VSIT | Vidyalankar School of Information Technology

V-Tech

Technical Magazine of Information Technology Department

Academic year
2019 - 20
Issue - II

Preface

Prof. Makarand Deshpande

I am pleased to present the second issue of V-Tech, the technical magazine by the Department of Information Technology of VSIT, for academic year 2019-20. Teachers contribute articles in their field of interest as well as current/upcoming areas which help in expanding the knowledge base of faculty members.

Continuing with this tradition, this issue deals with different aspects of IT field. Starting from journey of established technologies like Data Mining, Artificial Intelligence, and the way forward, to current trending topics in the field such as Machine Learning and artificial intelligence, Cryptocurrency, Augmented and virtual reality. This issue also talks about some of the advances in IOT and embedded technologies like Holter Monitor, robotics. Last but not the least; it also covers the COVID pandemic and challenges faced in academia, the concept of work from home.

I hope you will find this issue as interesting as I did. It will help all the readers in enriching their IT knowledge and hopefully strike a chord in at least one area where they can take a deep dive for their research activities.

Prof. Makarand Deshpande

Adjunct Faculty

Index

Augmented Reality in Real Estate

The speedy development of Augmented Reality(AR) technology is giving unique opportunities to consumers for making their life much easier and better. Augmented Reality (AR) is one of the biggest technology trends at present. AR ready smartphones and other devices become more accessible around the world. AR let us see the real-life environment right in front of us, for example trees swaying in the park, dogs chasing balls, kids playing soccer—with a digital augmentation overlapped on it.

With advances in AR technology, these examples are not that different from what might already be available for your smartphone. Augmented reality is, in fact, readily available and being used in many ways including as Snapchat lenses, in apps that help you find your car in a crowded parking lot, and in variety of shopping apps that let you try on clothes without even leaving home.

We find AR everywhere such as AR in health sector, Sports, manufacturing, security, travel. And, the same stands true for the real estate development. We all are aware of the technology's impact on the real estate industry. As most of the businesses today are getting benefited from the AR technology, it is significant to understand how Augmented Reality is transforming the real estate industry. Let us understand what is AR?

- It can be described as a 3D environment created using a computer which can also interact with User.
- It gives full view of an existing environment through video, audio, graphics & GPS Data.

Let us know the advantages of AR in Real estate:

- It creates a clear picture of the property whether its existing or not:
    - This means that virtual reality gives a clear picture of how the existing property looks like or the property which will be built in future would look like.
- Grabs customer's attention:
    - Who doesn't want to see the property by siting inside AC office or from sitting at home. This facility grabs customer's attention as the Customer get to see the actual picture of property.
- Best marketing tool as well as cost effective:
    - It is considered as the best marketing tool as Customer can get the glimpse of the property sitting inside the office.
- Using AR embedded mobile apps makes it easy to reach target audience:
    - With the advent of Mobile application in Augmented Reality has helped to reach to more customers or target audience.
- It saves time & cost effective:

- It saves time as the customer doesn't have to visit physically to the site. As with Augmented Reality the process becomes cheap and cost effective.

- Better engagement
  - This process engages customers as well as the estate agents or whoever involved in the process of buying and selling property.

Image Ref. Quepplin.com

How AR in Real Estate has helped both, the customer and the seller of the property?

1. Customer's Side:
   a. When there was no augmented reality in real estate, there was a lack of perfection in everything.
   b. For instance, if people were buying a plot, they had to go through fuzziness and they could only imagine what they can do with the plot.
   c. But, now with augmented reality technology, they can visualize everything. So it is making every buying process true happiness.
   d. Apart from the buyers' experience, it also refurbishes the agents' experience by transforming the selling and marketing efforts.
   e. Prospective customers of the builders can now see how the building will look after the construction is complete.
   f. The customers can have a virtual 3D walk in each flat and view each room and interior using Mockup Visualization.
2. Seller's Side:
   a. Augmented reality in the construction industry helps to visualize each and every element especially those which are hidden.
   b. For instance, engineers and construction workers can view the plan and blueprint of the city electricity and sewage pipes plan.
   c. Also, construction workers can get the detail plans for the electric wiring of a home or room that they are refurbishing.
   d. This would save time and effort by avoiding unnecessary drilling.

Augmented Reality provides great value at the early stages of a construction project, to the final selling and marketing tools when selling a property. This is a new way to engage and connect with its digital-savvy target audience. And lets your buyer know the real size of floor, open/constructed area and let them have the convenience of viewing and experiencing the property in 3D.

In this way we can conclude that, augmented reality is one of those sweeping positive changes that have the potential to revolutionize how the real estate industry operates, and we expect to see many more benefits from this technology in the future. Augmented reality and real estate have the potential to drive the growth together for the sector by assisting customers better to arrive at a conclusive decision.

Madhavi Amondkar
Asst. Professor

Blackle

Blackle is a website powered by Google Custom Search and created by Tony Heap of Heap Media, which aims to save energy by displaying a black background and using greyish-white font colour for search results. The concept behind Blackle is that computer monitors can be made to use less energy by displaying much darker colours. Blackle is based on a study which tested a variety of CRT and LCD monitors.

Blackle appeared taking into account the idea that when a PC screen is white, exhibiting an unfilled word or the Google home, your PC expends 74W. At the point when the screen is dark it devours just 59W.Based on this hypothesis if everybody changed from Google to Blackle, it would save750MW every year. This was a better than average usage of Green Computing.

Advantages of using Blackle are:

1. It is Energy efficient.
2. User can set Blackle.com as your home page, that way every time user logs into the computer, he can save a little bit of energy every time.
3. Secure Privacy Policy.

Blackle helps in saving energy and being more environmentally friendly.

Akshatha Jain

Assistant Professor

Clustering and its Applications in Life Science

CLUSTERING – a technique of Big Data Analytics
Clustering involves grouping items based on certain similarity measures, such that objects in a group are very similar and differ noticeably from those in other groups.

Clustering is the use of unsupervised techniques for grouping similar objects. In machine learning, unsupervised refers to the problem of finding hidden structure within unlabeled data. Clustering techniques are unsupervised in the sense that the data scientist does not determine, in advance, the labels to apply to the clusters. The structure of the data describes the objects of interest and determines how best to group the objects.

Some examples are :

- given news articles, cluster into different types of news
- given a set of tweets, cluster based on content of tweet
- given a set of images, cluster them into different objects
- discovery of patterns.

