

**HYPOTHYROIDISM PREDICTOR: STUDY OF DETECTING HYPOTHYROIDISM
USING DATA MINING**

BY

RUHUL QUDDUS AL QUADERY

ID: 152-15-5912

SHAWON BALA NATH

ID: 151-15-5181

AFIFA SULTANA

ID: 152-15-5575

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

Supervised By

REZWANA SULTANA

Lecturer

Department of CSE

Daffodil International University

Co-Supervised By

RUBAIYA HAFIZ UPOMA

Lecturer

Department of CSE

Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY
SHUKRABAD, DHANMONDI, DHAKA-1207**

APRIL 2019

APPROVAL

This Project titled “HYPOTHYROIDISM PREDICTOR”, submitted by Shawon Bala Nath Ruhul Quddus Al Quadery and Affa Sultana to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on April 2019.

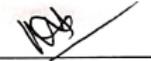
BOARD OF EXAMINERS



Dr. Syed Akhter Hossain
Professor and Head

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

Chairman



Dr. Md. Ismail Jabiullah
Professor

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

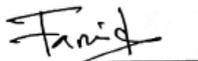
Internal Examiner



Dr. Sheak Rashed Haider Noori
Associate Professor & Associate Head

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

Internal Examiner



Dr. Dewan Md. Farid
Associate Professor

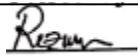
Department of Computer Science and Engineering
United International University

External Examiner

DECLARATION

I hereby declare that, this project has been done by us under the supervision of **REZWANA SULTANA, Lecturer, Department of CSE** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

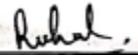


Rezwana Sultana
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Rubaiya Hafiz Upoma
Lecturer
Department of CSE
Daffodil International University

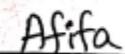
Submitted by:



Ruhul Quddus
ID: 152-15-5912
Department of CSE
Daffodil International University



Shawon Bala Nath
ID: 151-15-5181
Department of CSE
Daffodil International University



Afifa Sultana
ID: 152-15-5575
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **REZWANA SULTANA, Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of Data mining influenced me to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Dr. Syed Akhter Hossain, Professor and Head, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Thyroidism disease is extremely regular infection in human. These days most of the ladies experiencing thyroid infection than male. There are two sorts in thyroid ailment like hypothyroid and hyperthyroid illness. These sicknesses giving many symptoms, for example, weight gain, weight reduction, stress, etc. to our human body. If this ailment is recognized in before stage, at that point doctor can give legitimate treatment to the patients. Information Mining is assuming essential job in foreseeing numerous sicknesses. Grouping is one the most critical Technique in Data Mining. It is a directed learning. It is utilized to arrange predefined informational indexes. Medicinal services information is having exponential development in volume and multifaceted nature. Information mining Technique is for the most part utilized in social insurance associations for basic leadership, diagnosing illness and giving better treatment to the patients. By utilizing this system and to take care of this issue and make individuals comfort, we made this task. This venture introduces a site that named "**Hypothyroidism Predictor: Study of Detecting Hypothyroidism using Data Mining**" we can ready to distinguish the illness in all respects effectively. The dataset utilized for the investigation on hypothyroid is taken from UCI archive. Order of this thyroid ailment is a significant undertaking.

TABLE OF CONTENTS

CONTENTS	PAGE
Acknowledgements	ii
Abstract	iii
List of Figures	vi
List of Tables	vii
CHAPTER	PAGE
CHAPTER 1: INTRODUCTION	1-6
1.1 Introduction	1
1.2 Purpose	3
1.3 Proposed Methodology	4
1.4 Project Scope	5
1.5 Expected Outcome	6
CHAPTER 2: BACKGROUND	7-24
2.1 Introduction	7
2.2 Algorithms details	9
2.3 Evolution of Data Mining	13
2.4 Application	14
2.5 Data Mining Tools	16
CHAPTER 3: PROJECT REVIEW	25-29
3.1 Potential Challenges	25
CHAPTER 4: DESIGN SPECIFICATION	30-34
4.1 Project Perspective	30
4.2 Design and Implementation Constraints	33
CHAPTER 5: REQUIREMENT SPECIFICATION	35-36
5.1 Introduction	35

CHAPTER	PAGE
5.2 General System Requirements	35
5.3 Use Case Model	35
5.4 Use Case Description	36
CHAPTER 6: SYSTEM DESIGN IMPLEMENTATION	37
6.1 Data Flow Diagram	37
CHAPTER 7: CONCLUSION AND FUTURE SCOPE	38
7.1 Conclusion	38
7.2 Future scope	38
APPENDIX	39-40
Appendix A: Project Reflection	39
Appendix B: Related Diagram	40
REFERENCES	41

LIST OF FIGURES

PAGE

FIGURES

Figure 2.1.1: Data Mining Process	8
Figure 2.2.1: Data Mining Techniques	11
Figure 2.3.1: Evolution of Data Mining	14
Figure 2.5.1: Rapid Miner Tool	17
Figure 2.5.2: Weka	18
Figure 2.5.3: KNIME Tool	19
Figure 2.5.4: Orange	20
Figure 2.5.5: SQL Server Data Tool	21
Figure 2.5.6: Weka Classifier Output	23
Figure 2.5.7: Analyzing Instances	24
Figure 3.1.1: Prevalence of Hypothyroidism in Adults	26
Figure 3.1.2: Classification Tree Visualization	27
Figure 4.1.1: Supervised Learning	32
Figure 4.1.2: Unsupervised Learning	33
Figure 5.3.1: Use Case Model for Hypothyroidism Predictor	35
Figure 6.1.1: Data Flow Diagram	37

LIST OF TABLES

PAGE

TABLE

Table 2.5.1: Data Set	22
Table 4.1.1: Hypothyroidism Predictor Service	31
Table 5.4.1: Use Case Description of Input	36
Table 5.4.2: Use Case Description of Output	36

CHAPTER 1

INTRODUCTION

1.1 Introduction

Data mining is a procedure of dissecting and removing concealed data from huge informational indexes to discover a few examples. These examples are helpful in forecast technique. These information gives a premise to the investigation of hazard factors for some infection. There are different kinds of illnesses anticipated in information mining to be specific lung malignancy, liver issue, bosom disease, thyroid malady, diabetics and so on. Anticipating hypothyroidism disease is dissected in this undertaking.

Thyroid organ will stop thyroid hormones to keep up the body's metabolic rate. Thyroid issue is caused because of the glitch of thyroid hormones. Thyroid or thyroid organ discharges triiodothyronine (T3) and thyroxin (T4) into the circulation system as the imperative hormones.

There are two most common problems of the thyroid disorder:

- **Hyperthyroidism** (underactive thyroid or low thyroid) it discharges an excess of thyroid hormone into the blood due to over dynamic of thyroid. This can animate body's digestion fundamentally.
- **Hypothyroidism** (overactive thyroid) alludes to a condition in which the thyroid organ creates a lot of the hormone thyroxin. This miracle the ordinary equalization of compound responses in your body.

Worldwide too little iodine in the eating routine is the most reason for Hypothyroidism. The analysis of hypothyroidism can be affirmed with blood tests estimating thyroid-animating hormone (TSH) and thyroxin (FT4) levels. Hypothyroidism can affect the respiratory framework. Lung volumes are typically ordinary or somewhat decreased, however maximal breathing limit and diffusing limit are normally diminished.

Hyperthyroidism and Hypothyroidism are normal conditions that affect wellbeing. Hypothyroidism is a condition in which the body needs adequate thyroid hormone. Since the primary motivation behind thyroid hormone is to "run the body's digestion," it is justifiable that individuals with this condition will have indications related with a moderate digestion. Hypothyroidism is a condition in which your thyroid organ doesn't create enough of certain imperative hormones.

Around one billion individuals are evaluated to be iodine inadequate; be that as it may, it is obscure how regularly this outcome in hypothyroidism. In the US, hypothyroidism happens in 0.3– 0.4% of individuals. Subclinical hypothyroidism, a milder type of hypothyroidism described by typical thyroxine levels and a raised TSH level, is thought to happen in 4.3– 8.5% of individuals in the US Hypothyroidism is more typical in ladies than men. Individuals beyond 60 years old are more ordinarily influenced.

Arrangement is an information mining (AI) system used to foresee bunch participation for information occurrences [5, 6]. In this exploration LDA Calculation is utilized to predicate thyroid infection. An informational index downloaded from UCI store webpage is utilized for the exploratory reason, whole work is done with WEKA open source programming under Windows. K-overlap approval is additionally performed.

