

# Data Mining for Internet of Things

BY

Neeti Joshi

15MCEI09



DEPARTMENT OF <NAME OF THE DEPARTMENT>

AHMEDABAD-382481

January 2013

# Data Mining for Internet of Things

Seminar

Submitted in partial fulfilment of the requirements

For the degree of

Master of Technology in Computer science and Engineering [INS]

By

Neeti Joshi

15MCEI09



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

AHMEDABAD-382481

January 2016

## Certificate

This is to certify that the Seminar entitled "Data Mining For Internet of Things" submitted by Neeti Joshi, towards the partial fulfilment of the requirements for the degree of Master of Technology in Computer Science [INS] of Nirma University, Ahmedabad is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this Seminar, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Prof. Vijay Ukani

Assistant Professor,

Computer Science And  
Engineering ,

Institute OF Technology,  
Nirma University, Ahmedabad

Dr. Sharada Valivati

Associate Professor,

Computer Science And  
Engineering,

Institute OF Technology,  
Nirma University, Ahmedabad

Dr. Sanjay Garg

Professor and Head,

Computer Science And Engineering,

Institute of Technology,

Nirma University, Ahmedabad

## Abstract

In Today's environment smart things are come into the daily life. Everything is now smart, intelligent, and self –automated. Each smart component has been integrated with another smart component. By doing this now a day smart system has been introduced. This is area is called Internet of things. Internet of things contains large amount of datasets. So it is difficult to analyse easily. So here we will integrate the concept of data mining in internet of things to analyse the large set of data. By using data mining concept the application of internet of things can be workable. Because without any predication, classification, analysis of data it is impossible to work with large dataset.

So here we will applying a data mining concept on the application of echocardiogram. Which one is original dataset taken by medical institute, and we will apply data mining concept to analyse the situation of patients.

## Acknowledgements

I would like to express my sincere gratitude to my advisor **Prof. Vijay Ukani** for the continuous support of my seminar and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this report. I could not have imagined having a better advisor and mentor for my Research Seminar.

-Neeti Joshi

15MCEI09

# Contents

---

1	Introduction	7
1.1	IOT .....	7
1.2	Data Mining .....	8
1.3	Why Data ming for IOT?	
1.3	Objective of Study .....	9
1.4	Scope of Work .....	9
2	Literature Review	
2.1	Architecture	10
2.2	Application	11
3	IMPLEMENTATION	13
3.1	Problem Statement .....	13
3.2	Data Set .....	14
3.3	Implementation. ....	16
4	Conclusion and Future Scope	18
3.1	Conclusion .....	18
3.2	scope .....	18

## Chapter 1

---

# Introduction

---

As the time is going the things are going to be more and more smarter than previous versions. Our main concept is to make the application which reduce the working time of human beings. The internet of things follows this concept. But to make this idea possible there should be some intelligence like human being which can take the decision by its own with accurately and efficiency. So the data mining approach fulfil this idea to become a true.

### 1.1 Internet of things :

In informal way we can say that internet of things is anything can be connected to anything by using any service at any place using any network. Basically internet of things is nothing but the collection and connection of the sensors. The all the smart application can be possible by use of sensors. This sensors collect the data. And give it to some intelligent system for further procedure.

These are the features of Internet of things have.

- Intelligent interfaces
- Self- automation capability
- Dynamic network
- Distributed system
- Standard and interoperable communication
- Integration of the system