THE K-MEANS ALGORITHM

To illustrate the method to find k clusters from a collection of M objects with n attributes, the two-dimensional case (n = 2) is examined The k-means algorithm to find k clusters can be described in the following four steps.

1. Choose the value of k and the k initial guesses for the centroids.

   For example, k = 3, and the initial centroids are indicated by the points shaded in red, green, and blue in Figure 1-1.

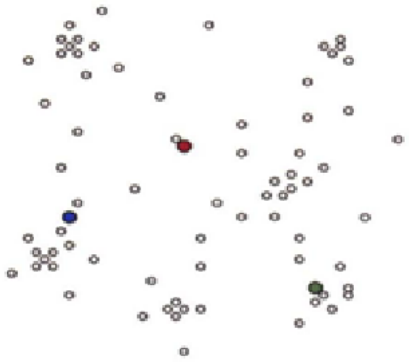


Fig 1-1 : initial starting points for the centroids

2. Compute the distance from each data point (x, y) to each centroid. Assign each point to the closest centroid.

   This association defines the first k clusters.

   In two dimensions, the distance, d, between any two points, (x_1, y_1) and (x_2, y_2) in the Cartesian plane is typically expressed by using the Euclidean distance measure provided in Equation 1-1.

d= $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ ……. 4-1.

In Figure 1-2, the points closest to a centroid are shaded the corresponding color.

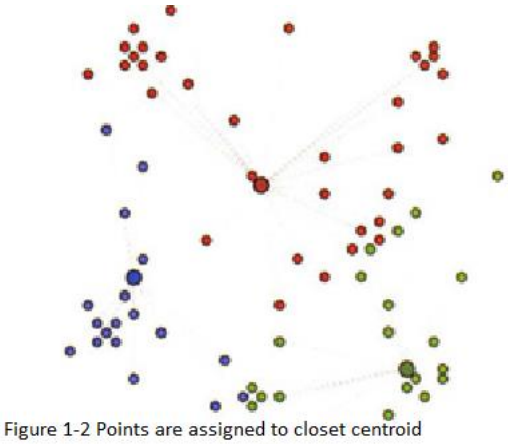


Figure 1-2 Points are assigned to closet centroid

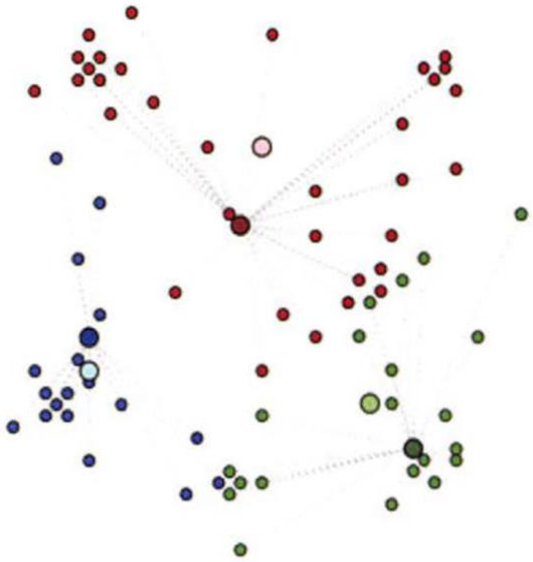


Figure 1-3 Compute the mean of each cluster

3. Compute the centroid, the center of mass, of each newly defined cluster from Step 2.

In Figure 1-3, the computed centroids in Step 3 are the lightly shaded points of the corresponding color. In two dimensions, the centroid (x_c, y_c) of them points in a k-means cluster is calculated as follows in Equation 1-2.

$$(x_c, y_c) = \left(\frac{\sum_{i=1}^{m} x_i}{m}, \frac{\sum_{1=1}^{m} y_i}{m}\right) \quad \text{….. 1-2}$$

Thus, (x_c, y_c) is the ordered pair of the arithmetic means of the coordinates of them points in the cluster. In this step, a centroid is computed for each of the k clusters.

4. Repeat Steps 2 and 3 until the algorithm converges to an answer.

Convergence is reached when the computed centroids do not change, or the centroids and the assigned points oscillate back and forth from one iteration to the next. The latter case can occur when there are one or more points that are equal distances from the computed centroid**.**

APPLICATIONS OF CLUSTERING IN LIFE SCIENCE

In 2015, lung cancer datasets were analyzed to find out which type of dataset and algorithm would be best for analyzing lung cancer. K-Means and Farthest First algorithms were used for the analyses.
The K-Means algorithm was found to be efficient for clustering the lung cancer dataset with Attribute Relation File Format (ARFF).
Sirinukunwattana et al used the Gaussian Bayesian hierarchical clustering (GBHC) algorithm. They tested the algorithm over 11 cancer and 3 synthetic datasets. They realized that in

comparison to other clustering algorithms the GBHC produced more accurate clustering results medically confirmed.
Karmilasari et al. implemented K-means algorithm on images from the Mammography Image Analysis Society (MIAS) to determine the stage of malignant breast cancer.
Moore et al. identified five distinct clinical phenotypes of asthma using unsupervised hierarchical cluster analysis.
In 2015, lung cancer datasets were analyzed to find out which type of dataset and algorithm would be best for analyzing lung cancer. K-Means and Farthest First algorithms were used for the analyses.
The K-Means algorithm was found to be efficient for clustering the lung cancer dataset with Attribute Relation File Format (ARFF).
Sirinukunwattana et al used the Gaussian Bayesian hierarchical clustering (GBHC) algorithm. They tested the algorithm over 11 cancer and 3 synthetic datasets. They realized that in comparison to other clustering algorithms the GBHC produced more accurate clustering results medically confirmed.
Karmilasari et al. implemented K-means algorithm on images from the Mammography Image Analysis Society (MIAS) to determine the stage of malignant breast cancer.
Moore et al. identified five distinct clinical phenotypes of asthma using unsupervised hierarchical cluster analysis.

Beena Kapadia
Assistant Professor

Cyber Security Framework – Components, Functions and Implementation

The internet has become the integral part of today's generation of people; from communicating through instant messages and emails to banking, travelling, studying and shopping, internet has touched every aspect of life. With the growing use of the internet by people, protecting important information has become a necessity. A computer that is not having appropriate security controls can be infected with malicious logic and thus any type of information can be accessed in moments. Hacking of important data, network outages, computer viruses and other cyber related threats affect our lives that range from minor inconvenience to serious incidents.