Anticipation in human services is a constant worry for the specialists and the right symptomatic at the correct time for a patient is critical, because of the suggested hazard. As of late, the standard therapeutic report can be joined by an extra report given by a choice emotionally supportive network or other propelled finding methods dependent on manifestations. Questions, for example, "what are the most critical components that influence thyroid?", "which is the class of the populace inclined to goiter malady?", "what is the most satisfactory treatment for a specific sickness?" and so forth may discover replies in applying information mining methods. Medicinal services information can be handled and after thorough utilization can give learning utilized in basic leadership, diagnosing illnesses all the more quickly and precisely, offering better drug for patients and limiting the demise hazard. The creators center their work around utilizing arrangement techniques and recognizing the best calculation for grouping thyroid issue.

1.2 Purpose

Data mining likewise called Knowledge Discovery in Databases (KDD) is the field of finding helpful data from a lot of information. It is an interdisciplinary subfield of software engineering and measurements with a general objective to extricate data from an informational collection and change the data into an intelligible structure for further use.

The distinction between information examination and information mining is that information investigation is to condense the history, for example, breaking down the viability of a showcasing effort, conversely, information mining centers around utilizing explicit AI and factual models to foresee the future and find the examples among information. Before information mining calculations can be utilized, an objective informational index must be gathered.

As data mining can just reveal designs really present in the information, the objective informational index must be sufficiently huge to contain these examples while staying compact enough to be mined inside a satisfactory time limit. A typical hotspot for information is an information shop or information distribution center. Pre-processing is basic to examine the multivariate informational indexes before information mining. The objective set is then cleaned. Information cleaning evacuates the perceptions containing commotion and those with missing information.

The finding of hypothyroidism, when suspected, can be affirmed with blood tests estimating thyroid-animating hormone (TSH) and thyroxine levels. In the US, hypothyroidism happens in 0.3– 0.4% of individuals. Subclinical hypothyroidism, a milder type of hypothyroidism described by typical thyroxine levels and a raised TSH level, is thought to happen in 4.3– 8.5% of individuals in the US. Hypothyroidism is more typical in ladies than men. The primary preferred standpoint of data innovation is that a tremendous information stockpiling of past patient's records is kept up and checked by emergency clinics consistently for different references. This therapeutic information encourages the specialists to look at changed examples in the informational index. The examples found in informational indexes might be utilized for order, forecast and finding of the sicknesses.

1.3 Proposed Methodology

Methodology ought to contain adequate detail for peruses to reproduce the work done and get comparative outcomes. This gives a depiction of the definite (explicit) activities, plan to be utilized to respond to the exploration questions.

Dataset and Sample

- The dataset of the project is formed by 900 data collected from the National Ambulatory Medical Care Survey 2010.
- We organized the dataset in a fashion that it can easily fit into the project database.

Techniques and Rules

- In this project we have use Data Mining technique for extracting hidden patterns and relations among datasets that are stored in a database that was previously unknown for the world.
- We have used an application of association mining in the temporal databases is that of prediction.
- A prediction rule must have relatively stable confidence with respect to the time frame determined by application domain. We develop a prediction mining algorithm for discovering the set of prediction rules.
- We have split the dataset into two parts. Two third of the populations for Training Dataset and remaining populations for the Testing Dataset.

Applications and Tools

Application is the basic part for executing the attempted strategies and procedures in a systematic approach to achieve the exploration works. We have utilized the most recognizable stage or application for Information Mining, which is called as Weka (Waikato Condition for Learning Investigation). An information mining programming that was created in the College of Waikato by utilizing Java programming language. Weka gives the offices to utilize every one of the tenets and calculations of information mining including: grouping, bunching and partner.

Statistical Processing of Outcomes

- Effectiveness percent analysis of the algorithms was overtaken for each of the four algorithms. We have gain a high presentence of outcome for each of them.
- For the J48 pruned tree algorithm we have got a result of **70%** (correctly classified instances), which we think more than we expected.
- For the Naïve Bayes algorithm, we have got a result of **66.67%** (correctly classified instances), which is little bit worse than J48 pruned tree algorithm.
- For the EM algorithm we have got a result of **57.29%** (correctly classified instances), which is too much worse than two Classification algorithms.
- For the Simple K-Means algorithm we have got a result of **56.069%** (correctly classified instances), which is little bit better than EM algorithm but still too much worse than two Classification algorithms.
- For Training set we have observed an outcome of **77%** (correctly classified instances) we observed an outcome of **67%** and **79%** (correctly classified instances) sequentially.

1.4 Project Scope

Hypothyroidism is a typical endocrine issue coming about because of lack of thyroid hormone. In our nation individuals don't know about this reality. These days the site become increasingly commonplace to us. By utilizing different informational indexes and the expectation tenets can be performed to recognize the malady whenever.

1.5 Expected outcome

The result of this task will foresee Hypothyroidism utilizing just TSH and FT4 and up to 70% occurrences our indicator will give genuine outcomes.

As we utilized WEKA as our information mentor, we are for the most part contingent upon its outcome. So, the rate it gives for every calculation, we have just chosen the calculation which gave us most ideal outcome.

The expected outcome is as following:

- To have number of models developed using the datasets.
- To having models that can predict successfully any test sets that relevant to our project dataset.
- To find out the correlation between the variables correctly.
- To recognize which variables influence Hypothyroidism the most.
- Able to understand the portion of males and females are most likely to affect by Hypothyroidism.
- Find out variables that can be controlled. If so, then it leads to a situation where we can minimize the threat of Hypothyroidism.

CHAPTER 2

BACKGROUND

2.1 Introduction

Information mining alludes to separating obscure examples from a tremendous volume of information including diverse techniques and calculations which exist at the crossing point of fields, for example, computerized reasoning, AI, measurements and database frameworks. Emergency clinics, facilities and restorative examination research centers collect a lot of patient information throughout the years. This information give a premise to the investigation of hazard factors for some ailments (different sorts of malignant growth, heart infections, diabetes, hypothyroidism, hepatitis and so forth.).

A large portion of points of reference insinuate diagnosing contaminations of thyroid using decision trees, fake neural frameworks, support vector machine, ace systems, etc. For example, the assurance of thyroid issue using ANN's.

Information mining can plan the patient profile inclined to build up a thyroid capacity and to distinguish the hazard factors for these classes of infections. On the first we present the analyses created on a database given by UCI AI Vault (UCI, 2016), containing information about clinical history of patients with thyroid issue. On the second subsection we utilized information about Romanian patients gathered from a site (tiroida.ro, 2016). The creators considered for grouping of information four information mining calculations: Guileless Bayes, Choice Trees, Multilayer Perceptron and RBF System to assemble a hearty classifier. The objective of this task is to locate the best characterization display to make future order of new patient information even more precisely.

Information mining (DM), the extraction of concealed prescient data from extensive databases, is a ground-breaking innovation with incredible potential to concentrate on the most critical data in the information distribution centers. Information mining instruments foresee future patterns and practices, permit making proactive, learning driven choices.

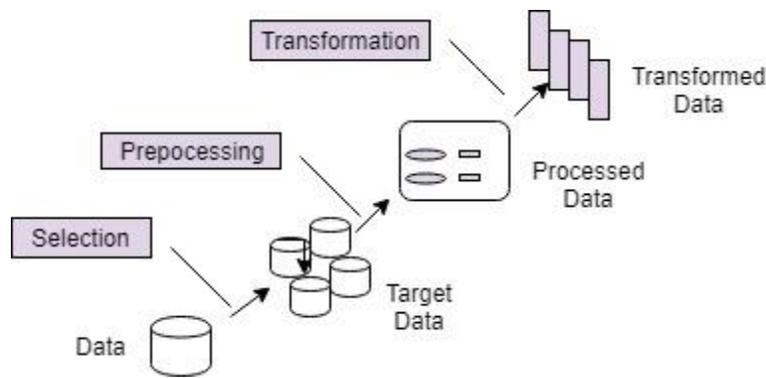


Figure 2.1.1: Data Mining Process

The analysis of data, on the data mining can be occurred in two ways:

1. Descriptive:

Elucidating investigation or insights does precisely what the name infers they "Portray" or outline crude information and make it something that is interpretable by people. They are investigation that portray the past. The past alludes to any purpose of time that an occasion has happened, regardless of whether it is one moment prior, or one year back. Spellbinding investigation are helpful on the grounds that they enable us to gain from past practices and see how they may impact future results. Three basic cornerstones of descriptive analytics are:

- Clustering
- Association rules discovery
- Sequential rules discovery

2. Predictive:

Prescient investigation has its underlying foundations in the capacity to "Anticipate" what may occur. That is, discover estimation of any trait, utilizing the estimation of other characteristic. These examinations are tied in with understanding what's to come. Prescient investigation furnishes organizations with significant bits of knowledge dependent on information. Prescient examination give gauges about the probability of a future result. Remember that no measurable calculation can "foresee" the future with 100% sureness. Organizations utilize these insights to conjecture what may occur later on. This is on the grounds that the establishment of prescient investigation depends on probabilities. Three basic cornerstones of predictive analytics are:

- Classification
- Regression
- Deviation detection.

