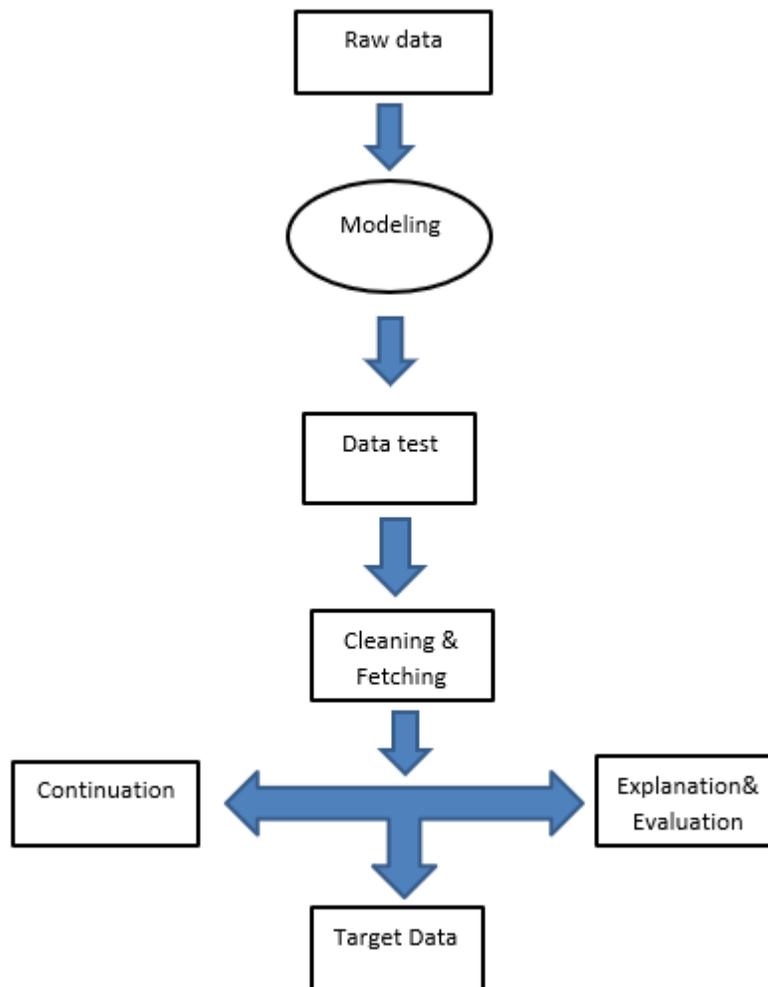


Figure 3.2: Use-case modeling machine learning.

In figure 3.2 we can see the interaction between raw data to target data . And they all are interacting with each other.

### **3.6 Logical Data Model**

There are always a structure of a collection of information. This is called Logical Data Model. They are most typically used in business processes. They seek to capture important things in an industry. They also show the relationship between these important things. This a model which showcases certain actions of a certain database. They can also show the structures of information capacity innovation. This always done mainly for an industry. But the logical data model can also be used in software to show the information structure of Database.

### **3.7 Design Requirements**

Our application has such design requirements and they are

1. In our application, we have not used any authentication protocol because it is a public application.
2. We decided to provide simple and easier user interface.
3. We kept the systemless complicated
4. We tried to keep the application more intuitive
5. We have provided technology to everything from computers, projectors and other smart devices to the Internet and virtual reality. Technology has significant effects on business operations. Technological infrastructure influences culture, efficiency and business relationships.
6. One of the biggest advantages of technology offered by businesses is the acquisition of productivity.

## **CHAPTER 4**

### **DESIGN SPECIFICATION**

#### **4.1 Front-end Design**

Front models exist in a pixel world, the world of logic. A frontend designer (which can be developed by a user interface, a client developer, a user interface engineer, a design engineer, a frontend architect, a designer / developer, a prototype, a unicorn, or Bo Jackson) Development global lazy software. A front-end designer only designs the application's user interface [15]. But it differs from person to person. Sometimes the person who creates the user interface is the person who creates the entire application. The tech market includes many other small markets. Technology covers everything from computers, projectors, and other smart devices, to the Internet and virtual reality. Technology has significant effects Technological infrastructure influences culture, efficiency, and business relationships.

In figure 4.1, we can see the first scene of our application. Everyon business operations. time the user starts the application, they will see this sence.