Cyber Security or Information technology security means protecting data, networks, programs and other information from unauthorized or unintended access, destruction or change. It can be also be defined as all operations and activities that are undertaken to minimize threats and any kind of vulnerabilities, and enforcing required policies for prevention, data assurance, recovery and other cyber security related operations. It encompasses all the mechanisms and processes that protect digital equipment, information and records from illegal or unintended access, manipulation or destruction.

Defining a Framework for Cyber Security is essential to assess and improve the ability of organisations and individuals to prevent, detect and respond to cyber-attacks.

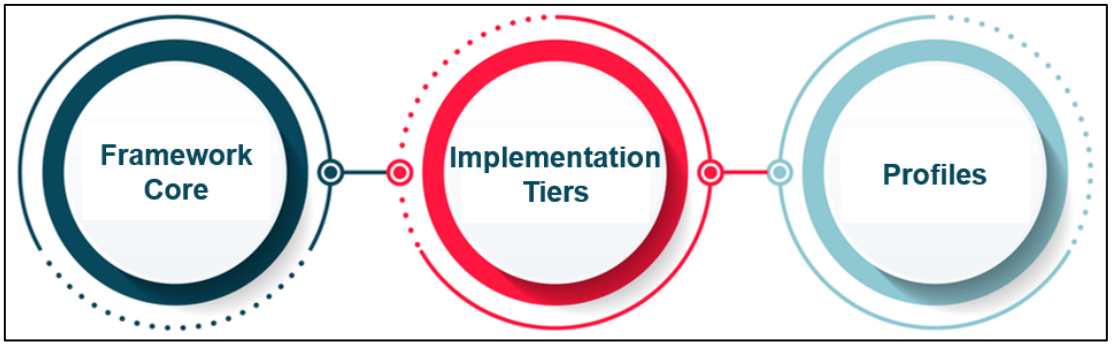
Existing Cyber Security Frameworks

The most frequently adopted frameworks are:

1. **PCI DSS (Payment Card Industry Data Security Standard)**: It is a set of security controls required to implement to protect payment account security. It is designed to protect credit card, debit card, and cash card transactions
2. **ISO 27001/27002 (International Organization for Standardization)**: Best practice recommendations for information security management and information security program elements.
3. **CIS Critical Security Controls**: A prescribed arrangement of activities for cyber protection that give particular and noteworthy approaches to stop the present most inescapable and perilous attacks. A key advantage of the Controls is that they organize and centre fewer activities with high outcomes
4. **NIST Framework**: A Framework for improving critical infrastructure Cybersecurity with a goal to improve organization's readiness for managing cybersecurity risk by leveraging standard methodologies and processes.

Components of Cybersecurity Framework

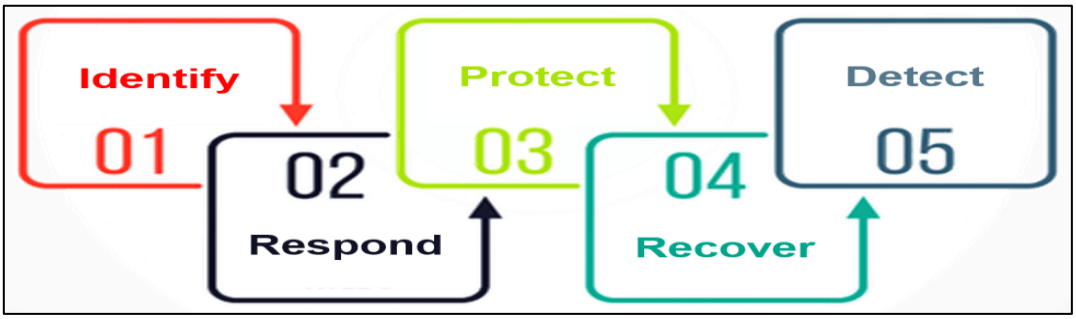
Most frameworks for Cyber Security include the following three key components:



1. **Framework Core:** It gives an arrangement of required Cybersecurity exercises and results utilizing normal understandable language. The Core guides associations in overseeing and decreasing their Cybersecurity chances in a way that supplements an association's current Cybersecurity and risk management processes.
2. **Implementation tiers:** It helps associations by giving setting on how an association sees Cybersecurity risk management. The tiers manage associations to consider the suitable level of thoroughness for their cybersecurity program and are regularly utilized as a specialized device to talk about hazard hunger, mission need, and spending plan.
3. **Profiles:** Profiles are an association's novel arrangement of their organizational prerequisites and goals, and assets against the coveted results of the Framework Core. Profiles are principally used to recognize and organize open doors for enhancing Cybersecurity at an association.

Cybersecurity Framework's Five Functions

The five functions included in the framework are:

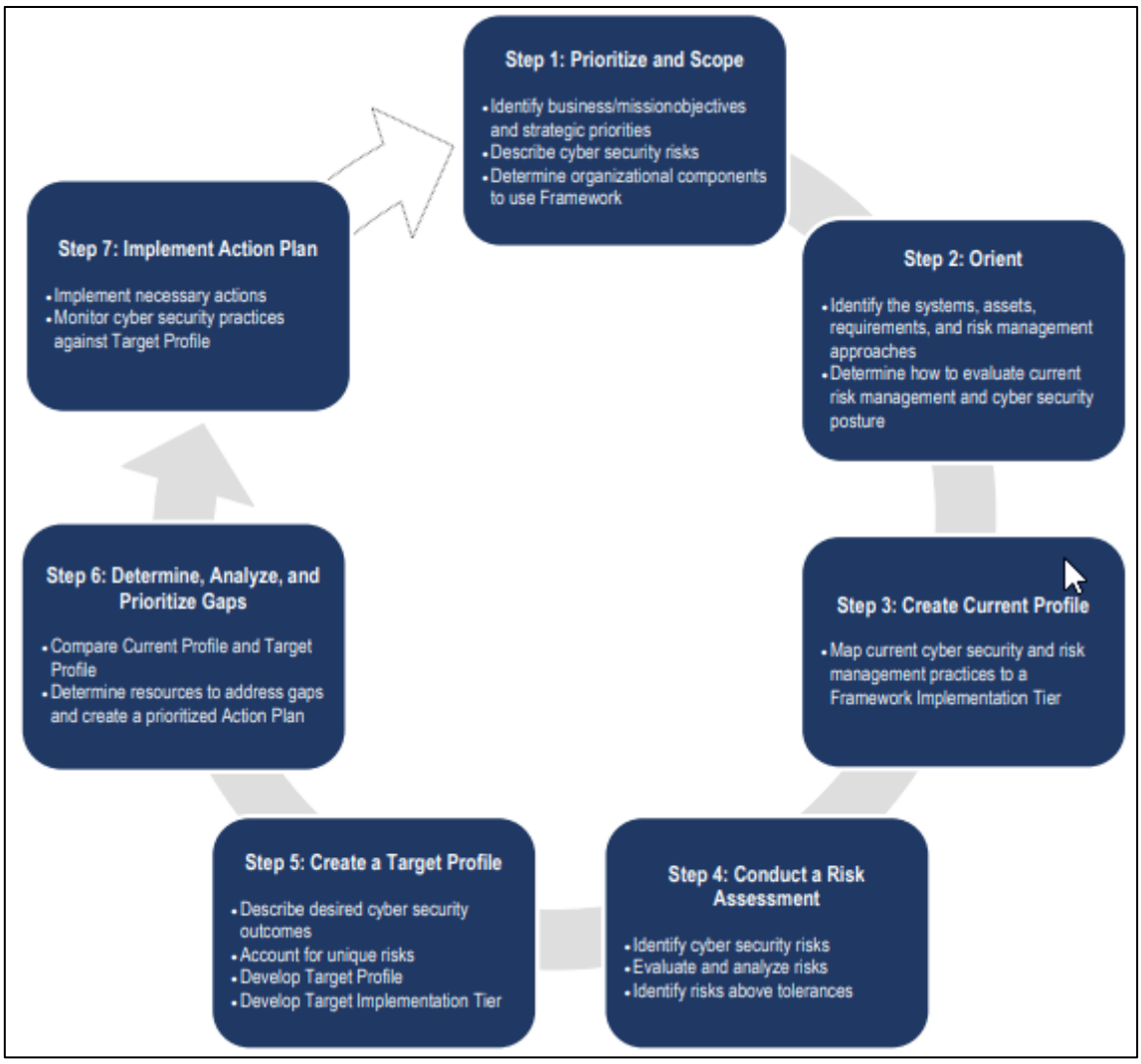


1. **Identify:** The Identify Function helps with building up a hierarchical comprehension in overseeing cybersecurity to frameworks, individuals, resources, information, and capacities.
2. **Protect:** The Protect Function underpins the capacity to restrict or contain the effect of a potential Cybersecurity occasion.

3. **Detect:** The Detect Function characterizes the fitting exercises to recognize the event of a Cybersecurity occasion.
4. **Respond:** The Respond Function incorporates proper activities to make a move regarding a distinguished Cybersecurity occurrence. The Respond Function bolsters the capacity to contain the effect of a potential Cybersecurity occurrence.
5. **Recover:** The Recover Function distinguishes proper exercises to keep up plans for versatility and to re-establish any abilities or administrations that were impeded because of a Cybersecurity event.

Steps to Implement Cybersecurity Framework

The Cybersecurity Framework defines 7 steps for establishing a cybersecurity program:



Implementing the Framework provides the following benefits to organizations:

- Evaluate and particularly depict its current and focused on digital security pose

- Identify gaps in its present projects, procedures, and workforce
- Identify and organize open doors for development utilizing a persistent and repeatable process
- Assess advance toward achieving its objective digital security act
- Demonstrate the association's arrangement with the Framework's broadly perceived accepted procedures
- Highlight any present practices that may outperform the Framework's prescribed practices
- Communicate its digital security act in a typical, perceived dialect to inside and outside partners—including clients, controllers, financial specialists, and approach producers

Mr. Bhavesh D. Shah

Assistant Professor

Embedded Device---Holter Monitor

In medicine, a Holter monitor (often simply Holter) is a type of ambulatory electrocardiography device, a portable device for cardiac monitoring (the monitoring of the electrical activity of the cardiovascular system) for at least 24 to 48 hours (often for two weeks at a time).

The Holter's most common use is for monitoring ECG heart activity (electrocardiography or ECG).Its extended recording period is sometimes useful for observing occasional cardiac arrhythmias which would be difficult to identify in a shorter period. For patients having more transient symptoms, a cardiac event monitor which can be worn for a month or more can be used.

The Holter monitor was developed at the Holter Research Laboratory in Helena Montana by experimental physicists Norman J. Holter and Bill Glasscock, who started work on radio telemetry in 1949. Inspired by a suggestion from cardiologist Paul Dudley White in the early 1950s.When used to study the heart, much like standard electrocardiography, the Holter monitor records electrical signals from the heart via a series of electrodes attached to the chest. Electrodes are placed over bones to minimize artifacts from muscular activity. The number and position of electrodes varies by model, but most Holter monitors employ between three and eight. These electrodes are connected to a small piece of equipment that is attached to the patient's belt or hung around the neck, keeping a log of the heart's electrical activity throughout the recording period. A 12 lead Holter system is also available when precise ECG signal information is required to analyse the exact nature and origin of the rhythm signal.

Figure

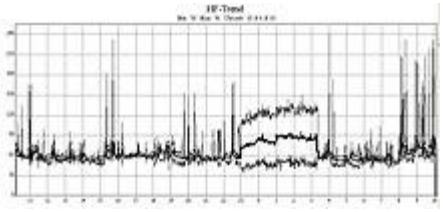
Why do people wear Holter monitors?

Regular electrocardiograms (ECGs or EKGs) let your doctor look at your heart's activity at one point in time during your ECG test. But abnormal heart rhythms and cardiac symptoms may come and go. That's why your doctor may want to evaluate your heartbeat over time while you go about your normal activities. You may be asked to wear a Holter monitor if you have fast, slow or irregular heartbeats called arrhythmias.

Wearing the monitor may tell your doctor:

1. If your medicines are working.
2. Why you have symptoms such as dizziness, faintness or the feeling that your heart is racing or skipping a beat.
3. If your heart is getting enough oxygen to meet its needs.

Data storage:



Older devices used reel-to-reel tapes or a standard C90 or C120 audio cassette and ran at a 1.7 mm/s or 2 mm/s speed to record the data. Once a recording was made, it could be played back and analyzed at 60x speed so 24 hours of recording could be analyzed in 24 minutes. More modern units record an EDF-file onto digital flash memory devices. The data is uploaded into a computer which then automatically analyzes the input, counting ECG complexes, calculating summary statistics such as average heart rate, minimum and maximum heart rate, and finding candidate areas in the recording worthy of further study by the technician.