2.2 Algorithm Details

Information mining is known as an interdisciplinary subfield of software engineering and essentially is a registering procedure of finding designs in huge informational indexes. It is considered as a fundamental procedure where shrewd techniques are connected so as to extricate information designs. To make a model, the calculation initially breaks down the information we give, searching for explicit sorts of examples or patterns. The calculation utilizes the aftereffects of this examination to characterize the ideal parameters for making the mining model. These parameters are then connected over the whole informational collection to separate noteworthy examples and nitty gritty measurements.

Given below is a list Data Mining Algorithms:

1. C4.5:

C4.5 [7] is an algorithm that is utilized to produce a classified as a choice tree and has been created by **Ross Quinlan** [7]. And so as to do likewise, C4.5 is given a lot of information that speak to things that have just been classified.

C4.5 that is frequently alluded to as a Statistical Classifier is fundamentally an expansion of Quinlan's ID3 calculation. The choice trees that are created by C4.5 can be additionally utilized for grouping. The C4.5 algorithm has additionally been portrayed as "a landmark decision tree program that is probably the machine learning workhorse most widely used in practice to date" by the authors of the Weka machine learning software.

2. k-means:

k-means [7] Clustering that is otherwise called Nearest Centroid Classifier or The Rocchio Algorithm is a strategy for vector quantization, that is extensively well known for group examination in information mining.

k-implies is utilized to make k bunches from a lot of items just so the individuals from a gathering are progressively comparable. It's a notable prominent group examination procedure utilized for investigating a dataset.

3. Support vector machines:

With regards to AI, support vector machines that are otherwise called help vector systems are fundamentally directed learning models that accompany related learning calculations which at that point break down information that are utilized for the investigation of relapse and grouping.

A SVM [7] model is made that is a portrayal of the models as focuses in space, that are additionally mapped with the goal that the instances of the different classes are then separated by an unmistakable hole that is should be as wide as could be allowed.

4. Apriori:

Apriori is an algorithm that is utilized for incessant itemset mining and affiliation rule learning value-based databases. The calculation is continued by the distinguishing proof of the individual things that are visit in the database and after that extending them to bigger datasets if adequately those thing sets show up frequently enough in the database. These regular datasets that are dictated by Apriori can be utilized for the assurance of affiliation rules which at that point feature general patterns.

5. EM(Expectation-Maximization):

An expectation– expansion (EM) calculation [7], with regards to insights is an iterative strategy that is utilized to discover most extreme a posteriori (MAP) or greatest probability appraisals of parameters in measurable models, that fundamentally relies upon in secret inactive factors.

6. PageRank (PR):

PageRank (PR) that was named after Larry Page who is one of the organizers of Google is a calculation that is utilized by Google Search to rank the sites in their web crawler results. PageRank that is the main calculation that was utilized by the organization isn't the main calculation that is being utilized by Google to arrange web index results, yet it is the best-known method for estimating the significance of site pages.

7. AdaBoost:

Versatile Boosting or AdaBoost, that has been defined by Yoav Freund and Robert Schapire is an AI meta-calculation, that won the originators the 2003 Godel Prize for the equivalent. The calculation can be utilized in arrangement with numerous different sorts of learning calculations so as to improve execution. AdaBoost is touchy to uproarious information just as exceptions.

8. k-NN:

The k-nearest neighbors calculation (k-NN) is a sort of languid learning or occasion based learning and is considered as a non-parametric technique that is utilized for characterization and relapse. In both the referenced cases, the information comprises of the k nearest preparing precedents in the component space and the yield relies upon whether the calculation is being utilized for characterization or relapse. This k-NN Algorithm is considered and is additionally among the least complex of all AI calculations.

9. Naive Bayes:

With regards to AI, Naive Bayes classifiers that are viewed as very versatile are known to be a group of basic probabilistic classifiers that depend on the utilization of Bayes' hypothesis with the assistance of solid free suppositions between the highlights.

10. CART:

CART is an algorithm that essentially represents characterization and relapse trees. It is a decision tree learning procedure that either yields arrangement or relapse trees and likewise like C4.5, CART is additionally a classifier.

A considerable lot of the reasons that a client would utilize C4.5 for additionally apply to that of CART, since them two are choice tree learning procedures and highlights like simplicity of elucidation and clarification are connected to CART too.



Figure 2.2.1: Data Mining Techniques.

We can utilize the Decision Trees calculation for forecast, yet additionally as an approach to decrease the quantity of segments in a dataset, on the grounds that the choice tree can recognize segments that don't influence the last mining model.

Investigation Services incorporates the accompanying calculation types:

Classification Algorithms

Classification algorithm is a data mining function that doles out things in an accumulation to target classifications or classes. The objective of order is to precisely anticipate at least one discrete factor, in light of different properties in the dataset.

A characterization task starts with an informational collection in which the class assignments are known. Orders are discrete and don't suggest request. Nonstop, gliding point esteems would show a numerical, as opposed to an unmitigated, target. A prescient model with a numerical target utilizes a relapse calculation, not a grouping calculation.

Typical algorithms of classifications are:

- Linear Classifiers.
- Support Vector Machines.
- Quadratic Classifiers.
- Decision Trees.
- Neural Networks.

Clustering Algorithms

Clustering Algorithm is a division calculation given by Analysis Services. The calculation utilizes iterative methods to gather cases in a dataset into bunches that contain comparative qualities. These groupings are valuable for investigating information, recognizing oddities in the information, and making expectations. Grouping models distinguish connections in a dataset that you may not sensibly infer through easygoing perception. For instance, you can sensibly perceive that individuals who drive to their employments by bike don't regularly carry on a long separation from where they work. Run of the mill Algorithms of Clustering are:

- K-Mean Method.
- Transactional Data.

Association Algorithms

Association rules are the same as grouping decides aside from that they can foresee any trait, not simply the class, and this gives them the opportunity to anticipate mixes of characteristics as well. Additionally, affiliation rules are not planned to be utilized together as a set, as characterization rules seem to be. Distinctive affiliation decides express unique regularities that underlie the dataset, and they for the most part anticipate diverse things.

Regression Algorithms

Regression algorithm predicts at least one consistent factor, for example, benefit or misfortune, in light of different characteristics in the dataset. It is a variety of the Decision Trees calculation that encourages you ascertain a direct connection between a reliant and autonomous variable, and afterward utilize that relationship for expectation.

The relationship appears as a condition for a line that best speaks to a progression of information. Every data point in the graph has a mistake related with its separation from the relapse line. There are different sorts of relapse that utilization various factors, and furthermore nonlinear techniques for relapse. Be that as it may, straight relapse is a helpful and understood strategy for displaying a reaction to an adjustment in some fundamental factor.

2.3 Evolution of Data Mining

DM is the aftereffect of a long transformative procedure of innovative work. DM is an interdisciplinary subject including information the executive's innovation as well as measurements and computerized reasoning. DM moves past review information access and route to forthcoming and proactive data conveyance. Figure 2.2 demonstrating the development of Data Mining.

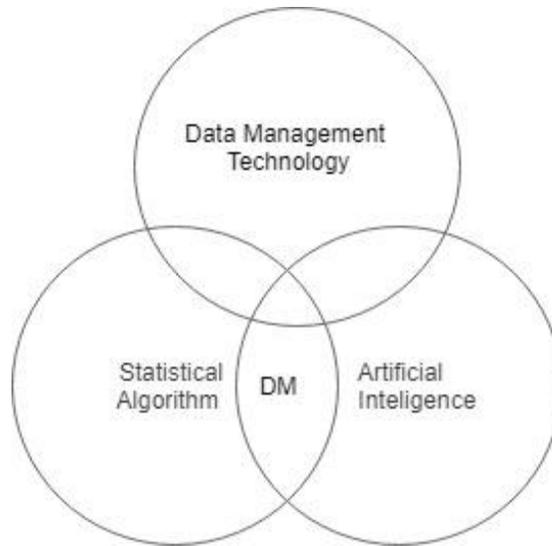


Figure 2.3.1: Evolution of Data Mining.

