1. **HEALTH MONITORING SYSTEM**
   1. **INTRODUCTION:**

* **Background of the study**

What is Health Monitoring System?

Health monitoring system is an extension of a hospital medical system where a patient’s vital body state can be monitored. Traditionally the detection systems were only found in hospitals and were characterized by huge and complex circuitry which required high power consumption. Continuous advances in the semiconductor technology industry have led to sensors and microcontrollers that are smaller in size, faster in operation, low in power consumption and affordable in cost.

This has further seen development in the remote monitoring of vital life signs of patients especially the elderly. The remote health monitoring system can be applied in the following scenarios:

i) A patient is known to have a medical condition with unstable regulatory body system. This is in cases where a new drug is being introduced to a patient.

ii) A patient is prone to heart attacks or may have suffered one before. The vitals may be monitored to predict and alert in advance any indication of the body status.

iii) Critical body organ situation

iv) Situation leading to development of a risky life threatening condition. This is for people at an advanced age and may be having failing health conditions.

v) Athletes during training. To know which training regimes will produce better results.

In recent times several systems have come up to address the issue of health monitoring.Some have even adopted a service model that requires one to pay a subscription fee. In developing countries this is a hindrance as some people cannot use them due to cost issue involved. There is also the issue of internet connectivity where some systems to operate good quality internet for a real-time remote connection is required. Internet penetration is still a problem in developing countries.

Many of the systems introduced work best in the developed countries where the infrastructure is working perfectly. In most cases the systems are adapted to work in developing countries. To reduce some of these problems there is need to approach the remote detection from a ground up approach to suit the basic minimal conditions presently available in developing countries.

In developing countries, just after retiring from their daily career routine majority of the elderly age group, move to the rural areas. In developed countries they may move to assisted living group homes. This is where the health monitoring system could be come in handy **.**

**MACHINE LEARNING:**

Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, machine learning has given us self-driving cars, practical speech recognition, effective web search, and a vastly improved understanding of the human genome. Machine learning is so pervasive today that you probably use it dozens of times a day without knowing it. Many researchers also think it is the best way to make progress towards human-level AI.

* **Purpose of the study**

Design a health monitoring system in order to predict the health condition of the patient based on the input parameters such as Systolic blood pressure(SBP),Diastolic blood pressure(DBP),Pulse, temperature.

**1.2 Objective of the study**

The objective of the project was to come up with a system that can monitor and provide physiological information in the home itself. The monitoring system would be useful for elderly or chronically ill patients who would like to avoid a long costly hospital stay. Wireless sensors would be used to collect and transmit signals of interest and a microcontroller was programmed to receive and automatically analyze the sensor signals. In the project, I was to select the appropriate sensors for signal detection and design algorithms to realize the detection.

The design process was covered from the beginning to the end how to predict the information about patients health condition.

Two advantages of the system that clearly come out are that the fall detector allows free movement of the patient or person of interest and the method of acquiring the cardiac signal is non-invasive.

**1.3 Problem Statement**

**“TO PREDICT THE HEALTH CONDITION OF PATIENT USING HEALTH MONITORING SYSTEM**”

health monitoring can provide useful physiological information in the home. This monitoring is useful for elderly or chronically ill patients who would like to avoid a long hospital stay. Wireless sensors are used to collect and transmit signals of interest and a processor is programmed to receive and automatically analyze the sensor signals. In this project you are to choose appropriate sensors according to what you would like to detect and design algorithms to realize your detection. Examples are detection of a fall, monitoring cardiac signals, brain signal monitoring (EEG), and in-home ultrasound.

Using a single parameter monitoring system an approach to a remote health monitoring system was designed that extends healthcare from the traditional clinic or hospital setting to the patient's home. The system was to collect a heartbeat detection system data and a fall detection system data. The data from the two single parameter monitoring systems was then availed for remote detection.

During design the following characteristics of the future medical applications were adhered

to

a) Integration with current trends in medical practices and technology,

b) Real-time, long-term, remote monitoring, miniature, wearable sensors. Long battery life of designed device

c) Assistance to the elderly and chronic patients. The device should be easy to use with minimal buttons

**2.REVIEW OF LITERATURE**

**Review on healthcare monitoring systems**

Subject of study “healthcare monitoring systems”, covers areas of interest in both electrical engineering and medical field of study. It has led to the direction of Biomedical engineering field of study.

Previous research referred to remote “health detection systems” as “mobile health” or “mHealth”. This was because they used mobile phones prior to smartphone era . At that time a mHealth alliance existed that identified barriers, gaps in scaling and use of mobile technology in healthcare . Proactive efforts have seen the barriers in healthcare and mobile communication technology being reduced.

According to the journal of Neuro Engineering and Rehabilitation most of non-invasive techniques used in acquiring critical signals from the human body are of microvolt (μV) nature signal. Signal processing by use of microcontrollers is then done on the detected signals to acquire the meaningful information from the signal data. Errors such as physical body modeling errors, source modeling errors and noise (instrumental or biological) are factored in the computation during signal acquisition and processing.

Patients suffering from chronic conditions such as Congestive Heart Failure (CHF), Chronic Obstructive Pulmonary Disease (COPD), diabetes, asthma, hypertension and some other health conditions can benefit from the remote health care management taking advantage of patient health monitoring technology

Once the patient is referred to the healthcare program by a healthcare professional in the hospital or at a primary care facility, they are introduced to a healthcare specialist team that will proceed to track and adjust the execution of the care plan as well as provide support and guidance for the patient.

**The healthcare system offers patients:**

i) Access to an application on a mobile tablet or a personal computer that can be used from their home.

ii) Use of an intuitive, step-by-step application based on pre-scheduled questions that they need to answer. In some cases, this can be several times a day.

iii) Seamless integration with electronic medical devices (blood pressure cuffs, etc.) that can capture health data that is shared with the healthcare provider.

**The healthcare system offers Healthcare specialist, Doctors:**

i) Access to a centralized view of all patients on the HHM program, allowing clinicians to tailor workflows, protocols and interventions, creating customized care plans according to a patient’s condition and status.

ii) Easy analysis of results, empowering them to adjust treatment based on best-practice guidelines and protocols.

iii) Alerts and reminders that trigger patient alerts which can be generated by forms created by the clinician or from data obtained from the patient (i.e. high blood pressure alert).

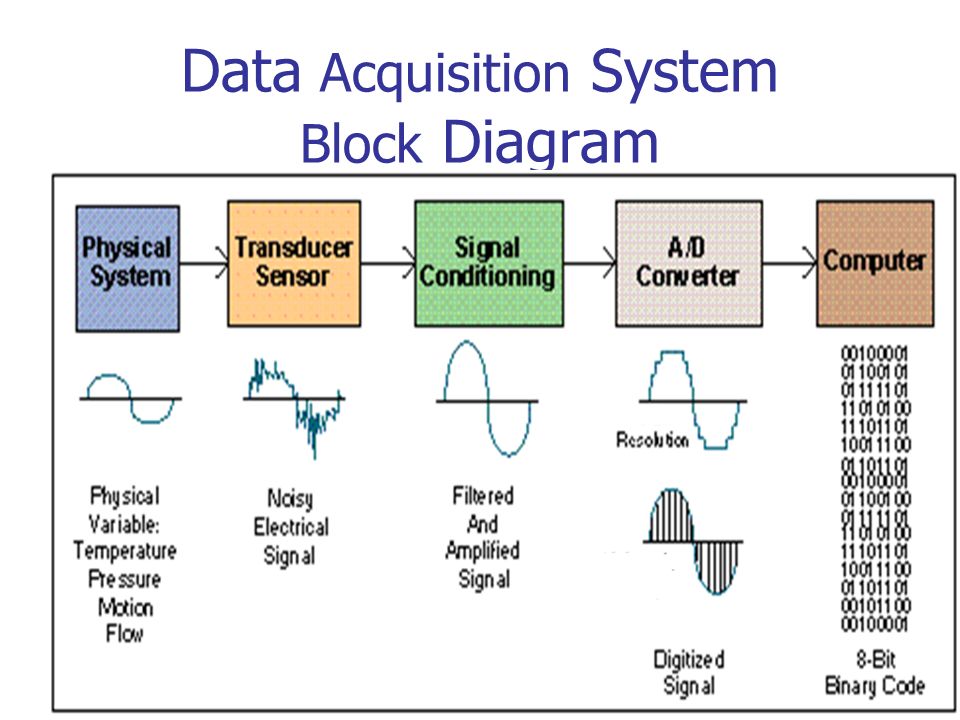
iv) Sophisticated care coordination through better organization of multidisciplinary teams, assignment of interventions and tasks, and the ability to view past, present and future interventions.

v) Asset management, making it easy to manage devices – including location and to whom they are assigned.