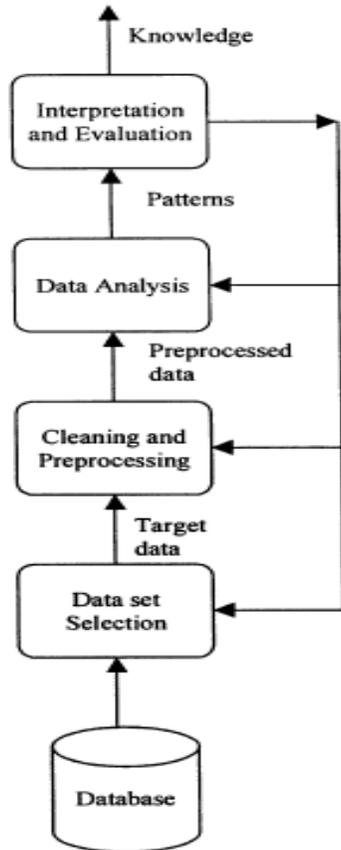


Figure 4.1: A Screenshot of the process data mining

## 4.2 Data mining application list

List of Data mining applications

Area	Application	Technique(s) used	Problem type
Finance	Forecasting bankruptcies [62,57]	NN [62]; RI [57]	Prediction
	Forecasting defaulting loans [40]	RI [40]	Prediction
	Forecasting stock price [6]	NN [6]	Prediction
	Credit assessment [14]	RI [14]	Prediction
	Portfolio management [27]	RI [27]	Prediction
	Forecasting interest rates [29]	NN and CBR [29]	Prediction
	Forecasting price of index futures [58]	RI and NN [58]	Prediction
	Corporate bond rating [13]	CBR and RI [13]	Prediction
	Loan approval [7]	RI and visualization [7]	Prediction
	Risk classification [52]	RI [52]	Classification
	Financial customer classification [43]	RI [43]	Classification
	Detecting delinquent bank loans [28]	NN and visualization [28]	Detection
	Identifying suspicious transactions [12,30,50]	NN [12]; RI [30,50]	Detection
Risk management [25]	Visualization [25]	Detection	
Telecom	Forecasting network behavior [38,46]	RI [38]; NN and RI [46];	Prediction
	Call tracking [1]	CBR [1]	Classification
	Churn management [9]	RI [9]	Classification
	Fraud detection [12,47,53]	RI [12]; Visualization [47,53]	Detection
Marketing	Market segmentation [53]	RI [53]	Classification
	Lifestyle behavior analysis [36]	Visualization and RI [36]	Classification
	Online sales support [61]	CBR [61]	Classification
	New opportunity analysis [48]	Visualization [48]	Prediction
	Customer reaction to promotions [11,49]	GA [11]; RI and visualization [49]	Prediction
Web analysis	Similarity assessment of user browsing patterns [16,17,63]	RI [16]; RI and visualization [17]; Rule-based heuristics [63]	Classification and association
	Identification of Web pages that are viewed together [63]	Rule-based heuristics [63]	Association
	Similarity assessment of Web page contents [19,56]	ILP [19]; rule-based heuristics [56]	Association
	Categorization of Web pages based on content [15,60]	Rule-based heuristics [15]; RI [60]	Classification
	Searching for specific Web pages [44]	Visualization [44]	Detection
Others	Litigation assessment [21,47]	RI [21]; NN and RI [47]	Prediction
	Political conflict resolution [24]	CBR and RI [24]	Prediction
	Insurance claim estimation [20]	CBR and statistics [20]	Prediction
	Detecting insurance claims fraud [37]	RI and statistics [37]	Detection
	Exception reporting in healthcare [39]	RI [39]	Detection
	Software cost estimation [35]	NN [35]	Detection
	Customer technical support [33,54]	RI and CBR [33]; CBR [54]	Classification
	Hypothesis formulation about illness [10]	GA [10]	Classification
	Mapping patient symptoms to surgical procedures [55]	RI and NN [55]	Classification
	Scheduling [32,59]	GA [32,59]	Classification
Expenditure of allocated budget [41]	GA [41]	Classification	
Software quality control [1]	CBR [1]	Classification	

## **CHAPTER 5**

### **IMPLEMENTATION AND TESTING**

#### **5.1 Algorithm analysis**

Machine learning is a data analysis technique that teaches computers what to do naturally. Humans and animals learn by experience. Machine learning algorithms use computer methods to "learn" information directly from the data, without depending on a machine. the default equation as a model. Deep learning is a specialized type of machine learning. Machine learning uses two types of supervised learning techniques, which create a model on the so-called input and output data, so that you can produce future results and unsupervised learning, which uncover patterns or structures. hidden intrinsic to the input data.

#### **5.2 Machine learning algorithm process**

Common regression algorithms include the linear model, the nonlinear model, regulation, phase regression, guided and planted decision trees, neural networks, and adaptive neuro-fusion learning.