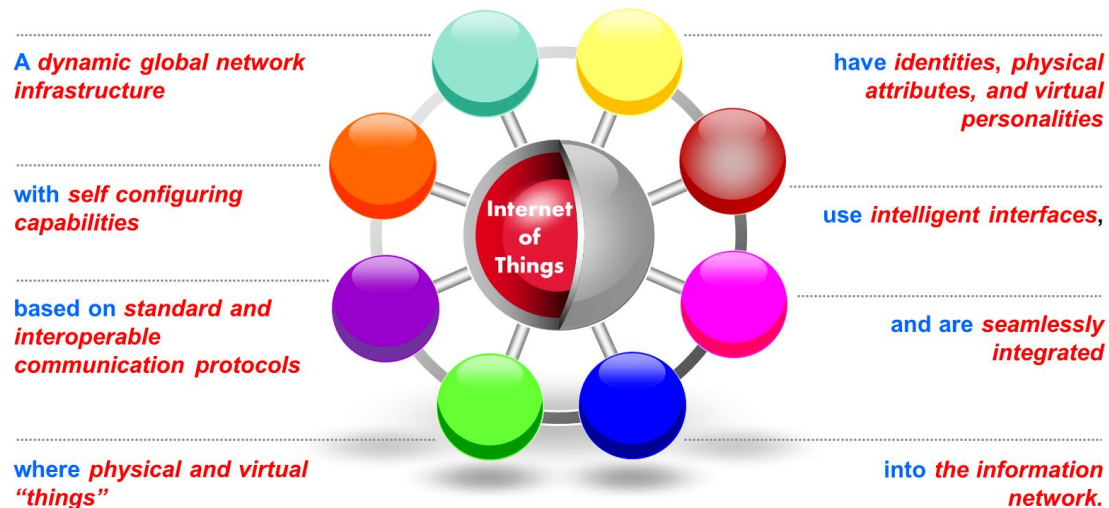


Fig: internet of Things

## 1.2 Data mining :

Data mining is the area which provides some intelligent activity to the system. It is the activity about mining fruitful from the system. As per the technical terms data mining is “It is the process of analysing the data from different resources.” This different resources can be different databases, data warehouses, data sets, different departments etc. and analysis can be classification, clustering, association, generalization, supervised, unsupervised etc.

## 1.3 Why Data Ming for IOT?

Applications of the IOT have so much complex datasets. It provides large amount of massive datasets. So data mining fetch the useful information from them.

Consider an Example of “SMART HOME” as an IOT Application, in which all the components of the smart home are self- automated. It has datasets of kitchen, drawing hall etc. talking about kitchen itself has large datasets of refrigerator, oven, spoon, bowl etc. so it these all the data sets are meaningless without any intelligent activities. By applying data mining algorithms the useful information will be fetch as per the need. Not all the data is needed for each activity.

So data mining provides one kind of brain to the IOT applications.



### 1.3 Objective of Study :

Objective of the study is to learn basic concepts of IOT and data mining. Also learn about the application of IOTs. Brief study of data mining algorithm and important thing is to relate the IOT application to data mining algorithm that how the data mining is performed. Which application need which kind of mining etc. Also learn data mining concept practically by taking one IOT application example and by doing implementation conclude the algorithm with the IOT dataset.

### 1.4 Scope of Work

Here I am concentrating only on the datasets of IOT and data mining algorithm. Here hardware implementation is neglected. The Scope of the work is to use the IOT dataset (which has been taken by sensors), Apply Data Pre-processing, Apply appropriate data mining algorithm which belongs to classification, clustering, regression, Association rules etc .

# Literature Review

## 2.1 Architecture Of Data mining for IOT:

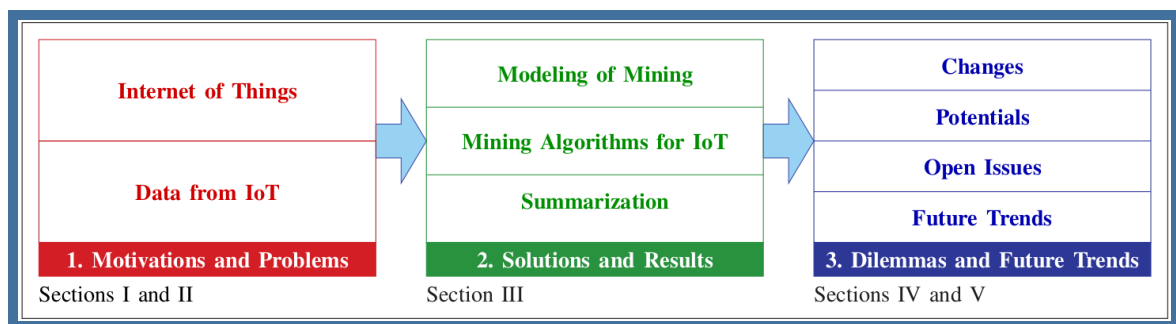


Fig2.1: Architecture of IOT and Data mining

### 2.1.1 Section 1 and 2:

In this section is over view of overall processes. Here the IOT application is to be selected and the collection of the datasets are to be prepared for further procedures. These data sets should be collected from the sensors. These data sets are streamline. They are vary at each instances.