**Block diagram of the symbolizing data acquisition system:**

They are the Data acquisition, data processing and data communication unit

Fig. 2-1 data acquisition block diagram



* **Literature review on Cardiac signal detection**

Heart is a vital organ whose physiological heartbeat needs to be monitored for everyone. This is more critical for people with heart conditions, athletes, chronically ill and elderly in society. As such it is important to monitor cardiac signals. In hospital a repeated representation of successive heartbeat waveforms is done by Electrocardiogram (ECG or EKG) detector.

Previous attempts at ECG detection had an introduction of basic microcontrollers for the complex signal analysis. With the advancement of technology powerful processors have been included in smartphones that can run apps to assist in signal processing. Smartphones equipped with infrared flash for camera can also be repurposed using an application as cardiac signal detectors. They can remotely do the signal acquisition/detection, processing and communication simultaneously. Arrhythmias caused by the heart's electrical conduction system disorders . Irregular heartbeats can cause irregular blood flow to other organs and cause damage to vital organs such as brain. There are three major conditions that can be detected from an ECG plot. Brachycardia (heart rate < 60), tachycardia (heart rate > 100) and an unstable heart rate.

**Literature review on fall detection**

A fall may be defined as unintentionally coming to the ground by a person. A fall if violent may cause loss of consciousness by a person or a fracture of the bone structure. Falls in the elderly or chronically ill patient is a major problem in public health as it causes many disabling fractures but also has dramatic psychological consequences which reduce the independence of the person. If a pattern of activities of a person can be detected they can be prevented. As a result fall detector is necessary. With recent developments in IC industry and telecommunication, system devices that encompass accelerometers and gyro-meters the medical profession has observed and through the help of some engineering some of the sensors can be repurposed. The accelerometer found in some of the smart phones available in shops today can be set to tell when a fall has occurred. This is more important for chronically ill or elderly people . For the chronically ill or people prone to falls it is important to monitor their particular daily behaviors. The same applies to elderly living on their own. Athletes can also benefit from this by monitoring and logging their training regimes. The human physiological data is collected for their normal situations of activity.

A popular commercial device available in the market that addresses the fall detection issue is the Life™ fall detector sensor by alert one . This is designed specifically for fall detection. An old Smartphone with an accelerometer in it can also be used for fall detection by use of an app. The app accesses the phone hardware interface to get the accelerometer data, process it and upload it to a database via its internet connection.

Previous fall detector systems were huge and cumbersome and had mechanical switches with physical damper springs as tilt sensors. Recent advances in development of Micro Electro Mechanical Systems (MEMs) has seen the sensors miniaturized to small ICs that fit into PCBs and can be made into wearable technology.

**3. DATA COLLECTION:**

While a range of health and health care entities collect data, the data do not flow among these entities in a cohesive or standardized way. Entities within the health care system face challenges when collecting race, ethnicity, and language data from patients, enrollees, members, and respondents. Explicitly expressing the rationale for the data collection and training staff, organizational leadership, and the public to appreciate the need to use valid collection mechanisms may improve the situation. Nevertheless, some entities face health information technology (Health IT) constraints and internal resistance. Indirect estimation techniques, when used with an understanding of the probabilistic nature of the data, can supplement direct data collection efforts.

Addressing health and health care disparities requires the full involvement of organizations that have an existing infrastructure for quality measurement and improvement. Although hospitals, community health centers (CHCs), physician practices, health plans, and local, state, and federal agencies can all play key roles by incorporating race, ethnicity, and language data into existing data collection and quality reporting efforts, each faces opportunities and challenges in attempting to achieve this objective.

To identify the next steps toward improving data collection, it is helpful to understand these opportunities and challenges in the context of current practices. In some instances, the opportunities and challenges are unique to each type of organization; in others, they are common to all organizations and include:

How to ask patients and enrollees questions about race, ethnicity, and language and communication needs.

How to train staff to elicit this information in a respectful and efficient manner.

How to address the discomfort of registration/admission staff (hospitals and clinics) or call center staff (health plans) about requesting this information.

How to address potential patient or enrollee pushback respectfully.

How to address system-level issues, such as changes in patient registration screens and data flow.

Previous chapters have provided a framework for eliciting, categorizing, and coding data on race, ethnicity, and language need. This chapter considers strategies that can be applied by various entities to improve the collection of these data and facilitate subsequent reporting of stratified quality measures. It begins by examining current practices and issues related to collecting and sharing data across the health care system. Next is a discussion of steps that can be taken to address these issues and improve data collection processes. This is followed by a review of methods that can be used to derive race and ethnicity data through indirect estimation when obtaining data directly from many patients or enrollees is not possible.

Collecting and Sharing Data Across The Health Care System

Health care involves a diverse set of public and private data collection systems, including health surveys, administrative enrollment and billing records, and medical records, used by various entities, including hospitals, CHCs, physicians, and health plans. Data on race, ethnicity, and language are collected, to some extent, by all these entities, suggesting the potential of each to contribute information on patients or enrollees. The flow of data illustrated in Figure 5-1 does not even fully reflect the complexity of the relationships involved or the disparate data requests within the health care system. Currently, fragmentation of data flow occurs because of silos of data collection (NRC, 2009).

It should be noted that a substantial fraction of the U.S. population does not have a regular relationship with a provider who integrates their care (i.e., a medical home) (Beal et al., 2007). For some, a usual source of care is the emergency department (ED), a situation that complicates the capture and use of race, ethnicity, and language data and their integration with quality measurement. While health plans insure a large portion of the U.S. population, their direct contact tends to be minimal, even during enrollment. Hospitals, which tend to have more developed data collection systems, serve only a small fraction of the country's population. As a result, no one setting within the health care system can capture data on race, ethnicity, and language for every individual.

Health information technology (Health IT) may have the potential to improve the collection and exchange of self-reported race, ethnicity, and language data, as these data could be included, for example, in an individual's personal health record (PHR) and then utilized in electronic health record (EHR) and other data systems.1 There is little reliable evidence, though, on the adoption rates of EHRs (Jha et al., 2009). While substantial resources were devoted to this technology in the American Recovery and Reinvestment Act of 2009,2 it will take time to develop the infrastructure necessary to fully implement and support Health IT (Blumenthal, 2009). Thus, the consideration of other avenues of data collection and exchange is essential to the subcommittee's task.

Until data are better integrated across entities, some redundancy will remain in the collection of race, ethnicity, and language data from patients and enrollees, and equivalently stratified data will remain unavailable for comparison purposes unless entities adopt a nationally standardized approach. Methods should be considered for incorporating these data into currently operational data flows, with careful attention to concerns regarding efficiency and patient privacy.