The benefit of machine learning are the predictions and models that make predictions .To have skill at applied machine learning means knowing how to consistently and reliably deliver high quality prediction after problem.

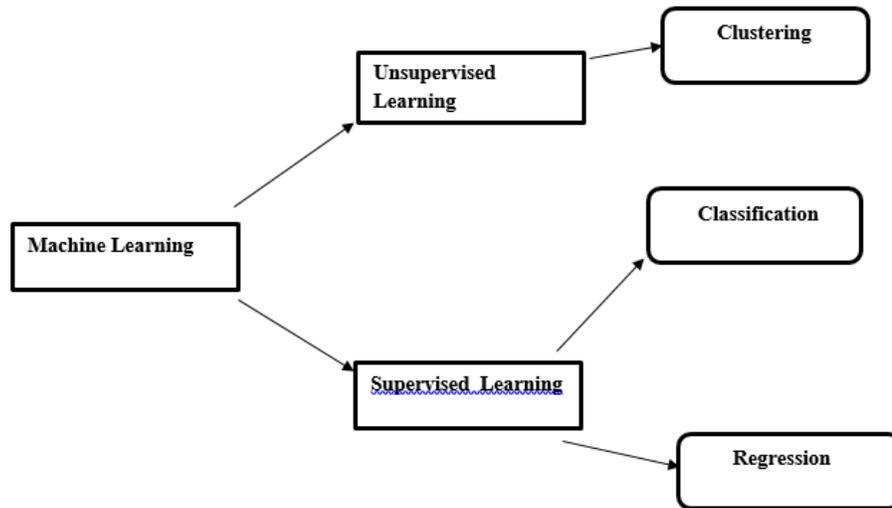


Figure 5.2: Machine learning techniques

Choose to know about monitoring if you need to form a model to make predictions, for example, the value of a future continuous variable, such as stock price or ranking, for example by identifying camera brands on the webcam video footage.

### 5.3 Formula Analysis:

Linear regression is a predictive modeling technique that finds a relationship between the independent variable and the independent variable. The independent variable can be categorical or continuous, while the dependent variable is continuous. The basic Mapping IV and DV function can be linear, quadratic, polynomial or other non-linear functions, but

this article is about linear technique. Regression techniques are widely used to predict product prices, functional forecasts and predicted arrival times

- a) **Categorical:** Takes distinct values: Spam/Not Spam email, Diabetes +ve/-ve
- b) **Continuous:** Can take infinite values, e.g. money, time, weight.
- c) **Dependent:** Outcome of the experiment, house values in this blog
- d) **Independent:** Variable independent of what the researcher does. Area, location, #bedrooms etc.

## 5.4 Regression Techniques:

### I. Ordinary Least Squares (OLS)

In OLS, the goal is to find the most suitable line at the data points. The best match line is obtained by minimizing the sum of the square distances of the prediction line data. This is an unbiased estimate (although the variation is not minimized) because it minimizes the bias for the dataset in this observation space.

$$Y = WX + b$$

$$\min_w ||Y - WX||^2$$

### II. Key Assumptions:

(i) OLS assumes a linear relation of dv and iv.

(ii) Suppose homoscedasticity. Therefore, he suffers from heteroskedasticity. In summary, it is the variability of the dependent variable relative to the independent variable as the value of iv increases. A conical shape in the data distribution diagram indicates heteroskedasticity.

### III. Multivariate Case

In multivariate OLS, the objective function remains similar as for a single variable OLS.

$$Y = W_1X_1 + W_2X_2 + \dots b$$

$$\min_{w,s} ||Y - W_1X_1 + W_2X_2 + \dots ||^2$$

The key problem with multivariate OLS is multicollinearity. This is the condition in which two or more predictors have a high correlation with each other ( $\pm 1$  indicates a 100% correlation, 0 does not indicate a correlation).

In multicollinear data, OLS does not produce a good estimate because the variance term error is significant and OLS only minimizes the bias error and not the variance error. Therefore, we use regularization techniques to minimize errors due to variations and improve our model.

#### IV. Ridge Regression

As discussed above, when the data suffers from multicollinearity, the unbiased estimator like OLS has a high error due to the variance term. The maximum likelihood regression solves this problem by introducing a parameter  $\alpha$  with a regularization term L2 to reduce the weight  $W$  in addition to the least square loss.

$$\min_w ||Y - WX||^2 + \alpha ||W||^2$$

It has a similar assumption as the OLS except for homoskedasticity. Ridge regression shrinks the value of coefficients which are not highly correlated (not exactly zero).