### 2.1.2 Section 3:

This section is core of IOT and Data mining. In this section the IOT and Data mining are integrated. The matching data mining algorithms are to be selected. The way of mining, generalization , summarization is to be done at this section. Each different data set may need different algorithm for each different dataset. This phase gives the solution and direction to the application by using datasets.

### 2.1.3 Section 4 and 5:

In this section the

- changes should be specified like change in the Thing-Oriented, Internet Oriented, Semantic Oriented
- Potentials like People-Oriented, self-Oriented, Thing Oriented

- Issues should be solved like Privacy, Security, How to combine data mining technology
- Future trends should be specified like Big Data, Smart Grid, Ontology, Cloud Computing

## 2.2 Application of IOT and Data mining:

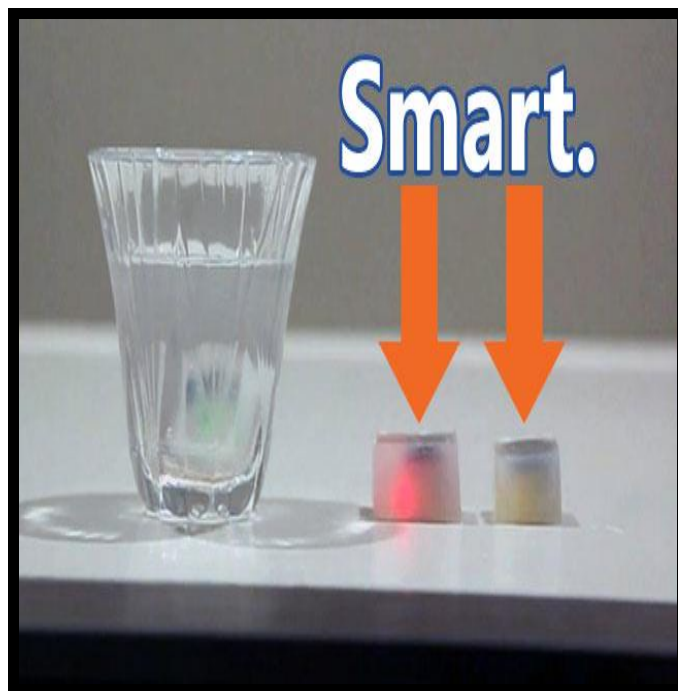
### 2.2.1 Traffic Analysis:

Traffic Analysis application directs the driver to which road he/she has to go. This application detects the traffic crowd on the road which we have mention as source to destination. It will detect from the setalite view and by the use of GPS system.

#### HOW THE DATA MINING WORKS?

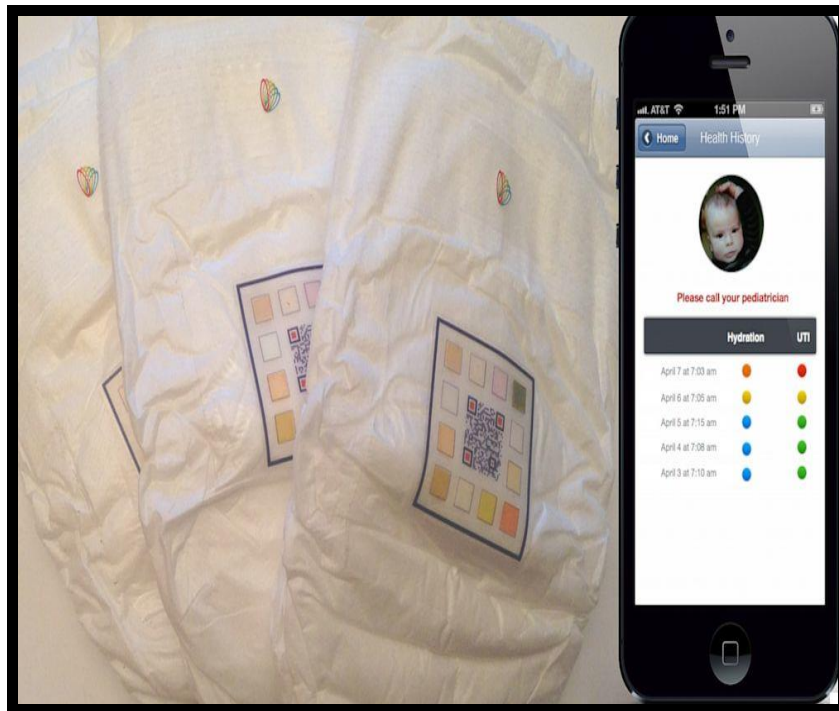
It will collect the data set of traffic and data mining algorithm will classify whether the route is convenient to go or not.