**Hospitals**

Because hospitals tend to have information systems for data collection and reporting, staff who are used to collecting registration and admissions data, and an organizational culture that is familiar with the tools of quality improvement, they are relatively well positioned to collect patients' demographic data. In addition, hospitals have a history of collecting race data. With the passage of the Civil Rights Act of 19643 and Medicare legislation in 1965,4 there was a legislative mandate for equal access to and desegregation of hospitals (Reynolds, 1997). Therefore it is not surprising that more than 89 percent of hospitals report collecting race and ethnicity data, and 79 percent report collecting data on primary language (AHA, 2008).

This culture of data collection has limitations, however. Historically, the data were never intended for quality improvement purposes, but to allow analysis to ensure compliance with civil rights provisions. Additionally, hospital data collection practices are less than systematic as the categories collected vary by hospital, and hospitals obtain the information in various ways (e.g., self-report and observer report) (Regenstein and Sickler, 2006; Romano et al., 2003; Siegel et al., 2007). Furthermore, compared with the number of people who are insured or visit an ambulatory care provider, a relatively small number of people are hospitalized in any one year (Figure 5-2). Thus, while hospitals are an important component of the health care system and represent a major percentage of health care expenditures, they are only one element of the system for collecting and reporting race, ethnicity, and language data.

Hospitals also face challenges associated with collecting accurate data and using these data for quality improvement and reduction of disparities. A 2006 National Public Health and Hospitals Institute (NPHHI) survey asked hospitals that collected race and ethnicity data whether they used the data to assess and compare quality of care, utilization of health services, health outcomes, or patient satisfaction across their different patient populations. Fever than one in five hospitals that collected these data used them for any of these purposes (Regenstein and Sickle, 2006). Additionally, only half of hospitals that collected data on primary language maintained a database of patients' primary languages that they could track over time (Hasnain-Wynia et al., 2006).

Many of the above challenges can be attributed largely to the many staff and departments or units that need to be engaged in the process to ensure systematic data collection and use. Hospitals have multiple pathways (inpatient, outpatient, ED, urgent care) through which patients enter the system. For example, the ED is the source of 45 percent of all hospital admissions (Healthcare Financial Management Association, 2007).

Systems changes can involve training a large number (possibly hundreds) of hospital registration/admission staff (many of whom may be off site) and modifying practice management and EHR systems to ensure that proper and consistent data fields are in place across multiple departments and units that serve as patient entry points. Ideally, these systems would be made interoperable through the development of interfaces that would make it possible to relay the data across different systems.

A Robert Wood Johnson Foundation initiative to reduce disparities in cardiac care required participating hospitals to systematically collect race, ethnicity, and language data and use the data to stratify quality measures. The ten hospitals in the collaborative initially cited the data collection requirement as one of the greatest challenges of the program, yet once they focused their efforts on these goals, they were able to bring together key stakeholders within each institution, implement needed IT changes, and train staff. As a result, they successfully began data collection within a relatively short time (Siegel et al., 2008). Other hospitals not part of this initiative are also successfully collecting race, ethnicity, and language data and linking them to quality measures (Weinick et al., 2008). Data collected at the hospital level are useful both for assessing the quality of hospital-provided services and, if shared with other entities, for facilitating analyses of quality across multiple settings. Box 5-1 provides an example of a statewide initiative to collect standardized race, ethnicity, and language data.

**Community Health Centers:**

CHSs are front-line providers of care for underserved and disadvantaged groups (Taylor, 2004) and therefore are good settings for implementing quality improvement strategies aimed at reducing racial and ethnic disparities in care. Yet while CHCs serve diverse patient populations and, as organizations, understand the importance of demographic data for improving the quality of care, the accuracy of the race, ethnicity, and language data they collect may be limited (Maizlish and Herrera, 2006). More than 87 percent of surveyed CHCs reported inquiring about a patient's need for language services, and 73 percent reported recording this information in the patient record (Gallegos et al., 2008); less is known, however, about the extent to which CHCs consistently collect patient race and ethnicity data beyond the basic Office of Management and Budget (OMB) categories included in their national Uniform Data System (HRSA, 2009).

**3. Methodology**

**RESEARCH METHODOLOGY**

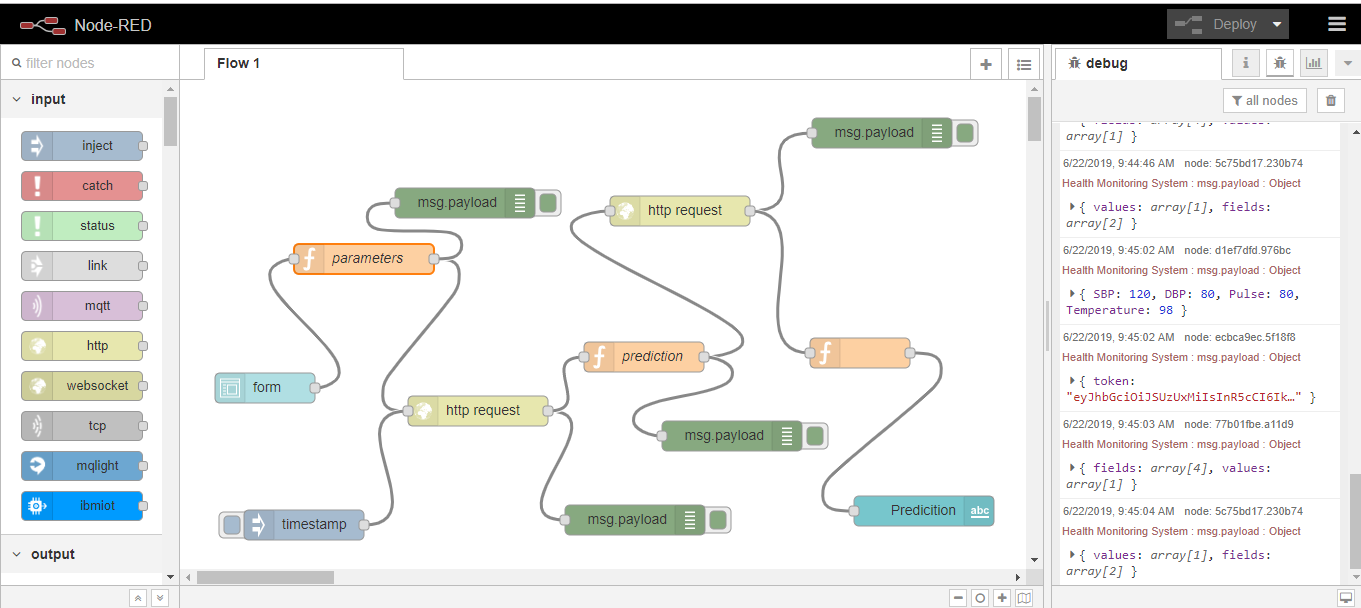
**Description:**

In this project a simple machine learning statistical techniques and algorithms are choosen.

Firstly we use the help of anaconda navigator as platform by using python as code.

Design is to predict the Health condition of the patient .

We Deploy the code to pedict paricular output in using IBM Cloud



**3.1 Exploratory Data Analysis**

Harmonizing diverse kinds of knowledge, diverse formats for knowledge, and their diverse modes of processing , via a universal framework for the representation and processing of knowledge.

**THE DISCOVERY OF STRUCTURE IN DATA:**

If we are trying to discover patterns of association or other structures in big data a diversity of formalisms and formats is a handicap.

**THE INTERPRETATION OF DATA:**

If we are trying to recognize objects in images, do scene analysis, or otherwise interpret what the images mean, it would make things simpler if we did not have to deal with the diversity of formats for images mentioned earlier. Likewise for other kinds of data. Big Data analytics has strengths in areas such as pattern recognition, information retrieval, parsing and production of natural language, translation from one representation to another, several kinds of reasoning, planning and problem solving.