#### 5.5 LASSO Regression:

Similar to maximum likelihood regression, it solves the problem of multicollinearity by regularization. The introduction of a contraction parameter with a standard L1 at weight  $W$  makes it possible to reduce the error of variation in the LASSO regression.

$$\min_w ||Y - WX||^2 + \alpha ||W||^1$$

LASSO makes similar assumptions as OLS except for homoscedasticity. LASSO shrinks the value of coefficients which are not highly correlated to exactly zero.

#### 5.6 Predicting data with formula:

The technique used to describe the relationship between two variables in which one variable (mentioned dependent variable  $y$ ) must be modified as the other variable (independent, explanatory or expected  $x$ ). Linear regression is the statistical technique for constructing a straight line of data, where the regression line is:  $y = a + bx$   $a =$  constant (initially ordered  $y$ ) and  $b =$  gradient (regression coefficient). The value of  $y$  is the expected value and the difference between  $y$  and the observed value is the error. The model is combined by choosing  $a$  and  $b$  to minimize the sum of the squares of the forecast errors (the least squares method). The method produces an estimate of  $b$ , as well as a standard

error and a confidence interval. From there, we can test the statistical significance of b. The regression coefficient (b) tells us that, on average, it changes according to the unit to change the unit to x (explanatory variable), y (response variable). B The variability of the error is considered constant. The error term is generally distributed with an average of zero. Observations are independent (for example, they do not occur during repeated sampling of materials). A similar correlation, the relationship is not symmetric, so if the dependent and independent variables had changed, if all the observations fell in a perfect straight line, a different equation would be obtained.

### 5.7 Calculation With Table :

This allows the analysis of several explanatory variables with a single response variable. The formula is:  $Y = a + b_1x_1 + b_2x_2 + \dots + b_kx_k$  where y is the expected value. The main use of multiple regression is to correct confusion. The observed result y is assumed to be continuous and the variables x are continuous or binary. The coefficients b1, b2 ..., bk are chosen again to reduce the sum of the squares of the difference y-y. When x1, as a treatment group, is a categorical variable and x2 is a continuous variable, such as age (unclear potential), we are talking about congruent analyzes.

Regression Statistics									
Multiple R	0.461299444								
R Square	0.068277399								
Adjusted R Square	0.040043381								
Standard Error	18.94992359								
Observations	35								
ANOVA									
	df	SS	MS	F	gnificance F				
Regression	1	868.3988	868.3988	2.418267	0.129467				
Residual	33	11850.29	359.0996						
Total	34	12718.69							
	Coefficients	andard Err	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	pper 95.0%	
Intercept	43.30503517	3.790597	11.42433	5.21E-13	35.59301	51.01706	35.59301	51.01706	
5000	0.000226099	0.000145	1.555078	0.129467	-7E-05	0.000522	-7E-05	0.000522	