2.4 Application

Data mining is broadly utilized in differing regions. There are various business information mining frameworks accessible today but then there are numerous difficulties in this field. In this instructional exercise, we will talk about the applications and the pattern of information mining. Here is the rundown of regions where information mining is broadly utilized –

Financial Data Analysis

The money related information in banking and budgetary industry is commonly dependable and of great which encourages deliberate information investigation and information mining. A portion of the common cases are as per the following –

- Design and development of information distribution centers for multidimensional information examination.
- what's more, information mining.
- Loan installment forecast and client credit arrangement examination.
- Classification and bunching of clients for focused showcasing.
- Detection of illegal tax avoidance and other money related wrongdoings.

Retail Industry

Information Mining has its extraordinary application in Retail Industry since it gathers vast measure of information from on deals, client acquiring history, merchandise transportation, utilization and administrations. It is regular that the amount of information gathered will keep on extending quickly due to the expanding simplicity, accessibility and fame of the web.

- Design and Development of information stockrooms dependent on the advantages of information mining.
- Multidimensional examination of offers, clients, items, time and district.
- Analysis of adequacy of offers battles.
- Customer Maintenance.
- Product suggestion and cross-referencing of things.

Telecommunication Industry

Here is the rundown of precedents for which information mining improves media transmission administrations –

- Multidimensional Examination of Media transmission information.
- Fraudulent design examination.
- Identification of uncommon examples.
- Multidimensional affiliation and successive examples examination.
- Mobile Media transmission administrations.
- Use of perception devices in media transmission information investigation.

Biomedical Engineering

Information mining can help pinpoint shrouded data in medicinal information and precisely separate obsessive from typical information. It can separate concealed highlights from patient gatherings and sickness states and can help in computerized basic leadership.

Other Scientific Applications

The applications talked about above will in general handle moderately little and homogeneous informational collections for which the factual procedures are proper. Colossal measure of information has been gathered from logical areas, for example, geosciences, space science, and so on.

- Data Stockrooms and information preprocessing.
- Graph-based mining.
- Visualization and space explicit information.

2.5 Data Mining Tools

Alongside the progress to an application-based world comes the exponential development of information. Notwithstanding, the greater part of the information is unstructured and thus it takes a procedure and technique to extricate helpful data from the information and change it into reasonable and usable structure. This is the place information mining comes into picture. A lot of apparatuses are accessible for information mining assignments utilizing computerized reasoning, AI and different strategies to remove information.

Here are some incredible open source Data Mining Instruments accessible:

RapidMiner

This is extremely well known since it is a readymade, open source, no-coding required programming, which gives progressed investigation. Written in Java, it fuses multifaceted information mining capacities, for example, information pre-handling, perception, prescient investigation, and can be effectively incorporated with WEKA and R-instrument to legitimately give models from contents written in the previous two. Other than the standard information mining highlights like information purging, separating, grouping, and so forth., the product additionally includes worked in formats, repeatable work processes, an expert perception condition, and consistent incorporation with dialects like Python and R into work processes that guide in quick prototyping.

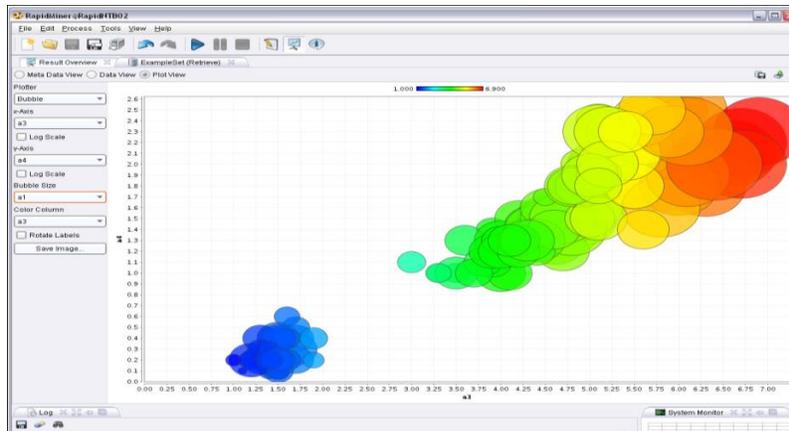


Figure 2.5.1: Rapid Miner Tool.

As indicated by Bloor Exploration, RapidMiner gives 99% of a progressed logical arrangement through layout-based structures that speed conveyance and decrease mistakes by almost wiping out the need to compose code. RapidMiner gives information mining and AI methods including information stacking and change information preprocessing and representation, prescient examination and measurable displaying, assessment, and sending. RapidMiner is written in the Java programming language. RapidMiner gives a GUI to structure and execute diagnostic work processes. Those work processes are classified "Procedures" in RapidMiner and they comprise of different "Operators".

Weka

Weka is an accumulation of AI calculations for information mining undertakings. The calculations can either be connected straightforwardly to a dataset or called from your own Java code. Weka contains instruments for information pre-preparing, characterization, relapse, grouping, affiliation guidelines and perception. It is likewise appropriate for growing new AI plans. Weka underpins a few standard information mining perceptions, relapse and highlight areas. It isn't equipped for multi-social information mining, yet there is discrete programming for changing over a gathering of connected database tables into a solitary table that is appropriate for handling utilizing Weka.

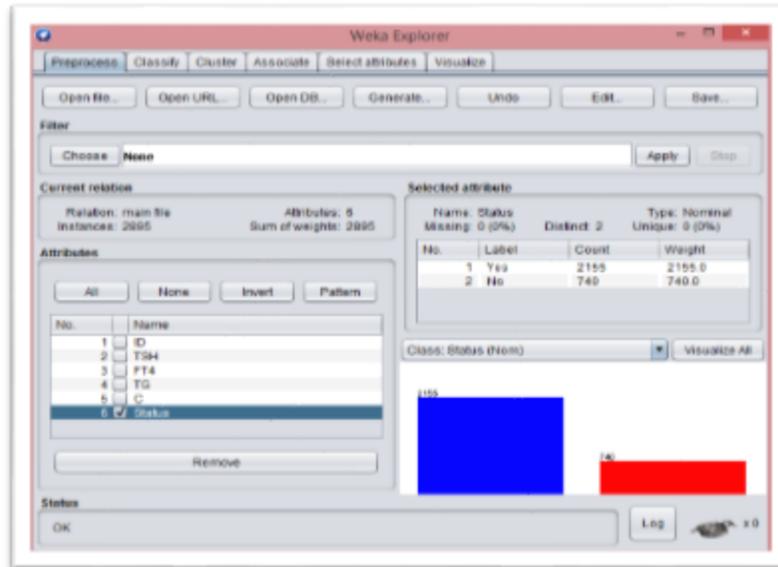


Figure 2.5.2: Weka.

Weka underpins a few standard information mining undertakings, all the more explicitly, information preprocessing, bunching, characterization, relapse, perception, and highlight choice. Most Weka's procedures are predicated on the supposition that the information is accessible as one level document or connection, where every datum point is portrayed by a fixed number of traits. Weka gives access to SQL databases utilizing Java Database Network and can process the outcome returned by a database inquiry.

Weka gives access to profound learning. It isn't equipped for multi-social information mining, however there is discrete programming for changing over a gathering of connected database tables into a solitary table that is appropriate for handling utilizing Weka. Another vital region that is at present not secured by the calculations incorporated into the Weka appropriation is arrangement displaying.

KNIME

KNIME is an easy to use graphical workbench for the whole examination process: information gets to, information change, introductory examination, incredible prescient investigation and announcing. The open incorporation stage gives more than 1000 modules (hubs). Figure 2.5.3 appearing working board of KNIME.