**THE UNDERSTANDING AND TRANSLATION OF NATURAL LANGUAGES :**

In our everyday use of natural languages we recognize that meanings are different from the words that express them and that, very often, two or more distinct sequences of words may mean the same thing or have the same referent: ``the capital of the United States'' means the same as ``Washington D. C.''; ``Ursus maritimus'' means the same as ``polar bear''; and so on. These intuitions corroborate the need for a UFK, or something like it, which is independent of the words in any natural language. 604 M. Srivathsan and K. yogesh Arjun / Procedia Computer Science 50 ( 2015 ) 602 – 609 .

**PATTERNS AND PATTERN RECOGNITION :**

The Prognitive Computer analyses all the data of diseases and groups into sectors and manages each sector in terms of tree structures and graphs .This similar to that of done in the brain. Multiple sources of information in the form of pictures(.TIFF,.JPEG,.PNG) or Reports(.DOCX,PDF) are gathered from networks of Hospitals ,Pharmaceutical companies ,Personal records and a link is created between each related node forming a directional graph. This contains all thorough information of single condition as separate data .

**4.2. STATISTICAL METHODS**

**1. Problem Framing**

Perhaps the point of biggest leverage in a predictive modeling problem is the framing of the problem.

This is the selection of the type of problem, e.g. regression or classification, and perhaps the structure and types of the inputs and outputs for the problem.

The framing of the problem is not always obvious. For newcomers to a domain, it may require significant exploration of the observations in the domain.

For domain experts that may be stuck seeing the issues from a conventional perspective, they too may benefit from considering the data from multiple perspectives.

Statistical methods that can aid in the exploration of the data during the framing of a problem include:

**2. Data Understanding**

Data understanding means having an intimate grasp of both the distributions of variables and the relationships between variables.

Some of this knowledge may come from domain expertise, or require domain expertise in order to interpret. Nevertheless, both experts and novices to a field of study will benefit from actually handling real observations form the domain.

Two large branches of statistical methods are used to aid in understanding data; they are:

**3. Data Cleaning**

Observations from a domain are often not pristine.

Although the data is digital, it may be subjected to processes that can damage the fidelity of the data, and in turn any downstream processes or models that make use of the data.

Some examples include:

* Data corruption.
* Data errors.
* Data loss.

The process of identifying and repairing issues with the data is called data cleaning

Statistical methods are used for data cleaning; for example:

* **Outlier detection**. Methods for identifying observations that are far from the expected value in a distribution.
* **Imputation**. Methods for repairing or filling in corrupt or missing values in observations.

**4. Data Selection**

Not all observations or all variables may be relevant when modeling.

The process of reducing the scope of data to those elements that are most useful for making predictions is called data selection.

Two types of statistical methods that are used for data selection include:

* **Data Sample**. Methods to systematically create smaller representative samples from larger datasets.
* **Feature Selection**. Methods to automatically identify those variables that are most relevant to the outcome variable.

**5. Data Preparation**

Data can often not be used directly for modeling.

Some transformation is often required in order to change the shape or structure of the data to make it more suitable for the chosen framing of the problem or learning algorithms.

Data preparation is performed using statistical methods. Some common examples include:

* **Scaling**. Methods such as standardization and normalization.
* **Encoding**. Methods such as integer encoding and one hot encoding.
* **Transforms**. Methods such as power transforms like the Box-Cox method.

**6. Model Evaluation**

A crucial part of a predictive modeling problem is evaluating a learning method.

This often requires the estimation of the skill of the model when making predictions on data not seen during the training of the model.

Generally, the planning of this process of training and evaluating a predictive model is called experimental design. This is a whole subfield of statistical methods.

* **Experimental Design**. Methods to design systematic experiments to compare the effect of independent variables on an outcome, such as the choice of a machine learning algorithm on prediction accuracy.

As part of implementing an experimental design, methods are used to resample a dataset in order to make economic use of available data in order to estimate the skill of the model. These two represent a subfield of statistical methods.

* **Resampling Methods**. Methods for systematically splitting a dataset into subsets for the purposes of training and evaluating a predictive model.

**7. Model Configuration**

A given machine learning algorithm often has a suite of hype-rparameters that allow the learning method to be tailored to a specific problem.

The configuration of the hype-rparameters is often empirical in nature, rather than analytical, requiring large suites of experiments in order to evaluate the effect of different hyper-parameter values on the skill of the model.

The interpretation and comparison of the results between different hyperparameter configurations is made using one of two subfields of statistics, namely:

* an assumption or expectation about the result (presented using critical values and p-values).
* **Estimation Statistics**. Methods that quantify the uncertainty of a result using confidence intervals.

**8. Model Selection**

One among many machine learning algorithms may be appropriate for a given predictive modeling problem.

The process of selecting one method as the solution is called model selection.

This may involve a suite of criteria both from stakeholders in the project and the careful interpretation of the estimated skill of the methods evaluated for the problem.

As with model configuration, two classes of statistical methods can be used to interpret the estimated skill of different models for the purposes of model selection. They are:

* **Statistical Hypothesis Tests**. Methods that quantify the likelihood of observing the result given an assumption or expectation about the result (presented using critical values and p-values).
* **Estimation Statistics**. Methods that quantify the uncertainty of a result using confidence intervals.

**9. Model Presentation**

Once a final model has been trained, it can be presented to stakeholders prior to being used or deployed to make actual predictions on real data.

A part of presenting a final model involves presenting the estimated skill of the model.

Methods from the field of estimation statistics can be used to quantify the uncertainty in the estimated skill of the machine learning model through the use of tolerance intervals and confidence intervals.

**10. Model Predictions**

Finally, it will come time to start using a final model to make predictions for new data where we do not know the real outcome.

As part of making predictions, it is important to quantify the confidence of the prediction.

Just like with the process of model presentation, we can use methods from the field of estimation statistics to quantify this uncertainty, such as confidence intervals and prediction intervals.

**DATA VISUALIZATION**

**Introduction to Data Visualization Methods in Python**

Sometimes data does not make sense until you can look at in a visual form, such as with charts and plots.

Being able to quickly visualize your data samples for yourself and others is an important skill both in applied statistics and in applied machine learning.

**Data Visualization**

Data visualization is an important skill in applied statistics and machine learning.

Statistics does indeed focus on quantitative descriptions and estimations of data. Data visualization provides an important suite of tools for gaining a qualitative understanding.

This can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more. With a little domain knowledge, data visualizations can be used to express and demonstrate key relationships in plots and charts that are more visceral to yourself and stakeholders than measures of association or significance.

Data visualization and exploratory data analysis are whole fields themselves and I will recommend a deeper dive into some the books mentioned at the end. In this tutorial, let’s look at basic charts and plots you can use to better understand your data.

There are five key plots that you need to know well for basic data visualization. They are:

* Line Plot
* Bar Chart
* Histogram Plot
* Box and Whisker Plot
* Scatter Plot

With a knowledge of these plots, you can quickly get a qualitative understanding of most data that you come across.

For the rest of this tutorial, we will take a closer look at each plot type.

**Introduction to Matplotlib**

There are many excellent plotting libraries in Python and I recommend exploring them in order to create presentable graphics.

For quick and dirty plots intended for your own use, I recommend using the matplotlib library. It is the foundation for many other plotting libraries and plotting support in higher-level libraries such as Pandas.