Figure 5.7: Regression statistics view

## 5.8 Data Calculation

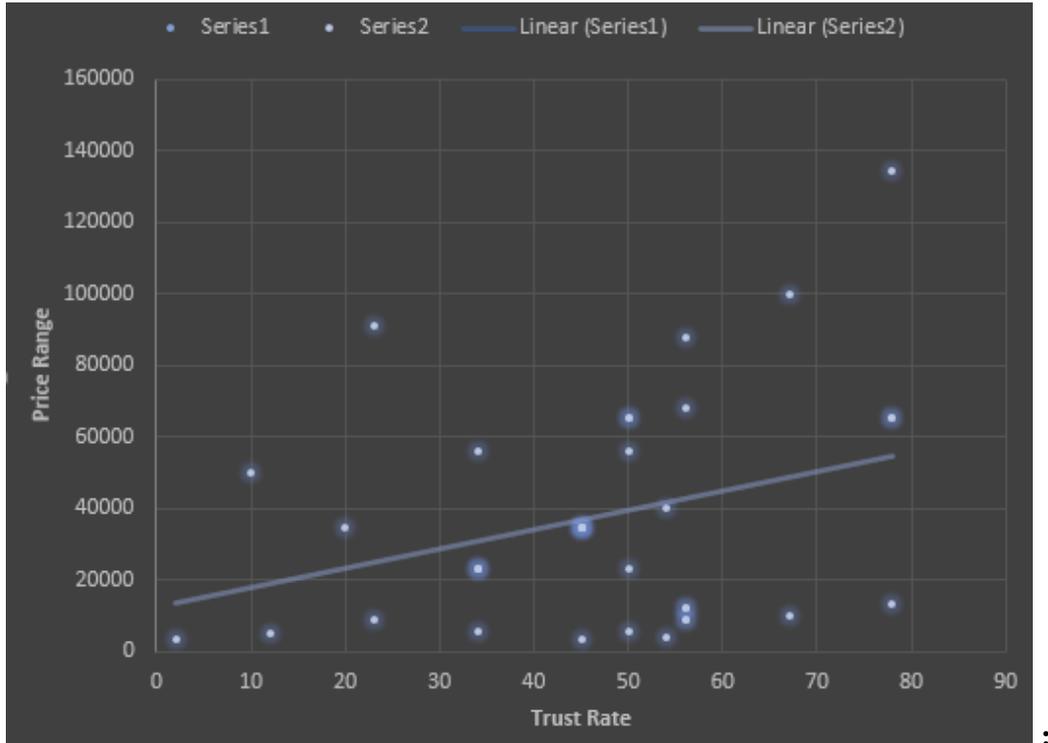


Figure 5.8: price range graph

## 5.9 Statical view

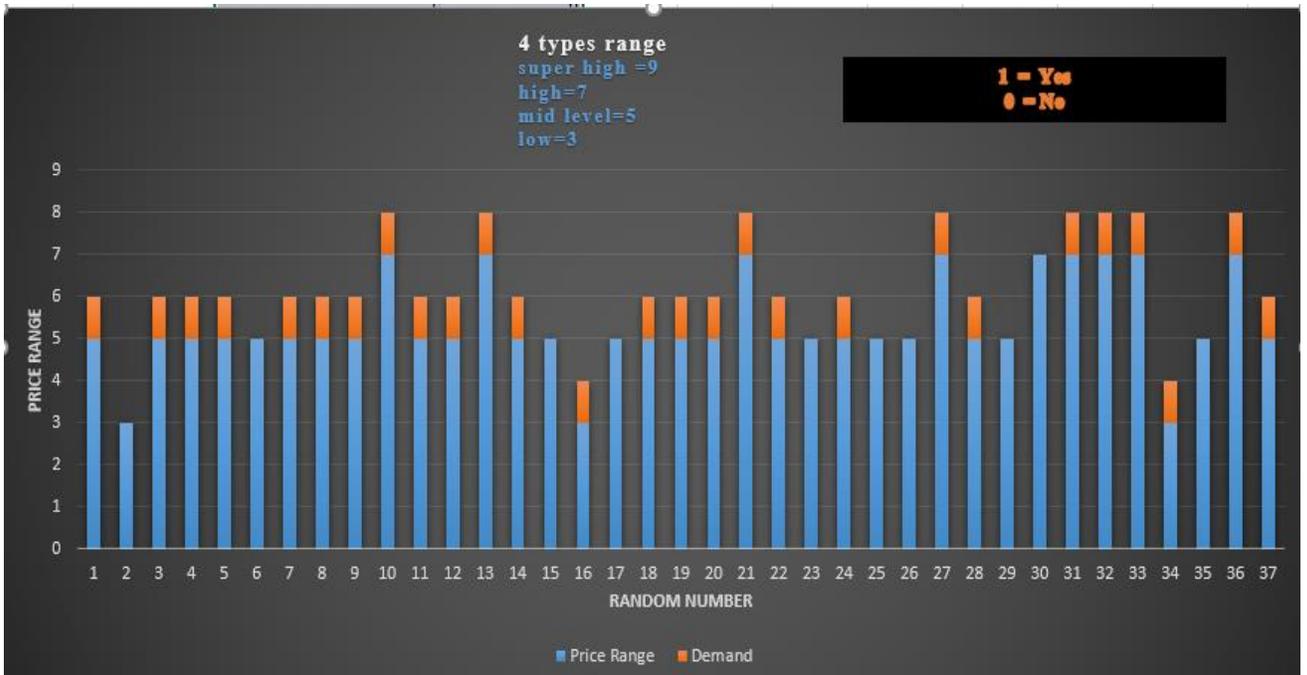


Figure 5.9: A Screenshot of the price range

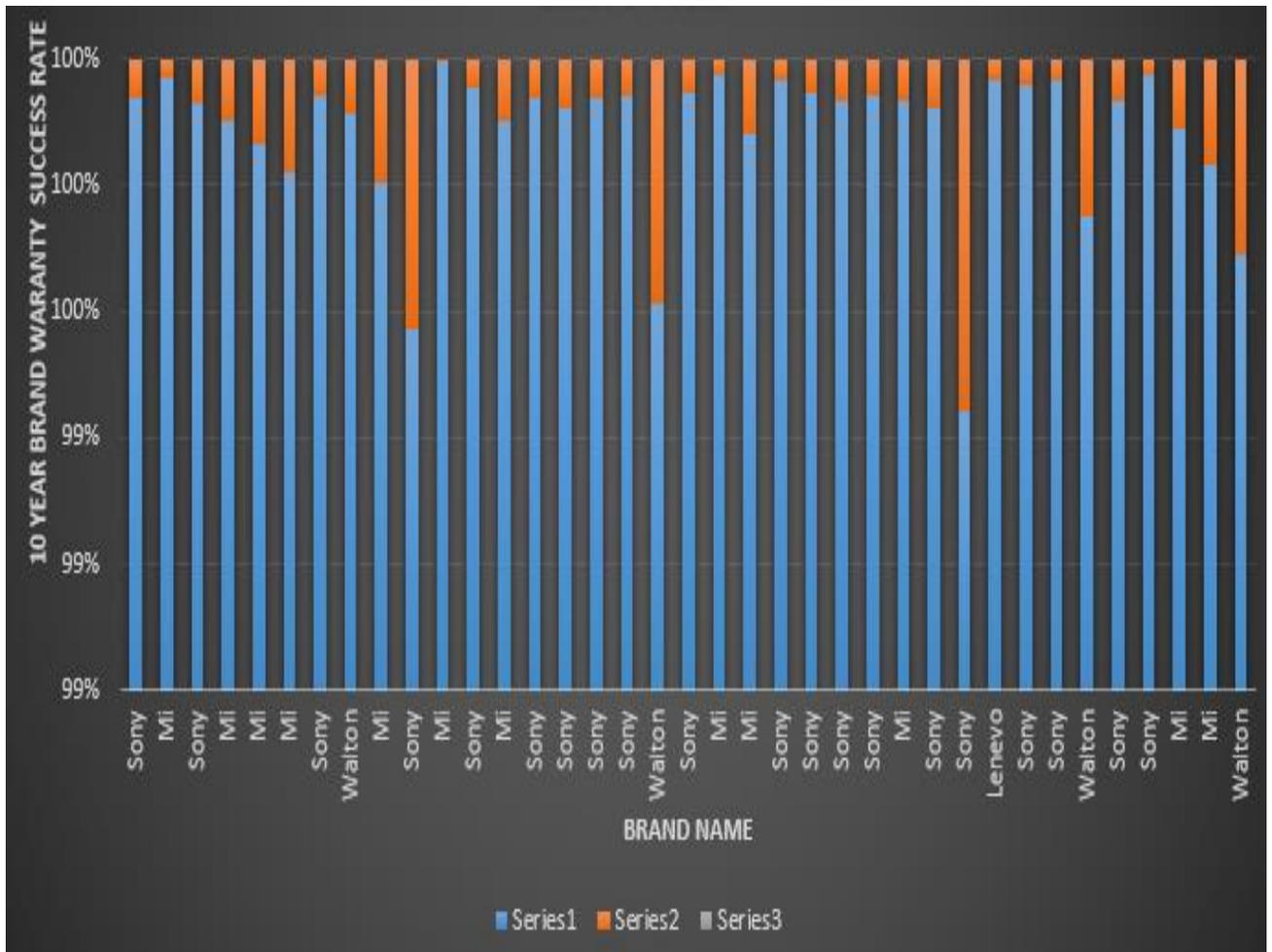


Figure 5.10: A Screenshot of the product percentage

### 5.10Result:

1. Trust rate vs Price Graph.

R value = 0.461299 that mean its moderate value in regression.

So, if the any brand price rate is high and this brand is trusted also in customer this product value rate increase .This types of product valuable more than the other product.

2. If any product use many years without any damage or problem , this product sell rate is too many high .
3. Mid level price rang product is more demandable. hear , R value is 0.654338

### **5.11 Discussion of result:**

Product value quality prediction data can be analysis products looks more standard a Brand new product is most popular. The following product is Television , Mobile , Projector . Many brand projector is market available just like Hitachi, Sony, Panasonic, View sonic, Optoma , Benq , Ricoh etc .Here is Many local projector is available here. But in this projector most popular is Brand Projector like Sony. Here is many Television and mobile is market available , But in time a new company goods on the market people trust level in Trust level(0-25%) is 29.7% , Trust level(26-50%) is 40.1% , Trust level(51-75%) is 19.2% , Trust level(76-100%) is 5.2% .Here is most acceptable product is Brand product .Brand product Television , Mobile, Projector compare in Sony ,Walton ,Mi , Lenevo . Sony product is most acceptable in people, Because of price, lasting of product ,brand product, Good looking. R value = 0.461299 that mean its moderate value in reggration .