Components:

Each Holter system consists of two basic parts – the hardware (called monitor or recorder) for recording the signal, and software for review and analysis of the record. Advanced Holter recorders are able to display the signal, which is very useful for checking the signal quality. Very often there is also a "patient button" located on the front site allowing the patient to press it in specific cases such as sickness, going to bed, taking pills, etc.; a special mark will be then placed into the record so that the doctors or technicians can quickly pinpoint these areas when analyzing the signal.

Recorder:

Most of the Holters monitor the ECG via only two or three channels (Note: depending on manufacturer, different counts of leads and lead systems are used). Today's trend is to minimize the number of leads to ensure the patient's comfort during recording. Although two/three

channel recording has been used for a long time in the Holter monitoring history, as mentioned above, 12 channel Holters have recently appeared. These systems use the classic Mason-Likar lead system, i.e. producing a signal in the same format as during the common rest ECG and/or stress test measurement. These Holters can occasionally provide information similar to that of an ECG stress test examination. They are also suitable when analyzing patients after myocardial infarction. Recordings from these 12-lead monitors are of a significantly lower resolution than those from a standard 12-lead ECG and in some cases have been shown to provide misleading ST segment representation, even though some devices allow setting the sampling frequency up to 1000 Hz for special-purpose exams such as detection of "late potential".

Another innovation is the inclusion of a triaxial movement sensor, which records the patient's physical activity, and on examination and software processing, extracts three movement statuses: sleeping, standing up, or walking. Some modern devices also have the ability to record a vocal patient diary entry that can be later listened to by the doctor. These data help the cardiologist to better identify events in relation to the patient's activity and diary.

Procedure:

Although some patients may feel uncomfortable about a Holter examination, the only hazards are potential minor skin abrasions to optimize signal quality, and it should have little effect on one's normal daily life.

The recording device can be worn in a case on a belt or on a strap across the chest. The device may be visible under light clothing, and those wearing a Holter monitor may wish to avoid shirts with a low neckline.

Persons being monitored should not limit normal daily activities, since its purpose is to record how a heart works under various actual conditions over an extended period. It is an electrical device, however, and should be kept dry; showering or swimming should probably be avoided. Monitors can be removed for a few minutes without invalidating collected data, but proper reattachment is critical to avoid degradation of its signals. Beyond changing batteries, one should leave its handling to trained personnel.

Sabir Moin.Moinuddin.Shaikh

Assistant Professor

How to “work from home”

The article talks about how to make the work from home situation work for everyone, who is not used to it. People don’t have ordinary structure of the workday routine when they are at home. No meeting to coworkers, no office desk, no working at office, our home becomes our office and we need to lead a whole new routine. This will end up wastage of more time in getting oneself used to this new mode of work

Some of the key point to be considered while working from home

1) Find spot for your work at home.
   Make a consistent desk or chair for working this will train your brain that these things are associated with work to be completed. Make the desk at home like the office one by keeping things that relates us with our office desk. This can include sticky notes, pen stands, favourite picture you keep on the office desk.
2) Use app for virtual meeting
   The are apps and software available in the market for virtual meeting like skype, hangout, zoom, gotomeeting etc. the meeting can also have some agenda less video conferencing just to have a good feel of meeting your friends and coworkers
3) Resist the temptation to work from bed
   Avoid putting your brain in working mode as soon as you get up. Follow the usual routines like waking up, getting dressed, eating breakfast, then moving to your new work space. Stay in your attires that you wear at home will make you feel lazy so try to dress up in the office clothes
4) Schedule
   Since a good amount of time which used to get wasted in commuting is in our hand now, try to make most of it. Use the time of the day according to your waking up schedule, like if you are a morning person start the work early and take a break in between. If you’re a night person than start your work little late in the afternoon.
5) Working all the time
   Don’t consider that work from home means you need to work whole day. Take mental breaks, like you would in the office. Its important to set boundaries to work from home. Leave the desk once the allotted duty hours are over.
6) To do list
   Plan your work schedule by creating a do to list and daily plan and stick to it. Try to prepare it a night before so that the next day you will know the things that you need to complete.
7) Finish the things that you feel add value to your skills
   Use the extra time that you are saving on travelling to complete task that add value to your skills. Complete and brainstorm on the long term projects.

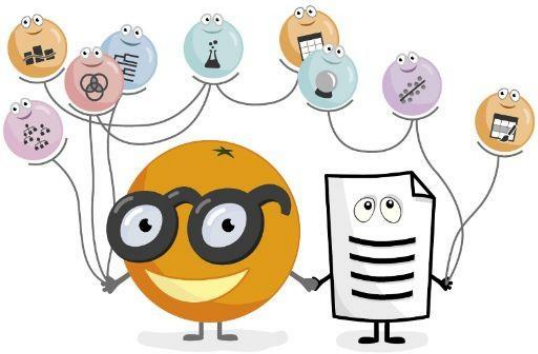
Working from home will add some values to our usual machine like thinking. Overcrowded trains and traffics on the lanes can be tackled by this option by allowing employees to take work from home at least few days in a week. It will also allow the employee to spend some good time with the family, show them the way they work and increase their emotional intelligence which is the need of the hours today.

Seema Vishwakarma

Assistant Professor

Orange-Open Source Tool

What is Orange?
Orange is an open source tool which provides machine learning and data visualization capabilities for novice and expert users.



Orange is an open source component-based visual programming software package used for data visualization, machine learning, data mining, and data analysis. Components of Orange are called widgets and they range from simple data visualization, subset selection, and pre-processing, to practical evaluation of learning algorithms and predictive modeling.

In Orange, visual programming is implemented through an interface in which workflows are created by linking predefined or user-designed widgets. While advanced users can use Orange as a Python library for data manipulation and widget alteration.

Orange is an open source data visualization and data analysis tool for data mining through visual programming or Python scripting. The tool has components for almost all well-known machine learning algorithms, add-ons for bioinformatics and text mining as well as features for data analytics also. So, for researchers it is a one stop solution for pre-processing of dataset, visualization of dataset using graphs, all inbuilt machine learning algorithms, test and score feature for measuring accuracy of algorithm on different datasets along with many more fantastic features.