The matplotlib provides a context, one in which one or more plots can be drawn before the image is shown or saved to file. The context can be accessed via functions on *pyplot* The context can be imported as follows:



|  |  |
| --- | --- |
| 1 | from matplotlib import pyplot |

There is some convention to import this context and name it *plt*; for example:



|  |  |
| --- | --- |
| 1 | import matplotlib.pyplot as plt |

We will not use this convention, instead we will stick to the standard Python import convention.

Charts and plots are made by making and calling on context; for example:



|  |  |
| --- | --- |
| 1 | pyplot.plot(...) |

Elements such as axis, labels, legends, and so on can be accessed and configured on this context as separate function calls.

The drawings on the context can be shown in a new window by calling the [show() function](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.show.html):



|  |  |
| --- | --- |
| 1  2 | # display the plot  pyplot.show() |

Alternately, the drawings on the context can be saved to file, such as a PNG formatted image file. The [savefig() function](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.savefig.html) can be used to save images.



|  |  |
| --- | --- |
| 1 | pyplot.savefig('my\_image.png') |

This is the most basic crash course for using the matplotlib library.

For more detail, see the [User Guide](https://matplotlib.org/users/index.html) and the resources at the end of the tutorial.

**Line Plot**

A line plot is generally used to present observations collected at regular intervals.

The x-axis represents the regular interval, such as time. The y-axis shows the observations, ordered by the x-axis and connected by a line.

A line plot can be created by calling the [plot() function](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html) and passing the x-axis data for the regular interval, and y-axis for the observations.



|  |  |
| --- | --- |
| 1  2 | # create line plot  pyplot.plot(x, y) |

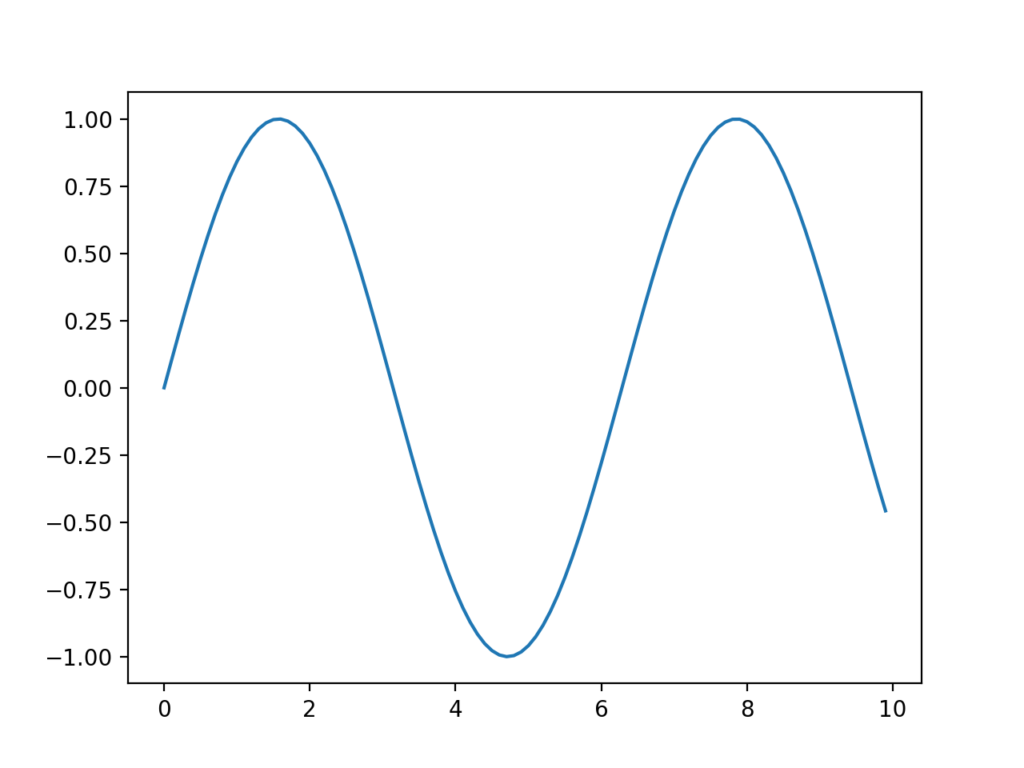
Line plots are useful for presenting time series data as well as any sequence data where there is an ordering between observations.

The example below creates a sequence of 100 floating point values as the x-axis and a sine wave as a function of the x-axis as the observations on the y-axis. The results are plotted as a line plot.



|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11 | # example of a line plot  from numpy import sin  from matplotlib import pyplot  # consistent interval for x-axis  x = [x\*0.1 for x in range(100)]  # function of x for y-axis  y = sin(x)  # create line plot  pyplot.plot(x, y)  # show line plot  pyplot.show() |

Running the example creates a line plot showing the familiar sine wave pattern on the y-axis across the x-axis with a consistent interval between observations.



Example of a Line Plot

**Bar Chart**

A bar chart is generally used to present relative quantities for multiple categories.

The x-axis represents the categories and are spaced evenly. The y-axis represents the quantity for each category and is drawn as a bar from the baseline to the appropriate level on the y-axis.

A bar chart can be created by calling the [bar() function](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.bar.html) and passing the category names for the x-axis and the quantities for the y-axis.



|  |  |
| --- | --- |
| 1  2 | # create bar chart  pyplot.bar(x, y) |

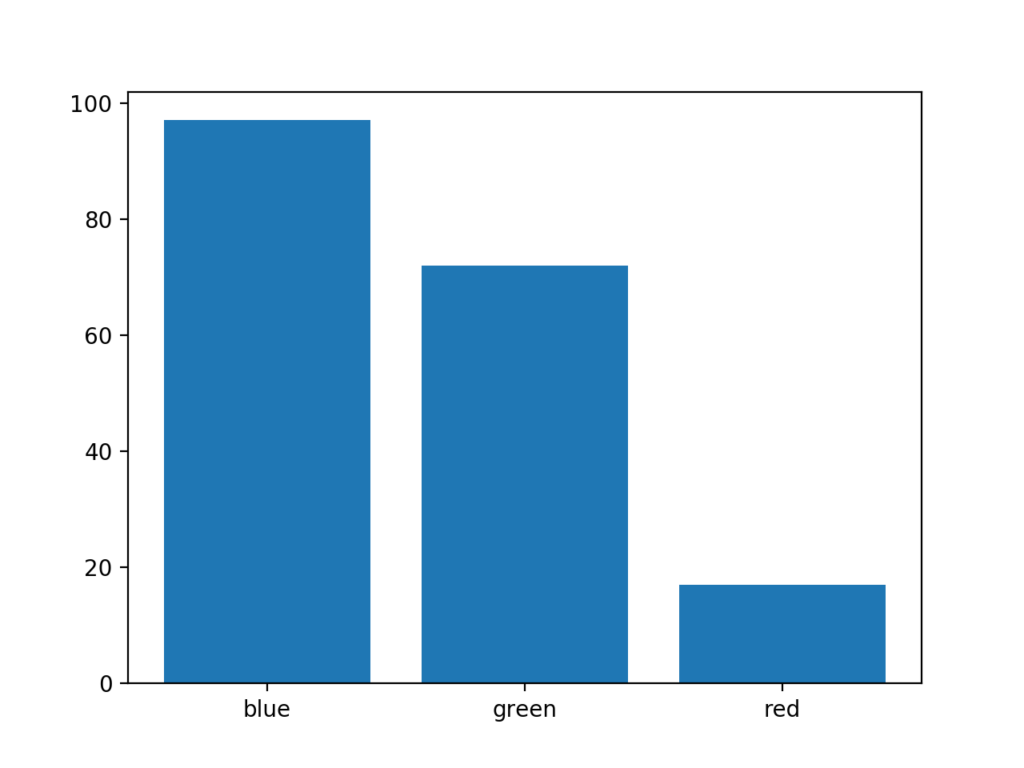
Bar charts can be useful for comparing multiple point quantities or estimations.