So, if the any brand price rate is high and this brand is trusted also in customer this product value rate increase .This types of product valuable more than the other product .Data calculation view a new product can buy in customer in approximately mid level price . The final step is to evaluate the performance of the algorithm.As we have discussed that the linear regression model basically finds the best value for the intercept and slope, which results in a line that best fits the data. To see the value of the intercept and slope calculated by the linear regression algorithm for our dataset.

## **CHAPTER 6**

### **CONCLUSION AND FUTURE SCOPE**

#### **6.1 Conclusion**

We study the linear regression of machine learning algorithms. We implemented multiple linear regression using the Scikit-Learn machine learning library. This article focuses on data mining and machine learning, predicting product value scenarios, and analyzing the cost of product value. Provide visualization and accuracy of the result. Therefore, the right quality prediction plays a key role in providing quality products to further improve competitiveness. Regression is a recent technique that generalizes and combines the characteristics of principal component analysis and multiple regression. Its purpose is to predict or analyze a set of variables dependent on a set of independent variables or predictors. These features will benefit the sector, new businesses, new customers and new products if implemented. Our country is growing every day. People are discovering new technologies every day. Thanks to the government, there are seminars on new technologies. People will take part in these seminars, they will learn new things, these seminars will inspire them.

#### **Appendix**

Journal list  
ACM transaction on database system  
ACM transaction on information system  
IBM system journal  
IEEE computer  
Database  
Machine intelligence  
Machine learning  
International journal of intelligent system  
Information system research  
Management science  
Business intelligence  
Data mining

## References

- [1] Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3), 559–569. doi:10.1016/j.dss.2010.08.006
- [2] Moro, S., Rita, P., & Vala, B. (2016). Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. *Journal of Business Research*, 69(9), 3341–3351. doi:10.1016/j.jbusres.2016.02.010
- [3] Yu, P. S. (n.d.). Data mining and personalization technologies. Proceedings. 6th International Conference on Advanced Systems for Advanced Applications. doi:10.1109/dasfaa.1999.765731.
- [4] Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, 31(1), 127–137. doi:10.1016/s0167-9236(00)001.
- [5] Bose, I., & Mahapatra, R. K. (2001). Business data mining — a machine learning perspective. *Information & Management*, 39(3), 211–225. doi:10.1016/s0378-7206(01)00091-x.
- [6] Kim, D.-H., Kim, T. J. Y., Wang, X., Kim, M., Quan, Y.-J., Oh, J. W., ... Ahn, S.-H. (2018). Smart Machining Process Using Machine Learning: A Review and Perspective on Machining Industry. *International Journal of Precision Engineering and Manufacturing-Green Technology*, 5(4), 555–568. doi:10.1007/s40684-018-0057-y
- [7] Hyundai & WayRay Unveil Next-generation Visual Technology at CES 2019, available at <<<https://www.hyundai.news/eu/brand/hyundai-wayray-unveil-next-generation-visual-technology-at-ces-2019/>>> last accessed on 01-11-2019 at 11:30 AM.

- [8] Argon4 The Augmented Reality Web Browser, available at <<<https://app.argonjs.io/>>>, last accessed on 01-11-2019 at 11:40 AM.
- [9] AR GPS DRIVE/WALK NAVIGATION, available at <<<https://play.google.com/store/apps/details?id=com.w.argps&hl=en>>>, last accessed on 01-11-2019 at 11:55 AM.
- [10] Learning will be fun with Augmented Reality | Ayman Sadiq, available at <<<https://www.youtube.com/watch?v=rGZjAl91WzY>>>, last accessed on 01-11-2019 at 12:05 PM.
- [11] Business process modeling, available at <<[https://en.wikipedia.org/wiki/Business\\_process\\_modeling](https://en.wikipedia.org/wiki/Business_process_modeling)>>, last accessed on 01-11-2019 at 1:00 PM.
- [12] Sundsoy, P., Bjelland, J., Iqbal, A. M., Pentland, A. “Sandy”, & de Montjoye, Y.-A. (2014). Big Data-Driven Marketing: How Machine Learning Outperforms Marketers’ Gut-Feeling. Lecture Notes in Computer Science, 367–374. doi:10.1007/978-3-319-05579-4\_45
- [13] Functional Requirement, available at <<[https://en.wikipedia.org/wiki/Functional\\_requirement](https://en.wikipedia.org/wiki/Functional_requirement)>>, last accessed on 01-11-2019 at 1:25 PM.
- [14] Ruey-Shun Chen, Kun-Chieh Yeh, Chan-Chine Chang, & Chien, H. H. (n.d.). Using Data Mining Technology to improve Manufacturing Quality - A Case Study of LCD Driver IC Packaging Industry. Seventh ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD’06). doi:10.1109/snspd-sawn.2006.75
- [15] frontend design, available at <<<https://bradfrost.com/blog/post/frontend-design/>>>, last accessed on 01-11-2019 at 2:00 PM.
- [16] Unity (game engine), available at <<[https://en.wikipedia.org/wiki/Unity\\_\(game\\_engine\)](https://en.wikipedia.org/wiki/Unity_(game_engine))>>, last accessed on 01-11-2019 at 3:00 PM.
- [17] Reiss, R., Wojsznis, W., Wojewodka, R., 2010. Partial least squares confidence interval calculation for industrial end-of-batch quality prediction. Chemometrics and Intelligent Laboratory Systems, 100(2), pp.75-82.