### 2.2.2 Smart ICE:



It will measures the speed of drinking of human being and if data mining find that the speed is high it will raised the alert that the speed is too high please drink slowly.

### 2.2.3 Smart Diapers:



This application checks the health of the child without the need of doctor. Data mining algorithm reports whether the health of the child is ok or not. It will check using the data of frequency of urine and amount of urine.

## Chapter 3

---

# Implementation

---

### 3.1 Problem Statement:

- Data set of patients who are suffering from the heart attack. Predict by using the data set of echocardiogram the patient will alive after the 1 month of the heart attack ?

### 3.2 Dataset Information:

- Source Information:
  - Donor: Steven Salzberg (salzberg@cs.jhu.edu)
  - Collector: Dr. Evlin Kinney
  - Location: The Reed Institute ,P.O. Box 402603 ,Maimi, FL 33140-0603,
  - Date Received: 28 February 1989
- Attributes: 13
- Instances: 59
- Missing Values: None ( I have removed missing values)
- Using Software: Weka ( version 3.6)

#### 3.2.2 Attribute information:

1. **Survival** -The number of months patient survived (has survived, if patient is still alive). Because all the patients had their heart attacks at different times, it is possible that some patients have survived less than one year but they are still alive. Check the

second variable to confirm this. Such patients cannot be used for the prediction task mentioned above.

2. **Still-Alive** -A binary variable. 0=dead at end of survival period, 1 means still alive
3. **Age-at-heart-attack** -Age in years when heart attack occurred
4. **Pericardial-Effusion-Binary**. Pericardial effusion is fluid around the heart. 0=no fluid, 1=fluid
5. **Fractional-Shortening** -A measure of contractility around the heart lower numbers are increasingly abnormal
6. **Epss** - E-point septal separation, another measure of contractility. Larger numbers are increasingly abnormal.
7. **Lvdd** - Left ventricular end-diastolic dimension. This is a measure of the size of the heart at end-diastole. Large hearts tend to be sick hearts.
8. **Wall-motion-score** -A measure of how the segments of the left ventricle are moving
9. **wall-motion-index** -Equals wall-motion-score divided by number of segments seen. Usually 12-13 segments are seen in an echocardiogram. Use this variable INSTEAD of the wall motion score.
10. **Mult** -A derivate var which can be ignored
11. **Name** - the name of the patient (I have replaced them with "name")
12. **Group** - meaningless, ignore it
13. **Alive-at-1** -Boolean-valued. Derived from the first two attributes.0 means patient was either dead after 1 year or had been followed for less than 1 year. 1 means patient was alive at 1 year.

### 3.2.3 Data set :

Here I am working on small data set for more efficient result.

Survival	Still Alive	Age at heart attack	name	alive-at-1
short	no	old	name	yes
avg	no	middle	name	no
long	no	middle	name	no
short	yes	middle	name	yes
long	no	young	name	no
short	no	young	name	yes
long	yes	old	name	yes
long	yes	old	name	no
long	yes	middle	name	no
short	no	young	name	no
long	no	old	name	yes
long	no	old	name	no

Fig 3.2.3 Data set

### 3.3 Implementation in weka :

Here I am using Decision tree algorithm because I have categorical attributes. And here the LABEL attribute is “ ALIVE AT 1”. here I am using ID3 algorithm for more specific and accurate result.