The example below creates a dataset with three categories, each defined with a string label. A single [random integer value](https://machinelearningmastery.com/how-to-generate-random-numbers-in-python/) is drawn for the quantity in each category.



|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14 | # example of a bar chart  from random import seed  from random import randint  from matplotlib import pyplot  # seed the random number generator  seed(1)  # names for categories  x = ['red', 'green', 'blue']  # quantities for each category  y = [randint(0, 100), randint(0, 100), randint(0, 100)]  # create bar chart  pyplot.bar(x, y)  # show line plot  pyplot.show() |

Running the example creates the bar chart showing the category labels on the x-axis and the quantities on the y-axis.



**4.3. Data modeling using supervised ML techniques**

**Supervised Learning**

In Supervised Learning, algorithms learn from labeled data. After understanding the data, the algorithm determines which label should be given to new data based on pattern and associating the patterns to the unlabeled new data.

Supervised Learning can be divided into 2 categories i.e Classification & Regression

**Classification predicts the category that data belongs to.**

eg: Spam Detection, Churn Prediction, Sentiment Analysis, Dog Breed Detection.

**Regression predicts a numerical value based on previous observed data.**

eg: House Price Prediction, Stock Price Prediction, Height-Weight Prediction.

**Classification**

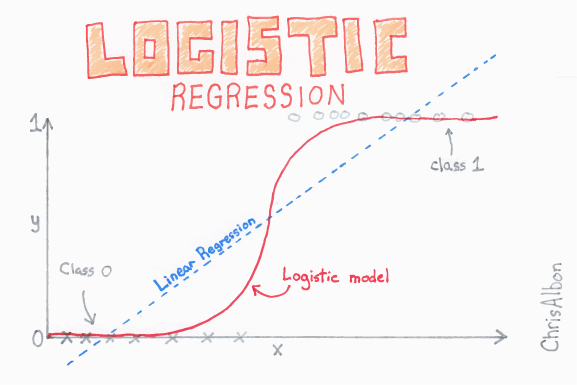
Classification is a technique for determining class the dependent belongs to the one or more independent variables.

Classification is used for predicting discrete responses.



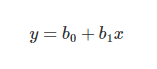
**Logistic Regression:**

Logistic regression is kind of like linear regression but is used when the dependent variable is not a number, but something else (like a Yes/No response). Its called Regression but performs classification as based on the regression it classifies the dependent variable into either of the classes.



Logistic Regression

Logistic regression is used for prediction of output which is binary, as stated above. For example, if a credit card company is going to build a model to decide whether to issue a credit card to a customer or not, it will model for whether the customer is going to “Default” or “Not Default” on this credit card.



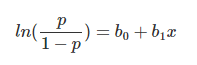
Linear Regression

Firstly, Linear Regression is performed on the relationship between variables to get the model .The threshold for the classification line is assumed to be at 0.5

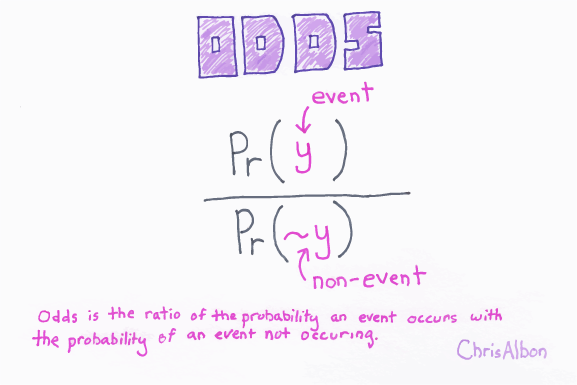
https://cdn-images-1.medium.com/max/1600/1*pwelaJ8-sZynXVDMON1Etw.png

Logistic Sigmoid Function

Logistic Function is applied to the regression to get the probabilities of it belonging in either class.



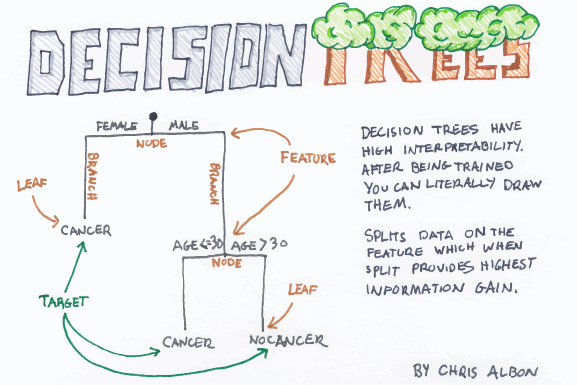
**odds**



It gives the log of the probability of the event occurring to log of the probability of it not occurring. In the end, it classifies the variable based on the higher probability of either class.

**Decision Tree Classification**

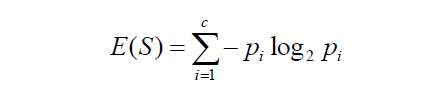
Decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with **decision nodes** and **leaf nodes**. It follows Iterative Dichotomiser 3(ID3) algorithm structure for determining the split.



Decision Tree uses *Entropy* and *Information Gain* to construct a decision tree.

**Entropy**

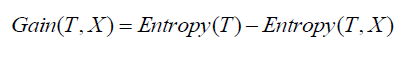
Entropy is the degree or amount of uncertainty in the randomness of elements or in other words it is a measure of impurity***.***



Intuitively, it tells us about the predictability of a certain event. Entropy calculates the homogeneity of a sample. If the sample is completely homogeneous the entropy is zero and if the sample is an equally divided it has an entropy of one.

[**Information Gain**](https://medium.com/@rishabhjain_22692/decision-trees-it-begins-here-93ff54ef134)

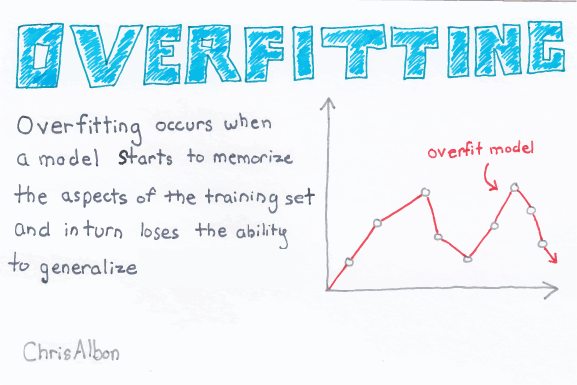
It measures the relative change in entropy with respect to the independent attribute. It tries to estimate the information contained by each attribute. Constructing a decision tree is all about finding the attribute that returns the highest information gain (i.e., the most homogeneous branches).



Where Gain(T, X) is the information gain by applying feature X. Entropy(T) is the Entropy of the entire set, while the second term calculates the Entropy after applying the feature X.

Information Gain ranks attribute for filtering at a given node in the tree. The ranking is based on the highest information gain entropy in each split.

The disadvantage of a Decision Tree Model is overfitting as it tries to fit the model by going deeper in the training set and thereby reducing test accuracy.

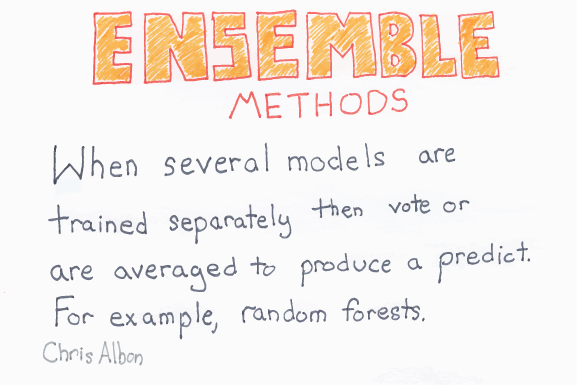


Over fitting

Over fitting in Decision Trees can be minimized by pruning nodes.