Eye-catching features of Orange

1. Open source
The best part of Orange is that it is open source so that you can get its code and can even modify the tool as per your requirement. This tool is undoubtedly a boon for people doing PhD or master's in data science and machine learning. Also, you can get the source code of almost all machine learning algorithms too. So, you can modify the algorithm as per your application and then you can add that modified algorithm in Orange and take the results. This is seriously amazing feature of the Orange tool.

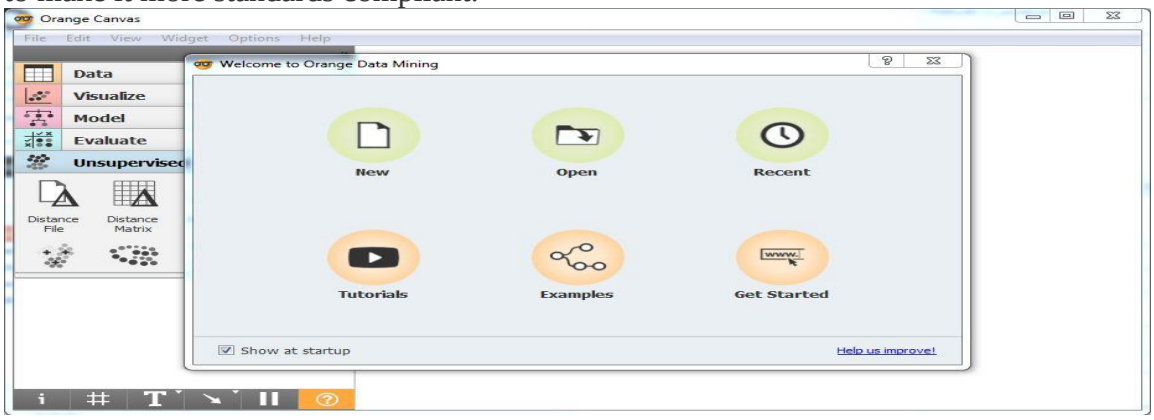
2. Visual Programming
This tool is not just meant for computer science professionals but even novice users can use it as it provides visual programming. It is as simple installing a game and then playing it.There is no need to learn any kind of programming languages like JAVA, C, C++ or Python etc, the only thing which you should know is data mining concepts and you should know that which algorithm should be used in a particular scenario. It provides drag and drop facilities. It even provides lines for connection. To plot a graph was never such playful as Orange has made it. You will definitely fall in love with this tool when you will experience its flexible and visual

environment. It provides dotted lines if connection is not proper. If you are not using proper machine learning algorithm or prediction algorithm, then it will not allow you to connect with the data.
In short, visual programming provides interactive data exploration for rapid qualitative analysis with clear visualizations. GUI allows users to focus on exploratory data analysis instead of coding, while smart defaults make prototyping of any data analysis workflow fast and extremely easy.

Just place the widgets on the canvas, connect them, load your datasets and yield the insight.

3. Supports Google sheet
Oftenly in data science tools, one can browse any file from the local hard disk. But with Orange, it is possible to fetch the data from a given URL. It also provides support to fetch the data from Google spreadsheet which is its most eye-catching feature. So, if your data is not in your local hard drive, no need to worry as Orange provides the facility of fetching the data from Google spreadsheets also.

4. Add-ons are available to extend the functionality
It is possible to extend the functionality of Orange through add-ons which are available online. In fact, Orange never provides the toolbox for crunching bioinformatical data as an integral part of it; rather than it has always been an add-on. The exact process of distribution of add-ons has changed considerably in the last year to streamline the process for add-on authors and to make it more standards compliant.



The above picture shows the first screen that appears upon opening the Orange tool.

Rohini Desai

Assistant Professor

Plagiarism

The Intellectual Property of individuals protected by removing the plagiarism from the literary work. It is necessary to publish the articles to claim for its own work, which, too, is published. The publisher has to ensure its original work. It is interesting to know that the word plagiarism is coming from the Latin word plagiarise. Plagiarism meaning is kidnapper [1].

What is Plagiarism?

Plagiarism is the act of presenting work or ideas of others as your own [2]. Plagiarism is an unethical activity, and the credit should be given to the original author, who has written it. This act is like theft—here, theft of literary. In academics, someone else's research papers and research projects are presented and taking credit for the work.

Plagiarism & User's Assumptions

The assumption of plagiarism that the use of a dead person's work is not plagiarism. While writing a research paper, even if you paraphrase some of the parts of other authors' research work, you are supposed to give a citation. It is assumed that paraphrasing is not plagiarism. Paraphrasing means restructured other's work in their own words. If the person is using word-to-word, exact content as it is there in the original document, then quotations should be used, and cite it.

There is a need to know something more about plagiarism. The author cannot use even his/her own previously published write up without citation. An author cannot publish the content of the friends, even though he permitted to use the literary work without citation.

Several things are available as open for free access; still, you should cite it. Literary is free of cost. But it is not free of citation. One should give credit to the original intellectual property owner. You should cite, in case you have paid for it.

If your project guide, professor, friends are giving you ideas, then also you must provide them acknowledgment or credit by citing the content with their names.

Once we understand the concept of plagiarism, we need to know about the citation and the knowledge about the tools used to find out the plagiarism.

Plagiarism Detection Software Tools

Many tools are available online. These tools help find out the plagiarism in the literary work. Plagiarism software finds out the plagiarised portion of the document and also providing the percentage of plagiarism. Useful journals and their publishers rejected the work if there is more than 15% similarity. Publishers like Springer check the plagiarism and decide on acceptance/rejection based on the percentage of plagiarism. Researchers and Authors should know about plagiarism. It is essential to follow the rules of plagiarism. Publishers are always taking the undertaking from the authors before publication of the article, research paper. The article and research work should be original. The most popular plagiarism software is

Grammarly, Whitesmoke, ProWritingAid, DupliCheker, Plagiarismcheck.org, Quetext, Plagiarism Checker, Copyleaks, Viper and Turnitin.

This software is helping the author to find out the duplication or similarity in the document. If it is more, then the author has to remove it or rewrite the entire content. Plagiarism is essential to protect the theft of literary and helpful to give credit to its original writer.