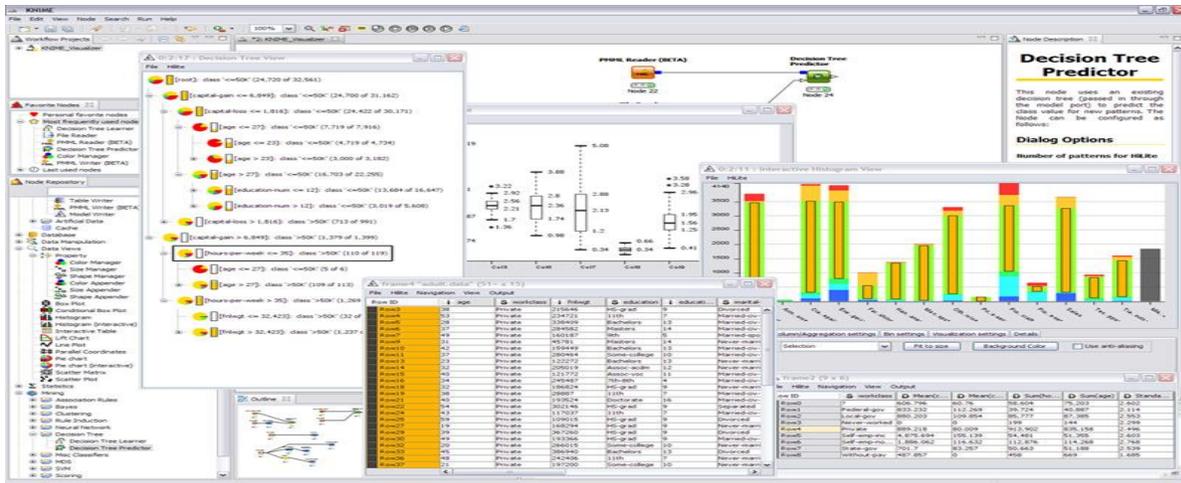


Figure 2.5.3: Overview of KNIME Tool

Orange

Orange is an ideal programming suite for AI and information mining. It best guides the information representation and is a part-based programming. It has been written in Python processing language.

As it is a part-based programming, the segments of orange are called 'gadgets'. These gadgets go from information perception and pre-handling to an assessment of calculations and prescient demonstrating.

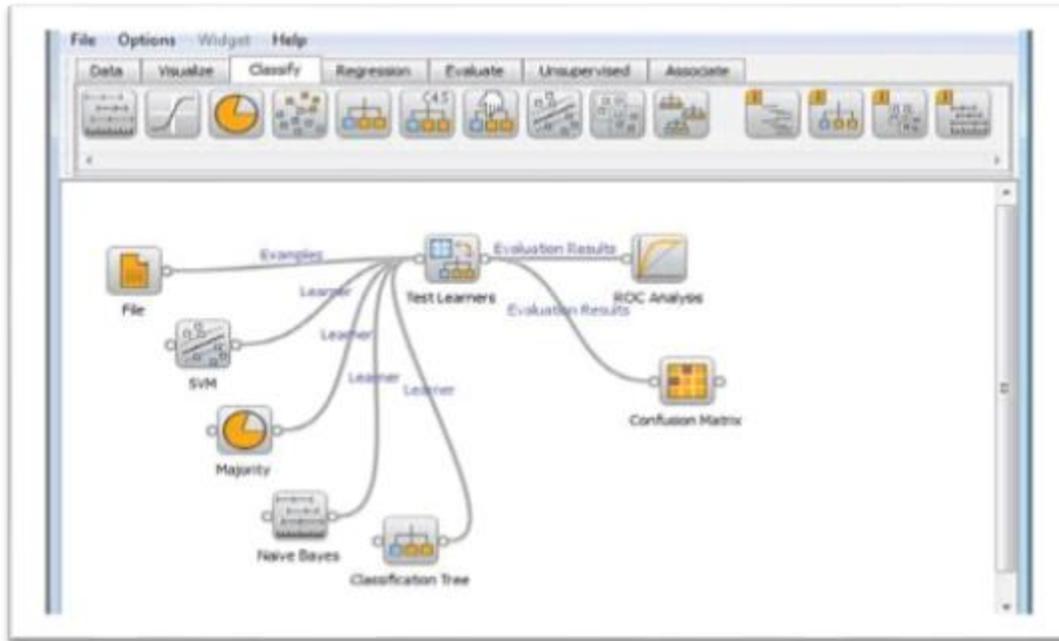


Figure 2.5.4: Orange.

Gadgets offer real functionalities like:

- Showing information table and permitting to choose highlights
- Reading the information
- Training indicators and to think about learning calculations
- Visualizing information components and so forth.

SSDT

SSDT is a widespread, decisive model that extends every one of the periods of database advancement in the Visual Studio IDE. Offers was the previous condition created by Microsoft to do information examination and give business knowledge arrangements. Engineers use SSDT execute a plan capacity of SQL, to manufacture, keep up, investigate and refactor databases. A client can work straightforwardly with a database or can work legitimately with an associated database, in this way, giving on or off-premise office.

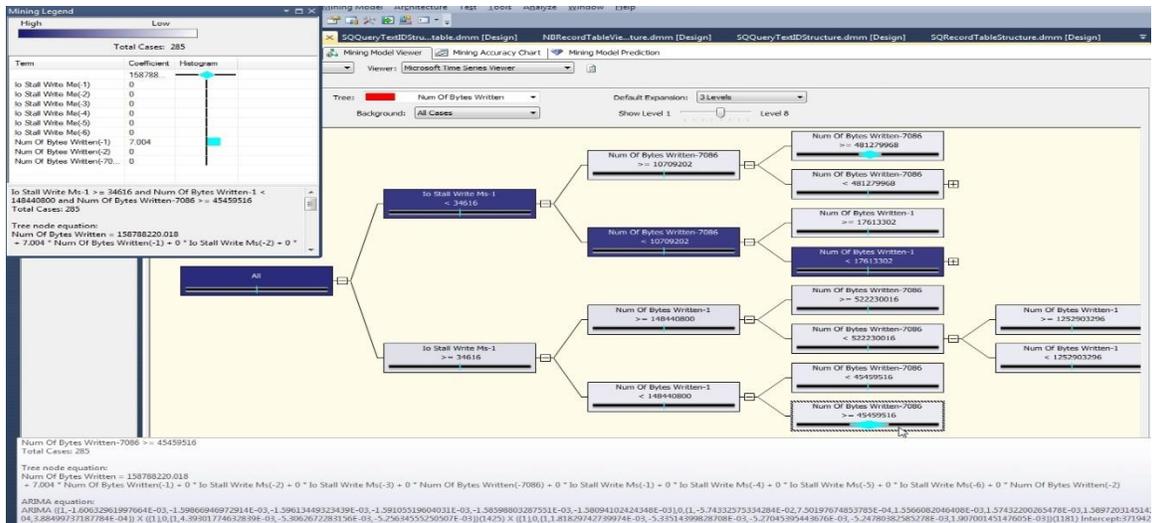


Figure 2.5.5: SQL Server Data Tools.

Clients can utilize visual studio devices for advancement of databases like IntelliSense, code route devices, and programming support by means of C#, visual fundamental and so on. SSDT gives Table Planner to make new tables just as alter tables in direct databases just as associated databases.

Comparative Study of DM Tools:

Information mining process includes a few different undertakings too separated from simply the extraction of data and examination of information. These distinctive errands fluctuate from information the executive's perspectives, and information pre-preparing to producing new guidelines and estimating the intriguing quality, relativity, unpredictability contemplations just as flag and picture handling, perception and internet refreshing of the database (Web Information Mining). A few of the functionalities given by these instruments incorporate portrayal and grouping of information, designs assessment, affiliations and relationships, and expectation over the information. Notwithstanding, they don't encourage for continuous.

Experimental Analysis

The execution of these instruments has been dissected by first running them with a few informational indexes accessible on UCI storehouse. A few distinct calculations for grouping and bunching were actualized in this investigation and execution of these calculations was watched.

In this area an example of the trial performed amid the exploration is exhibited and finish of the aftereffects of various apparatuses is examined.

Data Set:

The informational collection has 2896 cases from which 741 has a place with classification negative, 2155 has a place with classification remunerated hypothyroid. The last characteristic is the status, subsequently there are 2 includes altogether, which will be utilized to group the information.

TABLE 2.5.1: DATA SET.

Data Description	Attribute Name	Attribute Name	Attribute Name
Serial Number	TSH	FT4	Status
1.	(-1 – 400)	(-1 – 0.69)	Yes/No

(1) Audiology is utilized with information type multivariate, property type all out, number of characteristics 69, number of examples number of properties, number of occasions 226.

(2) Zoo: information type: multivariate, characteristic sort: straight out, whole number, number of qualities: 18, number of occurrences: 101.

Fundamentals: The arrangement was done of informational collection with rate split philosophy of 60% for preparing information and staying 40% for the test information. The acquired proportion of precision is utilized as the basis for the execution examination of the instruments.