- [18] Free3D, available at <<<https://free3d.com/>>>, last accessed on 01-11-2019 at 4:40 PM.
- [19] What is Interaction Design? available at <<<https://www.interaction-design.org/literature/article/what-is-interaction-design>>>, last accessed on 01-11-2019 at 4:45 PM.
- [20] Kumar, A. and Goyal, P. Forecasting of air quality in Delhi using principal component regression technique. Atmospheric Pollution Research,2011; 2(4) pp.436-444.

---

ORIGINALITY REPORT

---

**26%**  
SIMILARITY INDEX

**24%**  
INTERNET SOURCES

**10%**  
PUBLICATIONS

**%**  
STUDENT PAPERS

---

PRIMARY SOURCES

---

<b>1</b>	<a href="https://dspace.daffodilvarsity.edu.bd:8080">dspace.daffodilvarsity.edu.bd:8080</a> Internet Source	<b>8%</b>
<b>2</b>	<a href="http://www.healthknowledge.org.uk">www.healthknowledge.org.uk</a> Internet Source	<b>4%</b>
<b>3</b>	<a href="http://uta.influent.utsystem.edu">uta.influent.utsystem.edu</a> Internet Source	<b>3%</b>
<b>4</b>	<a href="http://it.mathworks.com">it.mathworks.com</a> Internet Source	<b>2%</b>
<b>5</b>	<a href="https://dspace.library.daffodilvarsity.edu.bd:8080">dspace.library.daffodilvarsity.edu.bd:8080</a> Internet Source	<b>2%</b>
<b>6</b>	<a href="https://docshare01.docshare.tips">docshare01.docshare.tips</a> Internet Source	<b>1%</b>

---

---

7

towardsdatascience.com

Internet Source

1%

---

8

sundsoy.com

Internet Source

1%

---

10

www.yohanli.com

Internet Source

1%

---

11

machinelearningmastery.com

Internet Source

1%

---

12

Dong-Hyeon Kim, Thomas J. Y. Kim, Xinlin

Wang, Mincheol Kim et al. "Smart Machining Process Using Machine Learning: A Review and Perspective on Machining Industry", International Journal of Precision Engineering and Manufacturing-Green Technology, 2018

Publication

1%

---

Ekaterina Kuleshova, Anastasia Levina, Rustam Esedulaev. "Reengineering of supply chain management integrated scheduling processes", MATEC Web of Conferences, 2018

Publication

<1%

---

Michael J Shaw, ChandrasekarSubramaniam, Gek Woo Tan, Michael E Welge. "Knowledge management and data mining for marketing", Decision Support Systems, 2001

Publication

<1%

---

---

15

[www.afbe.biz](http://www.afbe.biz)

Internet Source

<1%

---

17

[www.kdnuggets.com](http://www.kdnuggets.com)

Internet Source

<1 %

---

18

[repositorio.ufpe.br](http://repositorio.ufpe.br)

Internet Source

<1 %

---

19

[www.ijarcsse.com](http://www.ijarcsse.com)

Internet Source

<1 %

---

20

[docplayer.net](http://docplayer.net)

Internet Source

<1 %

---

21

M. Galster, A. Eberlein, M. Moussavi.

"Transition from Requirements to Architecture: A Review and Future Perspective", Seventh ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD'06), 2006

Publication

<1 %

---

22

[shura.shu.ac.uk](http://shura.shu.ac.uk)

Internet Source

<1 %

---

---

23 Indranil Bose, Radha K. Mahapatra. "Business data mining — a machine learning perspective", Information & Management, 2001

Publication

<1%

---

"Find out a new product value quality prediction in market using Data mining"

[www.repjusticeconference3.co.za](http://www.repjusticeconference3.co.za)

Internet Source

1%

[productdevelop.blogspot.com](http://productdevelop.blogspot.com)

16 Internet Source

<1%