References

1. http://130.65.109.143/plagiarism/tutorial/introduction.php
2. http://www.politicsnj.com/plagerism90903.htm

Ujwala Sav

Assistant Professor

RPA (Robotic Process Automation)

In programming when we have find factor of a number what we do is write a function to get an answer but what shall be done if the situation calls for a repeated task like invoice generation by the customer to pay the bill in billing center or at bank branches for such a situation instead having a human; RPA are been or can been implemented.

RPA defines itself to be as automated task process just like human beings doing it for various applications and systems. The RPA is a software program which runs on an end user pc, laptop and mobile device. It holds sequence of commands which are executed by bots under some defined set of business rules. It is a non-humanoid software robot offering a 24/7 virtual workforce that can perform any rule-based, workflow-driven or mundane time-consuming and costly clerical tasks, 10-20 times faster than a human. These software bots operate only on the UI level like a human, without API instrumentation or any security bypass deeming them perfect for various operational processes which require a security clearance.

Let's talk about the two Use cases of RPA :

Education: The space which is evolving from chalk and board to multimedia data transmission to the students now involves multidisciplinary study to cope up with this pressure RPA can be used so let's see at the challenges:

•Course Registration, Shortlisting and Enrolment Process
With automation, we will be able to check student's eligibility criteria, validate information, shortlist candidates which indeed consumes a lot of time. Also, this could be the best solution during peak hours.

•Attendance management
Manage the attendance, send automated notifications and reports to parent and students.

•IT Operations
IT operations can get rid of a lot of time-consuming tickets like password reset, unlock accounts, reboot systems, restart service, install updates, monitor alerts with automation.

•Schedule meeting scheduling, timetable updates, equipment reservations, updates scheduling
Institution staff and students must attend a lot of events and meeting. Notifying students and faculty about meeting, events and sending updates can be easily automated to save time.

•Human resources, Admin, Finance
Automate processes like payroll processing, vendor management, accounts payable-receivable, employee onboarding-offboarding, inventory management, vendor management and many more.

•Chatbot

Chabot can help automate the general queries from students, staff and website visitors like admission schedule, admission process, contact person, course information. With NLP and AI, chatbot can understand the natural language, learn from past data and reply to queries like a human.

Healthcare: With increasing population and its need of primary and secondary care Healthcare industry must be always be ahead, to cope with this demand.Hence RPA can be implemented for the following.

Patients' scheduling

Software robots can streamline online scheduling. Factors received via the appointment request, like diagnosis, location, insurance carrier, personal preferences, etc., can be gathered in a report, and forwarded to a referral management representative who makes the appointment.

Management of supply processes (claims and billing)

Claims management takes up a lot of healthcare administrators' time, if only because of the many sub-processes that must be handled: inputting, processing, evaluating, dealing with appeals. Automation can lead to substantially faster and error-free processing, thereby laying down administrators of appreciable burden. Moreover, consider that a significant amount of healthcare claims (as much as 30-40%, according to a KPMG study) don't comply with official demands. Software bots can easily identify those exceptions, and thus save many unnecessary payments.

Implementation of health plans

By improving patient records management (due to technology features such as high accuracy and minimization of human error), RPA in healthcare results in tracking more standardized, patient-specific routes towards patients' desired health objectives. Automated processes allow more timely identification of patients who deviate from the plans, and consequently, they make it easier to bring them back on track. In fact, continuous remote monitoring of individual records at the level of population health without robotic process automation involves very large costs.

Although the above-mentioned cases are limited there are multiple benefits of RPA which can help business flourish and customers too.

Mithila Chavan

Assistant Professor

Tiny AI – Would it be greener?

Artificial Intelligence has been in trend since the past few years and is growing at a rapid rate. But AI has one problem. All the AI programmers are building more and more powerful algorithms. Researchers around the globe are using great amounts of data and computing power and are relying on cloud services. Has anyone thought of what are the possible drawbacks? Yes, you read it right! There are drawbacks – this is not only generating large amounts of carbon emissions but is also limiting the speed and privacy of AI applications.

In order to avoid these situations, the AI research community is working on reducing the size of the algorithms, especially those that require large amounts of datasets and computational power. **Tiny AI** is the term used to describe these efforts. These researchers develop methods that are called distilled methods, that not only reduce the size of a model but they do this while accelerating inference and maintaining high levels of accuracy. Using these methods, a model can be scaled down, so much to an extent of factors reaching up to 10x. Also if the algorithm is reduced to a smaller size, then it can be deployed on the edge instead of sending its data to the cloud. Which means you can now make decisions on the device itself.

All this could bring about several benefits. Multiple existing services such as voice assistants, autocorrect, digital cameras will get better and even faster by not having to connect to cloud every time that they need access to an algorithm.

Tiny AI can also make new applications possible. Applications like mobile based medical image analysis or self - driven cars with faster reaction times. On top of it, you have your data stored on edge improves privacy and security.

If all these experiments and studies fit together, AI will then may also be called as a greener AI, because by shrinking the algorithm, we are also putting an indirect control on other technical aspects which are leading to many such hidden environmental costs.

Ms. Ketaki Ghawali

Assistant Professor

Usefulness of Association Rule Mining

Association Rule Mining (ARM) has become a developed field of exploration. So many research papers, articles are surveyed in the field of ARM. ARM can be divided into the following areas: Positive rule mining, Negative rule mining and Interestingness measures.

Association rules have been broadly used in many applications fields for discovery outline in data. The outline reveals mixtures of actions that occur at the same time. One of the best area is commercial ground, where learning of pattern or association helps in real conclusion creation and marketing. Other zones where association rule mining can be applied are discovery pattern in organic records, market basket study of public library flow data, protein composition, population and economic census etc. Recent studies have shown that there are various algorithms for finding association rule. Apriori algorithm is the best known algorithm for association rule mining. However the complication and performance of mining algorithms is subject to research area, as they have to mine a greater set of facts i.e. most of the study are based on how to streamline association rule and to progress the algorithm presentation.

Let us understand the basis of association rule and their implication. Association rule mining finds stimulating relations and correlation among great set of data items. Association rules show abilities value conditions that occur frequently together in a given dataset. Association rules offer evidence of this type in the form of "if-then" statements. These rules are calculated from the data and unlike the if-then rules of logics, association rules are probabilistic in nature. In addition to the "if" part and "then" part, an association rule has two numbers that express the degree of improbability about the rule. In association study the antecedent and consequent are sets of items that are disjoint (do not have any items in common).

Support: The support means the number of relations that include all items in the antecedent and consequent parts of the rule. (The support is stated as a percentage of the total number of records in the list.)