Trial: The k nearest neighbor algorithm did not keep running on the informational index, the conceivable reason being the inconsistent information type as the information was of discrete qualities.

```

Classifier output
ZeroR predicts class value: Yes

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      2155           74.4387 %
Incorrectly Classified Instances    740           25.5613 %
Kappa statistic                    0
Mean absolute error                 0.3806
Root mean squared error             0.4362
Relative absolute error             100 %
Root relative squared error         100 %
Total Number of Instances          2895

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                1.000   1.000   0.744     1.000   0.853     ?       0.499    0.744    Yes
                0.000   0.000   ?         0.000   ?         ?       0.499    0.255    No
Weighted Avg.   0.744   0.744   ?         0.744   ?         ?       0.499    0.619

=== Confusion Matrix ===

  a  b  <-- classified as
2155 0 | a = Yes
 740 0 | b = No

```

Figure 2.5.6: Weka Classifier Output.

Database system support:

Weka supports perusing of documents from a few distinct databases. It additionally gives highlight to import the information over web, from site pages or from a remotely found SQL database server by entering only the URL of asset. This subsequently enables Weka to help wide range of information designs.

Graphical representation:

It enables clients to utilize R application as can legitimately associate with R bundle. Weka likewise furnishes with isolated GUI for learning stream, for Experimenter, used to contrast different outcomes and traveler with examine diverse informational collections. It additionally accommodates making own channels for sifting through occurrences.

Analysis and Processing Capabilities:

Be that as it may, Weka isn't better appropriate choice for the extensive informational indexes as they are generally dealt with. The databases with expansive unstructured information are not appropriate as it impedes the pre-preparing and processing time of Weka. It can perform well and furnish progressively precise outcomes with littler databases.

Correctly Classified Instances	1821	62.945 %
Incorrectly Classified Instances	1072	37.055 %
Kappa statistic	0.0206	
Mean absolute error	0.3705	
Root mean squared error	0.6087	
Relative absolute error	97.3045 %	
Root relative squared error	139.5191 %	
Total Number of Instances	2893	

Figure 2.5.7: Analyzing Instances.

Issues:

Tree based learning algorithms are viewed as a standout amongst the best and generally utilized administered learning techniques. Tree based techniques enable prescient models with high exactness, soundness and simplicity of translation. In contrast to direct models, they map non-straight connections great.

CHAPTER 3

PROJECT REVIEW

3.1 Potential Challenges

Predictive Analytics incorporates an assortment of factual systems from information mining, It abuses designs found in chronicled and value-based information to distinguish dangers and openings. The critical errand of this calculation is Supervised realizing where the information is utilized legitimately (no unequivocal model is made) to anticipate the class estimation of another case.

Predictive analytical process:

1.**Define task:** Define the venture results, deliverable, extent of the exertion, business targets, distinguish the informational indexes that will be utilized.

2.**Data gathering:** Data digging for prescient investigation plans information from different hotspots for examination. This gives a total perspective on client associations.

3.**Data examination:** Data Analysis is the way toward reviewing, cleaning and displaying information with the goal of finding valuable data, touching base at end

4.**Statistics:** Statistical Analysis empowers to approve the presumptions, theory and test utilizing standard factual models.

5.**Modelling:** Predictive demonstrating gives the capacity to naturally make precise prescient models about future. There are likewise choices to pick the best arrangement with multi-modular assessment.

6.**Deployment:** Predictive model arrangement gives the alternative to send the expository outcomes into ordinary basic leadership procedure to get results, reports and yield via robotizing the choices dependent on the displaying.

Decision Tree:

Decision Trees are broadly utilized strategy in Data mining. These innovations pursue Divide and Conquer arrangement to separate the space for choice regions. The principal hub structured is known as the root hub and the properties which have part from it are called as leaf hubs.

It likewise separates the information in dataset as per portrayed disarranges. This technique gives five diverse part conditions to the development of choice tree. The conditions are Information Gain, Gain Ratio, Gini Index, Likelihood Ratio Chi-Squared Statistics, and Distance Measure. Among, the above part conditions three has a place with Polluting influence-based part condition and other two are Normalized Impurity based part criteria.

Thus, the Decision Tree Classifies the Thyroid informational collection into three classes of scatters. At the point when the informational collection is little, part criteria will make it. In any case, for extensive informational collection, it is hard to create increasingly exact choice tree. Different part rule for Decision Tree Attribute determination had been investigated and thought about. This serves to finding the thyroid maladies through the extricated principles. From this test, it is clear, that standardized based part manages have high exactness and affectability or genuine positive rate. This work can likewise be reached out for any medicinal datasets. Further improvement can be made by utilizing different streamlining Algorithms or guideline extraction Algorithms.

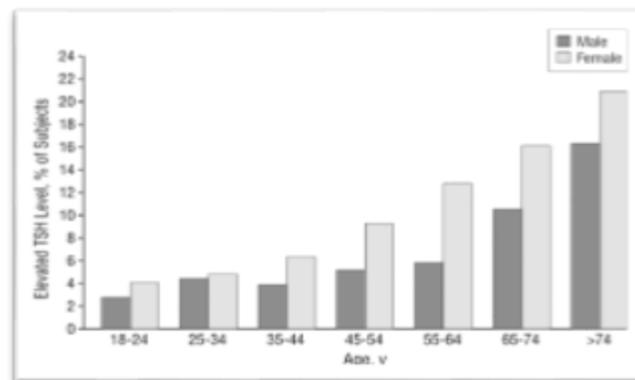


Figure 3.1.1: Prevalence of hypothyroidism in adults.

Suppositions of settling on Decision Tree:

- At the starting, the entire preparing set is considered as the root.
- Feature values are liked to be absolute. On the off chance that the qualities are constant, at that point they are discretized preceding structure the model.
- Records are dispersed recursively based on trait esteems
- Order to setting traits as root or inner hub of the tree is finished by utilizing some factual methodology. Order to placing attributes as root or internal node of the tree is done by using some statistical approach.

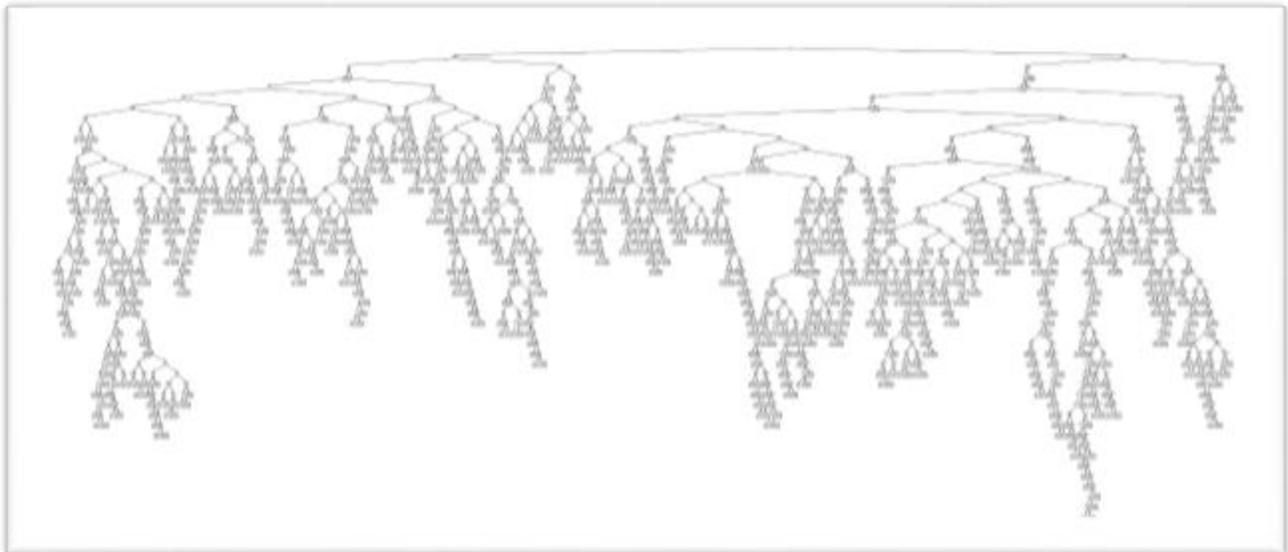


Figure 3.1.2: Classification Tree Visualization.

Neural Network

Artificial Neural Network (ANN) has been created dependent on back proliferation of mistake utilized for before forecast of thyroid ailment. The proposed trial chooses the back-engendering calculation for expectation examination of Thyroid sickness. Back proliferation is a neural system learning calculation. Neural system is a lot of associated input/output units in which every association has a weight related with it. The back-spread calculation performs learning on a multilayer feed-forward neural system. The back-engendering calculation works in two stages: proliferation and weight update.