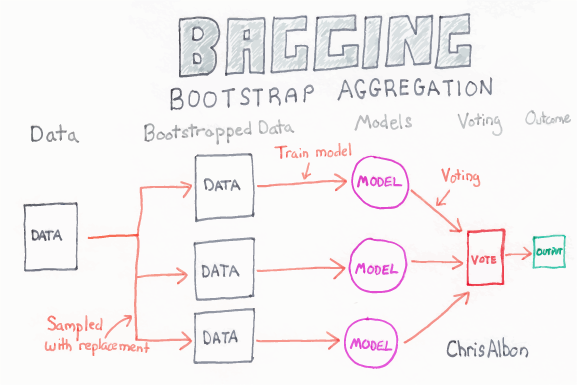
**Ensemble Methods for Classification**

An ensemble model is a *team of models*. Technically, ensemble models comprise of several supervised learning models that are individually trained, and the results are merged in various ways to achieve the final prediction. This result has higher predictive power than the results of any of its constituting learning algorithms independently.



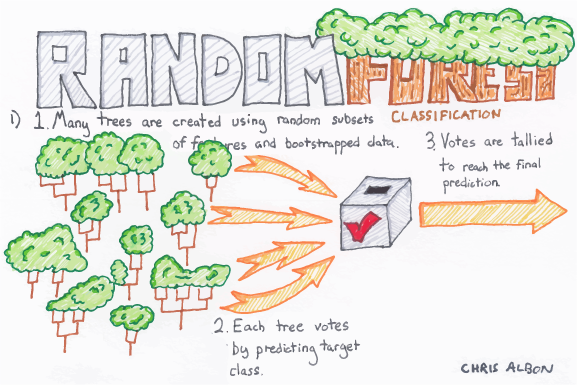
**Random Forest Classification**

Random Forest Classifier is an ensemble algorithm based on bagging i.e bootstrap aggregation. Ensemble methodscombines more than one algorithms of the same or different kind for classifying objects i.e an ensemble of SVM, Naive Bayes or Decision Trees.



Bagging

The general idea is that a combination of learning models that increases the overall result is selected.



Random Forest Classification

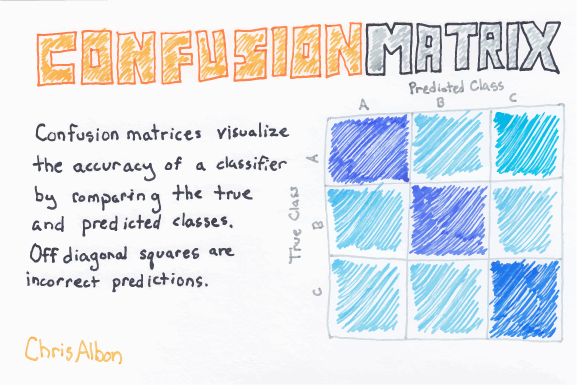
Deep decision trees may suffer from over fitting, but random forests prevent from over fitting by creating trees on random subsets. The main reason is that it takes the average of all the predictions, which cancels out the biases.

Random Forest adds additional randomness to the model while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model.

**Classification Model Performances**

**1.Confusion Matrix**

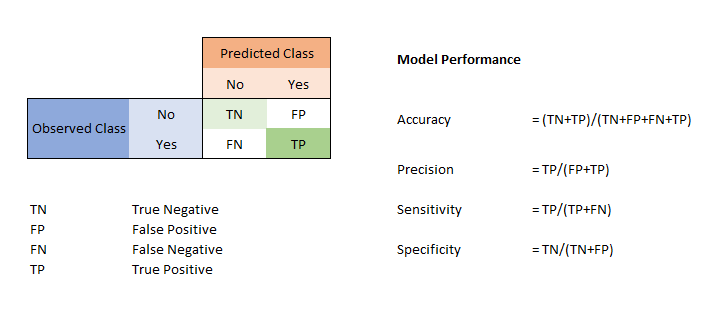
A confusion matrix is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known. It is a table with 4 different combinations of predicted and actual values in the case for a binary classifier.



General Multiclass Confusion Matrix

The confusion matrix for a multi-class classification problem can help you determine mistake patterns.

For a Binary Classifier,



Binary Confusion Matrix

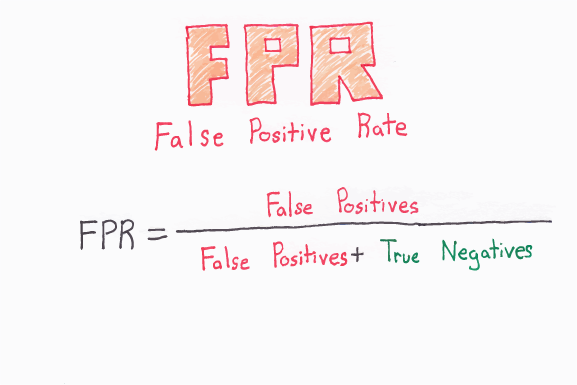
A **true positive** is an outcome where the model *correctly* predicts the *positive* class. Similarly, a **true negative** is an outcome where the model *correctly* predicts the *negative* class.

The terms False Positive and False Negative are very in determining how well the model is predicted with respect to classification .A **false positive** is an outcome where the model *incorrectly* predicts the *positive* class. And a **false negative** is an outcome where the model *incorrectly* predicts the *negative* class. The more values in main diagonal, better the model whereas the other diagonal gives the worst result for classification.

**False Positive**

An example in which the model mistakenly predicted the [**positive class**](https://developers.google.com/machine-learning/glossary/#positive_class). For example, the model inferred that a particular email message was spam (the positive class), but that email message was actually not spam. It’s like a warning sign that the mistake did should be rectified as it’s not much of a serious concern compared to False Negative.

**False positive***(type I error)*— when you reject a *true* null hypothesis

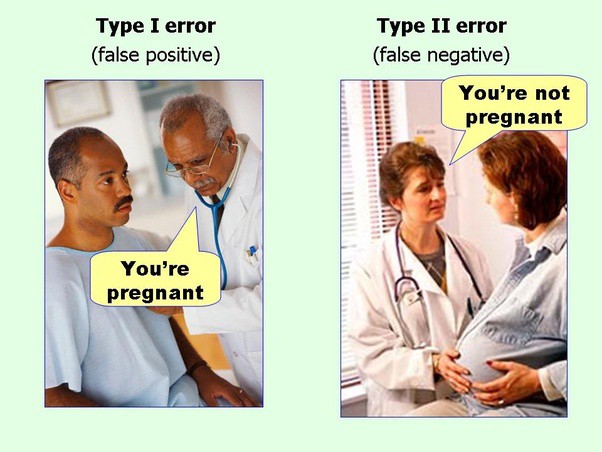


False Positive Rate

**False Negative**

An example in which the model mistakenly predicted the [**negative class**](https://developers.google.com/machine-learning/glossary/#negative_class). For example, the model inferred that a particular email message was not spam (the negative class), but that email message actually was spam. It’s like a **danger**sign that the mistake did should be rectified at the earliest as it’s of a much serious concern compared to False Positive.