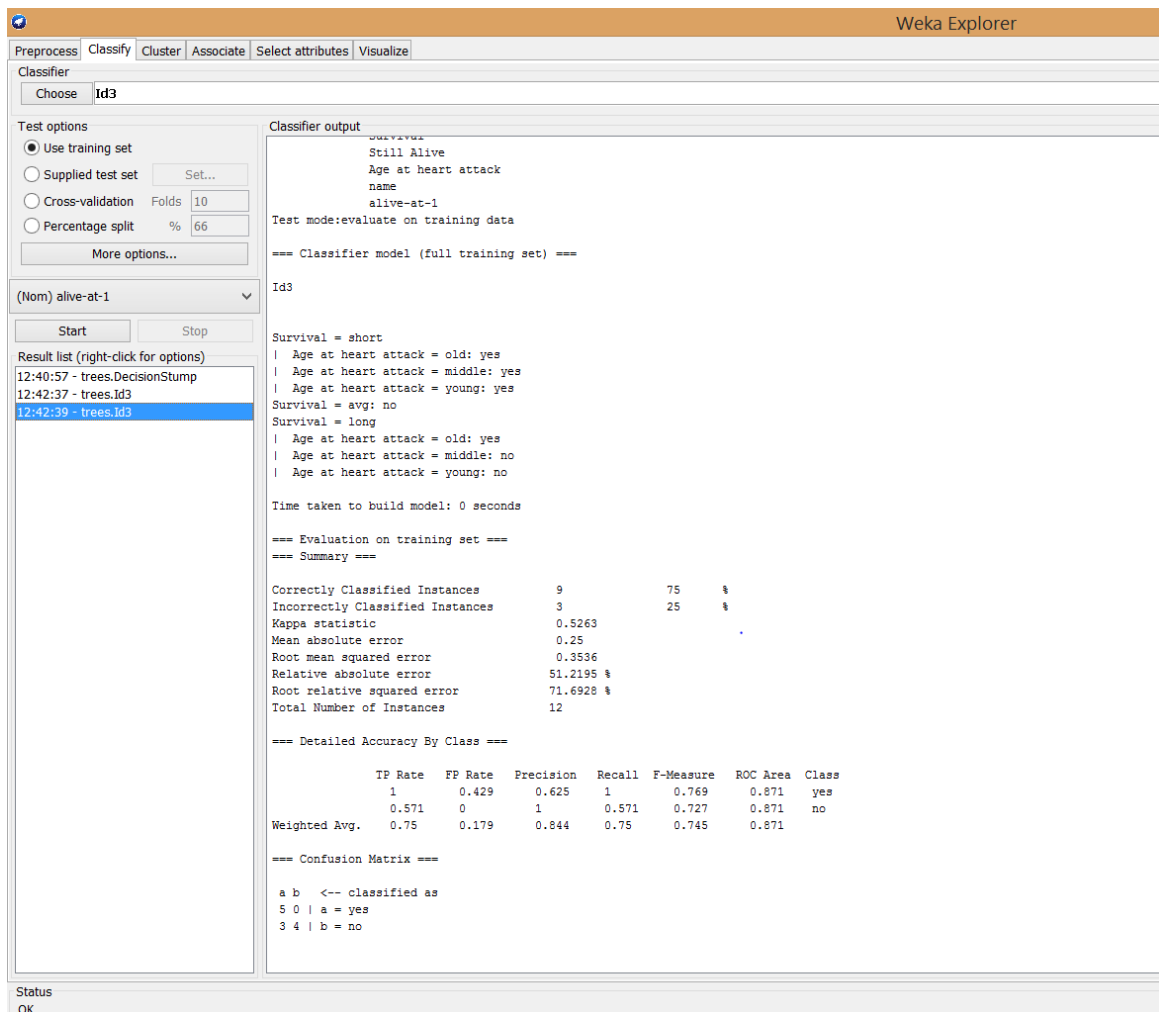
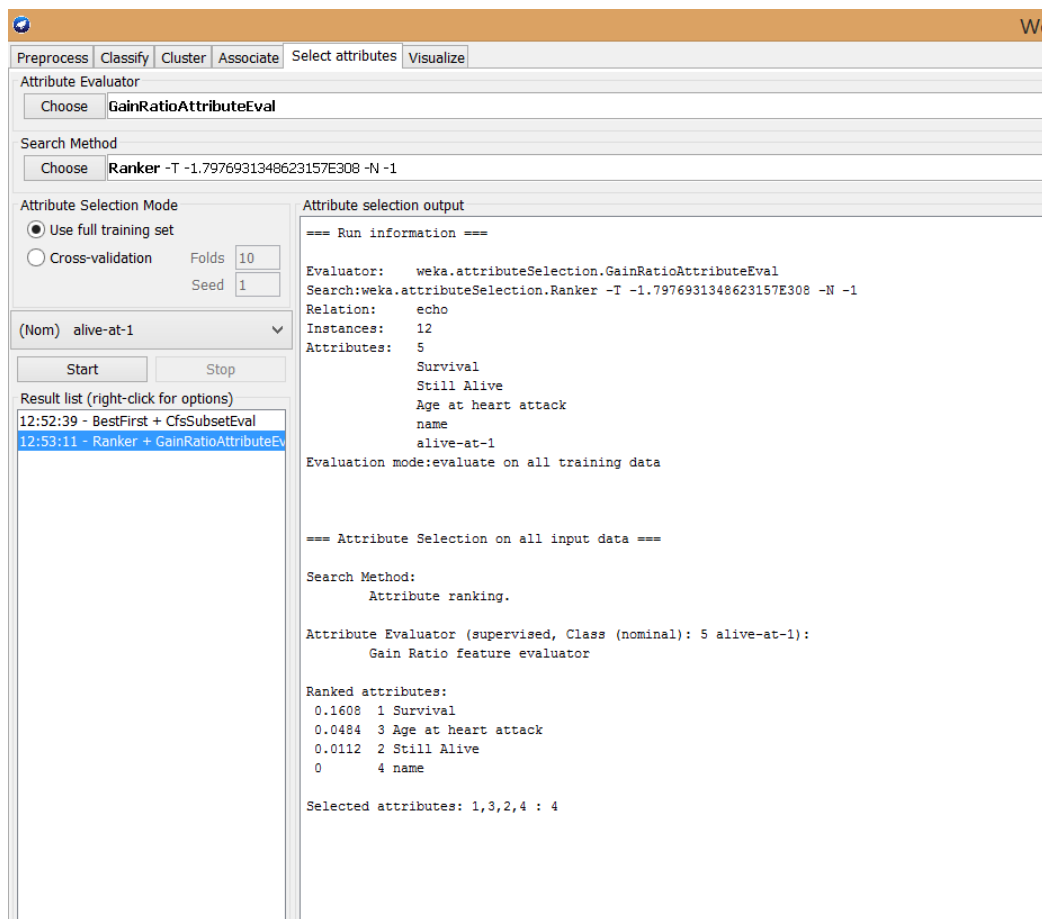


Fig: 3.3.1 output of the data set :

Here the main root node is survival which has large information gain.

- Correctly classified instances: 9
- Incorrectly classified instances: 3
- Mean absolute error: 0.25





**Fig: 3.3.2 rank wise information gain:**

Rank wise information gain is:

1. Survival:= 0.16608
2. Age at heart attack=0.0484
3. Still Alive=0.0112
4. Name=0

# Conclusion and Future Scope

---

### 4.1 Conclusion

On the basis of the research seminar , we have shown the how the IOT applications are used by using data mining algorithm. Also learn the problems and issues related to data mining using the IOT application. The biggest challenge is to select the appropriate algorithm for IOT dataset and how to integrate it with another algorithm is another biggest challenge.

### 4.2 scope

The scope of the seminar project is only on dataset of the IOT applications and implementation of algorithms by using appropriate data mining algorithms.