Hypothyroidism in Bangladesh:

Thyroid organ is an imperative endocrine organ which begins working in early fetal life and keeps up the dimension of digestion in the tissues that is ideal for their typical capacity and related with physical and mental development. Awkwardness of thyroid capacity can emerge numerous serious issues. There can be numerous types of thyroid illness and Hypothyroidism is one of its significant types. An investigation was led, and this examination was meant to discover the pervasiveness of Hypothyroidism, its normal side effects, conclusion and treatment example, and learning dimension of patients with respect to this infection in Bangladesh.

In any case, alongside TSH test, T4 testing is likewise exceptionally normal in hypothyroid patients. In Bangladesh most of the hypothyroidism containing patients is treated with thyroid hormone treatment (Levothyroxine).

We have seen that the predominance of under responsive thyroid sickness among endocrine infection is high however learning dimension of patients of our nation is generally poor and should be improved. The most widely recognized reason for hypothyroidism worldwide is iodine insufficiency and in our nation's iodine lack is exceptionally normal, particularly at rustic locales of the nation. So, Government should find a way to decrease iodine lack just as there ought to be appropriate strides in our nation to make the hypothyroid patients and overall population cognizant about this malady.

Hypothyroidism is characterized as the inadequate generation of thyroid hormones from the thyroid organ. Hypothyroidism is extensively delegated an essential, optional, or tertiary infection relying upon the fundamental reason. In essential sickness there is weakened hormone discharge from the thyroid organ; in optional ailment, there is damaged TSH motioning from the pituitary; in tertiary or focal infection, the nerve center neglects to invigorate thyroid hormone discharge. Hypothyroidism runs in seriousness from subclinical infection, where patients might be asymptomatic, to out and out clinical malady, where patients are seriously influenced within the sight of numerous research facility variations from the norm.

On account of the scope of manifestation seriousness and the moderately normal and non-explicit nature of clinical discoveries, finding of hypothyroidism is exceedingly subject to research facility testing. The cutting-edge research facility test for hypothyroidism is thyroid-animating hormone (TSH).

TSH is raised in essential hypothyroidism as the pituitary reacts to the overall absence of flowing T3 and T4; TSH is irregular in all clinical and subclinical instances of essential hypothyroidism. As the illness advances toward clinical or obvious hypothyroidism, T4 and T3 become quantifiably diminished. In optional and tertiary hypothyroidism, TSH, T4, and T3 levels are dynamically irregular relying upon the span, cause, and seriousness of malady. Because of the multifaceted nature of hypothyroid etiology, lab testing for hypothyroidism is intricate and past the extent of this audit.

CHAPTER 4

DESIGN SPECIFICATION

4.1 Project Perspective

Information is the center and obligatory piece everything being equal. For our case, it very well may be saying that we need to manage the hardest of information source that can manage Hypothyroidism patient's information. This informational index for the most part has all the wonder and qualities that has been intending to use for this venture.

For our venture, it tends to be said that we need to manage the hardest of information source that can be as it manages hypothyroidism patient's information. Although we realize that hypothyroidism has turned out to be typical situation in our day by day life now daily, one of each five individuals are influenced by hypothyroidism around the globe. In any case, the kinds of information we need are not only accessible on the web or somebody's scope. It must be gain from a Medical Center, which manages hypothyroidism. There is a hypothyroidism Association in Bangladesh, which called as Bangladesh hypothyroidism (BADAS), that has broadened they're across the nation essential and tertiary social insurance administration focuses, research facility and HR to convey AMCARE administrations to all hypothyroidism patients in Bangladesh.

This task is a perception instrument for various sorts of datasets with complete outlines drawing examination arrangement. It will give the reasonable prerequisites to test all sort of datasets in physical condition. Also, it will be incredible approach to examine and think about custom datasets and their execution and separation on site-based stage.

Project Function:

The Hypothyroidism Predictor venture is a site-based information representation apparatus for a wide range of datasets, which expects to be easy to actualize and look at any information, greatly versatile and highlight rich. Hypothyroidism Predictor gives administrations to its client by putting away their information in a fixed database and helping them to picture the information examinations.

The accompanying table depicts the Hypothyroidism Predictor benefits that have designed to make up the Hypothyroidism Predictor engineering in venture:

TABLE 4.1.1: HYPOTHYROIDISM PREDICTOR SERVICES

Service Name	Disease status.
User activities	User will give their information according to inquiry.
Prediction of Hypothyroidism	Provides probability of Hypothyroidism to the user.
Upload Data	User will upload their information.

Real component of that executed frameworks is given beneath:

- 1) It has an amazingly adaptable design
- 2) It is anything but difficult to alter
- 3) It is anything but difficult to incorporate with other informational indexes.

User Classes and Characteristics

The user classes will be user and admin. A user can input new data sets and get result and a message.

Machine Learning Algorithms

Supervised Learning

All data is labeled and the algorithms learn to predict the output from the input data[1].

A model is prepared through a training process in which it is required to make predictions and is corrected when those predictions are wrong.

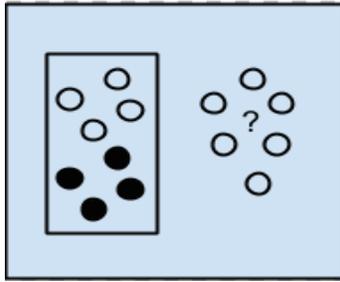


Figure 4.1.1: Supervised Learning.

All information is named, and the calculations figure out how to anticipate the yield from the information data [1].

A model is set up through a preparation procedure in which it is required to make forecasts and is rectified when those expectations aren't right. The preparation procedure proceeds until the model accomplishes an ideal dimension of exactness on the preparation information. To take care of a given issue of directed learning, one needs to play out the accompanying advances:

1. Determine the kind of preparing models. The client ought to choose what sort of information is to be utilized as a preparation set.
2. Gather a preparation set. The preparation set should be illustrative of this present reality utilization of the capacity. Consequently, a lot of info objects is assembled and comparing yields are likewise accumulated, either from human specialists or from estimations.
3. Determine the info include portrayal of the educated capacity. The precision of the scholarly capacity depends unequivocally on how the info object is spoken to.
4. Determine the structure of the educated capacity and relating learning calculation. For instance, the architect may utilize choice trees.
5. Evaluate the precision of the scholarly capacity. After parameter modification and learning, the execution of the subsequent capacity ought to be estimated on a test set that is independent from the preparation set.

Unsupervised Learning

All data is unlabeled, and the algorithms learn to inherent structure from the input data [1].

A model is prepared by deducing structures present in the input data. This may be to extract general rules. It may be through a mathematical process to systematically reduce redundancy, or it may be to organize data by similarity.

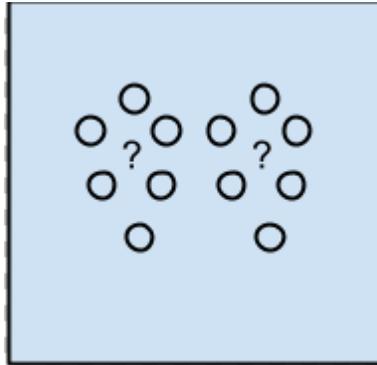


Figure 4.1.2: Unsupervised Learning.

Operating Environment:

Requirements for Hypothyroidism Predictor

- ❖ Internet Connection
- ❖ Browser [In pc/mobile (any)]

4.2 Design and Implementation Constraints

General constraints for the system are given below:

- The interface will be in English only.
- There is no maintainability or backup so availability will get affected.
- The system is a single user system.

Challenges

We face a huge amount of challenges in this section, as data collection is the most difficult phase of a research work. Not just data collection, but data organizing, data filtering, data fitting all are very tough job to do. Let's us mentions shortly some of the challenges of this section.

- To determine the types of data that we need.
- To determine the types of attributes fittest best.
- To collecting information about where to get this data (particular organization).
- How to get there and by which sources.
- Rearranging data in the tabular formats.
- Designing the correct structure for the research database
- Shorting data's into the research database.
- Fitting data's into the research database.
- Learning the way to convert an excel file into arff (Attribute Relational File Format) file, that's what Weka support as input file.
- To introduce with Weka and its associates facilities.
- To introduce with all classification algorithm that provide by Weka, especially with J48 and Naïve Bayes algorithms.
- To learn how to interpret the outcomes of the classification algorithms.
- Determining what experiments we should do with Weka.
- Learning how to create Training and Testing Dataset by using the original Dataset and Weka.
- Knowing to run Training Dataset and interpret result using Weka.