**False negative***(type II error) —*when you accept a *false* null hypothesis.

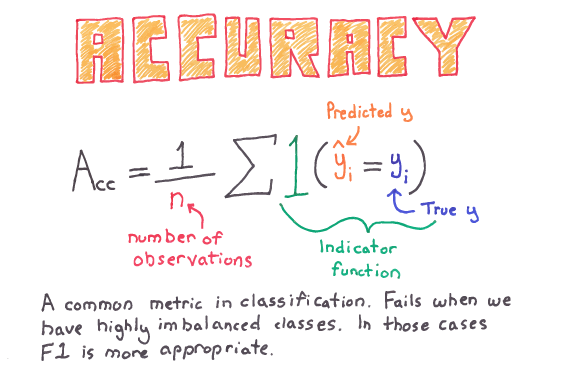


This picture easily illustrates the above metrics . The man’s test results say “You ’re pregnant” is False Positive as a man cannot be pregnant, and a pregnant woman’s test results say “Pregnant” is False Negative as from the image it easily identified that the woman is pregnant.

From the Confusion Matrix, we can infer Accuracy, Precision, Recall, F-1 Score.

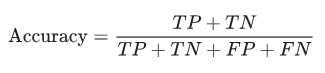
**Accuracy**

Accuracy is the fraction of predictions our model got right.



Accuracy Term

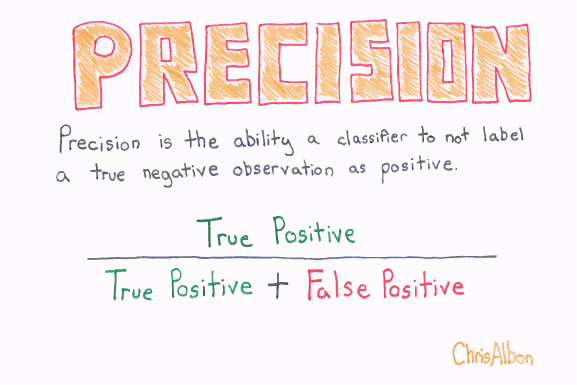
Accuracy can also be written as



Accuracy alone doesn’t tell the full story when you’re working with a class-imbalanced data set, where there is a significant disparity between the number of positive and negative labels. Precision and Recall are better metrics for evaluating class-imbalanced problems.

**Precision**

Out of all the classes, how much we predicted correctly.



Precision should be as high as possible.

**5. FINDINGS AND SUGGESTIONS**

Information technology offers many potential benefits to health care. Electronic medical records (EMRs) facilitate cost-effective access to more complete, accurate health data with which providers can make better decisions about patient care. Advanced communications networks can enable the sharing of data among distributed elements of integrated health care delivery systems and can enable telemedicine programs to overcome geographic boundaries between patients and providers. Electronic data processing techniques can enable managed care providers, health services researchers, and public and private oversight organizations to conduct more sophisticated analyses of health care utilization and outcomes. Electronic billing and administration systems may help reduce the administrative costs of health care. Computer-based decision support tools can help reduce variation in health care quality across providers, improve adherence to standards of care, and reduce costs by eliminating duplicative or non-efficacious tests and therapeutic procedures.

To obtain the benefits of electronic medical records, the nation must address and mitigate concerns regarding the privacy and security of electronic health care information. As the recommendations in this chapter describe, health care providers have to adopt a range of technical and organizational practices to protect health care information, and the health care industry will have to work with government to create a legal framework and proper set of incentives for heightening interest in privacy and security and for ensuring industry-wide protection of health information.

**Finding 1: Information technology is becoming increasingly important in improving the quality and lowering the costs of health care; attempts to protect patient privacy must therefore center on finding ways to protect sensitive electronic health information in a computerized environment rather than on opposing the use of information technology in health care organizations.** As the site visits conducted for this study attest, the shift to integrated health care delivery systems and managed care creates a growing demand for electronic health information and for data networks capable of transferring data within and across organizations. Electronic health information allows such organizations to better analyze data for such purposes as improving care, monitoring the quality of care, analyzing the utilization of health care resources, and managing health benefits. Care providers claim that the availability of health information on-line helps them enhance the quality of health care delivery, as well as its efficiency. Patients will see the advantages of integrating and sharing data across the institution as they begin to receive a greater proportion of their care within integrated delivery systems. The application of information technology to health care is expected to help reduce the cost of administering care.

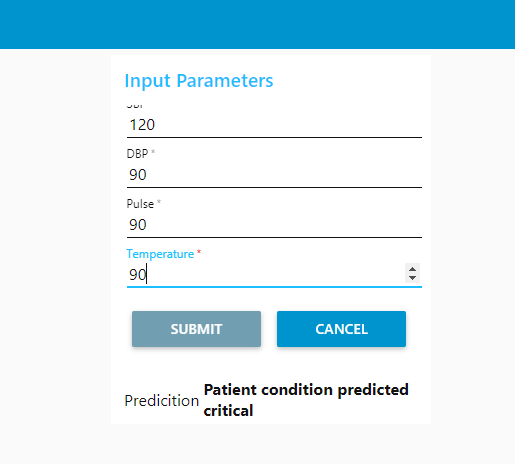
Each of the organizations visited as part of this study has ongoing programs to expand the use of information technology for clinical care and administration; all reported positive benefits of such applications. As long as health care organizations continue to find value in these activities, whether by improving the quality or reducing the costs of care, strong incentives will exist to pursue them. Thus, although opposition to the use of electronic medical records may succeed in delaying their widespread adoption, in the long run expectations of enhanced quality and improved efficiency, combined with economic pressures, are likely to dominate. From a policy perspective, it therefore makes far more sense for the health care system to find ways to handle legitimate privacy and security concerns without foregoing the benefits of information technology.

Furthermore, properly implemented EMRs offer great potential for improving the security of health information and the privacy of patients. EMRs allow the use of technical mechanisms to either impede unauthorized access or deter potential abuses. For example, authentication and access control technologies can help ensure that access to health information

**Finding 2: Health care organizations need to take a more aggressive approach to improving the security of health information systems in order to better protect electronic health information.** Little is known about the extent of existing violations of privacy and security in the health care industry. Although some sites were aware of some cases in which authorized users had intentionally or unintentionally released health information inappropriately (from both electronic and paper record systems), the sites visited as part of this study reported no incidents in which outside attackers breached system security and produced large-scale violations of patient privacy. Most health care organizations therefore continue to perceive insider abuse as the primary problem to be solved; however, evidence from other industries indicates that organizations with Internet connections or other kinds of remote access (e.g., modem connections) are prone to outsider attacks.[1](https://www.nap.edu/read/5595/chapter/8#p200063478960162001) As health care organizations put more information on-line and begin to transmit patient information electronically, they will have to ensure that adequate security protections have been developed to protect against new vulnerabilities.

**6. CONCLUSION**

Health care is moving into the home increasingly often and involving a mixture of people, a variety of tasks, and a broad diversity of devices and technologies; it is also occurring in a range of residential environments.



Here we can see the patients condition is “**critical”** by giving the input parameters such as:

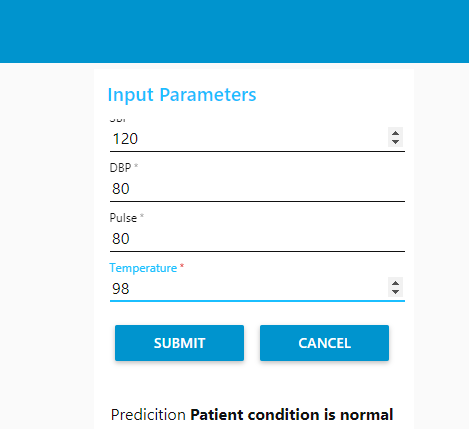
1.Sbp

2.Dbp

3.Pulse

4.Temperature

And predict the 5. Level of health condition



Here we can see the patients condition is “**Normal” ”** by giving the input parameters such as:

1.Sbp

2.Dbp

3.Pulse

4.Temperature

And predict the 5. Level of health condition using **“NODE-RED”** as a tool for Flow-Based Programming.