CHAPTER 5

REQUIREMENT SPECIFICATION

5.1 Introduction

Each application needs some Hardware segments and certain Software assets to be available on the framework to run proficiently. These essentials are known as Requirement Specification and are frequently utilized as a rule rather than an outright principle.

5.2 General System Requirements

General system requirement deals with both minimum amounts of things. To run “Hypothyroidism Predictor” one should need the following:

1. Internet connection
2. Browser [pc/mobile(any)]

With above required things one should be able to work with “Hypothyroidism Predictor”.

5.3 Use Case Model

The accompanying figure demonstrates the utilization case show for the task. We have made it so basic and simple to use for anybody. No compelling reason to enroll. Clients can without much of a stretch give contribution of THS and FT4 to get their outcome.



Figure 5.3.1: Unsupervised Learning.

5.4 Use Case Description

It contains data pretty much all the utilization instance of the Use case Model appeared in the above picture. We have depicted each utilization case through a table in the beneath area. Each table shows five sorts of detail data about a utilization case. They are Use case name, Actor, Pre-condition, Primary Path and Exceptional Path. Use case name implies the name of the procedure. Performing artist characterizes will's identity the client. Pre-condition tells whether another procedure is expected to enter or utilize this procedure.

Table 5.4.1 portrays about "Information" procedure of our task and it gives us a seeing how the procedure functions in venture. Register process has essential ways.

TABLE 5.4.1: USE CASE DESCRIPTION OF INPUT

Use case name:	INPUT
Actor:	User
Pre-condition:	Website, Internet connection
Primary Path:	<ol style="list-style-type: none">1. Go to the link given2. Give input
Exceptional Path:	<ol style="list-style-type: none">1. Check Internet connection.

Table 5.4.2 describes about "OUTPUT" process of our project and it gives us an understanding how the process works in project.

TABLE 5.4.2: USE CASE DESCRIPTION OF OUTPUT

Use case name:	OUTPUT
Actor:	User
Pre-condition:	Input
Primary Path:	<ol style="list-style-type: none">1. Click result
Exceptional Path:	<ol style="list-style-type: none">1. Check input

CHAPTER 6

SYSTEM DESIGN IMPLEMENTATION

In this area we will talk about our proposed strategy for anticipating hypothyroidism hazard levels. Here we investigated the information accumulation way, information preparing, worked stream methodology, factor scoring technique and exactness correlation of various calculations.

6.1 Data Flow Diagrams

Data Flow Diagrams (DFD) is an approach to speak to framework necessities in a graphical structure which lead to plan. A DFD portrays what information stream (consistent) instead of how they are prepared, so it doesn't rely upon equipment, programming, information structure.

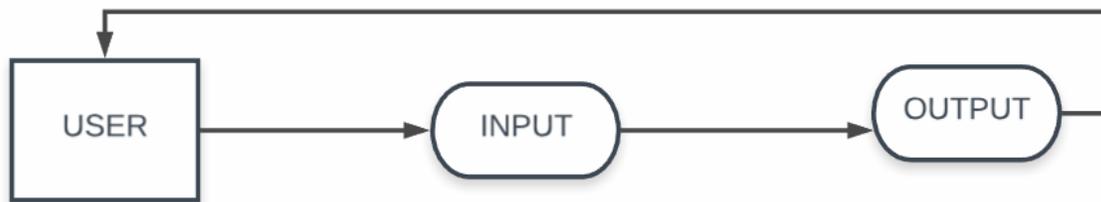


Figure 6.1.1: Data Flow Diagram

CHAPTER 7

CONCLUSION AND FUTURE SCOPE

7.1 Conclusion

As end, this undertaking is to build up the framework that ready to distinguish whether the individual is decidedly or contrarily having Hypothyroidism sickness.

7.2 Future Scope

Also, in future we are wanting to reach out venture and attempt to build the level of precise outcomes.

And furthermore, we will redesign our site and include some increasingly vital information. Which can almost certainly anticipate by possess utilizing AI. Furthermore, this is our principle focus for overhauling our framework with no blunders.

APPENDIX

Appendix A: Project Reflection

The purpose of this appendix is about project reflection. From semester we have started journey to make this website. The main feature of our website is, give every information about hypothyroidism, a total guideline.

By using this a patient or people can easily know any information about this diseases, can see results, can get notice about any update.

The project "Hypothyroidism Predictor" will be very helpful. So, we think and we also believe that our website will be a positive application for users.

Appendix B: Related Diagram Plagiarism Check Report

Turnitin Originality Report

Processed on: 18-Apr-2019 00:00 +06
 ID: 1114415444
 Word Count: 8708
 Submitted: 1

Similarity Index	Similarity by Source	
14%	Internet Sources:	14%
	Publications:	3%
	Student Papers:	N/A

Hypothyroidism indicator: Study of
 Detecting Hypothyroidism utilizing Data
 Mining By Rezwana Sultana

5% match (Internet from 15-Apr-2019) https://www.techleer.com/articles/438-a-list-of-top-data-mining-algorithms/
2% match (Internet from 22-Mar-2019) https://mmfinfotech.com/Blog/tag/data-mining-services/
1% match (Internet from 02-Jan-2019) http://biglistofwebsites.com/list-top-websites-like-iiu.edu.pk
1% match (Internet from 24-Oct-2012) http://research.ijcaonline.org/volume44/number8/pxc3878474.pdf
1% match (Internet from 07-Apr-2019) https://www.ivanschmidt.com/articles/show/a-tour-of-machine-learning-algorithms---228603
< 1% match (Internet from 09-Apr-2019) https://scholar.sun.ac.za/bitstream/handle/10019.1/104971/walters_demonstration_2018.pdf?sequence=2
< 1% match (Internet from 10-Sep-2018) http://mdpi.com/2313-433X/4/7/95/htm
< 1% match (Internet from 01-Aug-2016) https://docs.com/flavio-bernardo/2291/data-mining-practical-machine-learning-tools-and
< 1% match (Internet from 17-Jul-2018) https://www.ijcaonline.org/archives/volume154/number3/26474-2016912066
< 1% match (Internet from 22-Mar-2016) http://ijarcse.com/docs/papers/Volume_5/6_June2015/V5I6-0460.pdf
< 1% match (Internet from 15-Apr-2019) https://estudogeral.sib.uc.pt/bitstream/10316/39072/1/Knowledge%20elicitation%20by%20merging%20heterogeneous%20data%20sources%20in%20a%20die%20casting%20process.pdf
< 1% match (Internet from 22-Oct-2017) https://documents.mx/documents/rashmi.html
< 1% match (Internet from 28-Mar-2019) https://cv.archives-ouvertes.fr/yasmine-bouagga
< 1% match (publications) John Hunt. "The Unified Process for Practitioners", Springer Nature, 2000

References

- [1] R. Quinlan, "UCI Machine Learning.," 10 January 2016. [Online]. Available: <http://archive.ics.uci.edu/ml/datasets/thyroid+disease>. [Accessed 3 November 2018 at 12:35 AM].
- [2] "Data Mining Software," June 2016. [Online]. Available: <https://www.cs.waikato.ac.nz/ml/weka/index.html>. [Accessed 13 September 2018 at 10:00 PM].
- [3] P. R. Orlander, "Data processing," 21 Nov 2018. [Online]. Available: <https://emedicine.medscape.com/article/122393-overview>. [Accessed 03 February 2019 at 2:30 PM].
- [4] R. Li, PWM Industry, 6-7 May 2015. [Online]. Available: <https://www.kdnuggets.com/2015/05/top-10-data-mining-algorithms-explained.html>. [Accessed 20 March 2019 at 4:00 AM].
- [5] J. McIntosh, "Facts of Hypothyroidism.," 2 January 2018. [Online]. Available: <https://www.medicalnewstoday.com/articles/163729.php>. [Accessed 10 November 2018 at 5:00 AM].
- [6] A. Samuel, "Machine Learning," [Online]. Available: https://en.wikipedia.org/wiki/Machine_learning. [Accessed 6 December 2018 at 1:00 AM].
- [7] L. J. Jitender Deogun, "Prediction Mining- An Approach to Mining association.," [Online]. Available: https://link.springer.com/chapter/10.1007/11548706_11. [Accessed 12 June 2018 at 8:00 AM].