Confidence: Confidence is the ratio of the number of relations that include all items in the consequent as well as the antecedent (namely, the support) to the number of relations that include all items in the antecedent.

Lift: Lift is the ratio of confidence to probable confidence. Lift is a value that gives us evidence about the rise in chance of the "then" (consequent) given the "if" (antecedent) part.

DATA MINING TECHNIQUES

Data mining uses Supervised and unsupervised learning. Supervised learning involves a parameter that helps to learn. The learning predicts a result based on certain criteria. Classification and regression are examples of supervised learning. Similarly, unsupervised learning is a technique that does not involves a supervisor. It outlines class of data without his obligation. Common example is the clustering. Clustering: It identifies a category of data in a finite set which describe the task. It can be used to predict an outcome. K-means and x-means are some of the algorithms that have been used in clinical process and diagnosing results.

APPLICATION AREAS

Medical diagnosis: Medical Diagnosis is a complicated process. ARM can help the doctors or physicians to cure patients. The general problem of the induction of reliable diagnostic rules is hard because theoretically no induction process by itself can guarantee the correctness of induced hypotheses. Diagnosis involves unreliable tests and the presence of noise in training examples. This may result in hypotheses with unsatisfactory prediction accuracy which is too unreliable for critical medical applications.

Protein sequences: The study of protein sequence evolution is the foundation of molecular evolution and it offers insights in the mechanism of evolution. The evolution rate of protein is weakly influenced by the functioning of proteins. For any organism, Proteins are vital constituents of cellular machinery. Recombination of DNA technology has provided tools for the rapid resolve of DNA sequences and, by deduction, the amino acid sequences of proteins from structural genes. Proteins sequences are made up of 20 types of amino acids.

Identification of Diseases: Data mining based on association rule for finding incidence of ailment by patients. Medical stat association forecast was devised to extract information from healthcare database which predict the relationship among primary disease and secondary disease. A methodology to identify the locally frequent diseases by implementing Apriori mining technique is suggested. The data may be collected data from medical centre and generate a frequent pattern to identify the frequent disease. The current system of knowledge discovery in medical database using data mining can be studied. They also carried out some

comparative study among techniques and concluded that the association rule mining perform well for medical data set. Various applications of association rule mining in medical data in areas of infections, adverse drug reactions, etc. By understanding the previous work carried out by various researchers in the area of data mining of disease analysis, the current work of the author is unique and it is need and important for the society.

Clinical Observations: System-prototype, named CSCP system, using the association rules of data mining technique applied to a patients' (assumed) database for discovering patterns of diseases that might be carried by a patients. As a novel idea of mining the data capturing process can further be modified in the clinics as well as in the data-warehouses which should further be involved to enhance the CSCP system we have proposed. The recognised pattern by this implementation definitely can improve the healthcare services along with medical researchers for further exploring trends of diseases that are correlated.

In this work a revision on the main function areas of union rules has been focused. It is all about to find some kind of pattern or relationship between various datasets. The outcome is association rules, and it is an iterative refinement process. Further work can be done on the employee database for finding association rules related to job stability.

Pushpa Mahapatro

Assistant Professor

5 Emerging AI and Machine Learning Trends to Watch In 2020.

This news shouldn't surprise you: the market for artificial intelligence solutions continues to grow at a fast clip and represents tens of billions of dollars in revenue.

Case in point: a report by research firm IDC in September said global spending for AI systems will reach $97.9 billion in 2023, a staggering increase from the projected $37.5 billion that will be spent this year. That means the annual growth rate will be 28.4 percent over the next several years.

This means 2020 will be a critical year to set the tone for the next decade of innovations in the AI space and continue the existing momentum. But what does that mean for organizations who are selling and buying AI solutions? In which areas should they invest?

What follows are six emerging AI and machine learning trends organizations should watch and consider for their own strategies in 2020.

1. Explainable AI:

   As artificial intelligence becomes a bigger part of how organization make decisions, there will be a greater need for AI-fuelled applications to explain how they reached such decisions. According to Gartner, 30 percent of government and large enterprise contracts will require these solutions by 2025. That means 2020 will be an important year in creating the foundation for this emerging market. When it comes to data science and machine learning platforms, they will need to list out in clear terms the accuracy, attributes and model statistics used to reach conclusions, according to the research firm. In a separate report, Gartner said rules and legislation like the European Union's General Data Protection Regulation will push the need for such explainable AI solutions.

Explainable AI (XAI) refers to methods and techniques in the application of artificial intelligence technology (AI) such that the results of the solution can be understood by human experts. It contrasts with the concept of the "**black box**" in machine learning where even their designers cannot explain why the AI arrived at a specific decision. XAI is an implementation of the social right to explanation. Some claim that transparency rarely comes for free and that there are often trade-offs between the accuracy and the explain-ability of a solution.

The technical challenge of explaining AI decisions is sometimes known as the interpretability problem. Another consideration is info-obesity (overload of information), thus, full transparency may not be always possible or even required.

AI systems optimize behaviour to satisfy a mathematically-specified goal system chosen by the system designers, such as the command "maximize accuracy of assessing how positive film reviews are in the test dataset". The AI may learn useful general rules from the test-set, such as "reviews containing the word 'horrible'" are likely to be negative". However, it may also learn inappropriate rules, such as "reviews containing 'Daniel Day-Lewis' are usually positive"; such rules may be undesirable if they are deemed likely to fail to generalize outside the test set, or if people consider the rule to

be "cheating" or "unfair". A human can audit rules in an XAI to get an idea how likely the system is to generalize to future real-world data outside the test-set.

2. Autonomous AI:

While self-driving vehicles are getting the most attention in the autonomous space, autonomous technology is expected to become far broader in application, thanks to continued progress made in advanced AI systems by researchers and technologists. Gartner said autonomous AI will be among the top technology trends in 2020 that will shape the future. This will enable things like collaborative swarms of drones or robots to move on their own and in concert with each other, automating tasks that were once manual or semi-automated. Business use cases include autonomous shipping and advanced agriculture, where robots can operate farms autonomously.

Self-driving cars combine a variety of sensors to perceive their surroundings, such as radar, lidar, sonar, GPS, odometry and inertial measurement units. Advanced control

systems interpret sensory information to identify appropriate navigation paths, as well as obstacles and relevant signage.

Long distance trucking is seen as being at the forefront of adopting and implementing the technology.

3. AI Security:

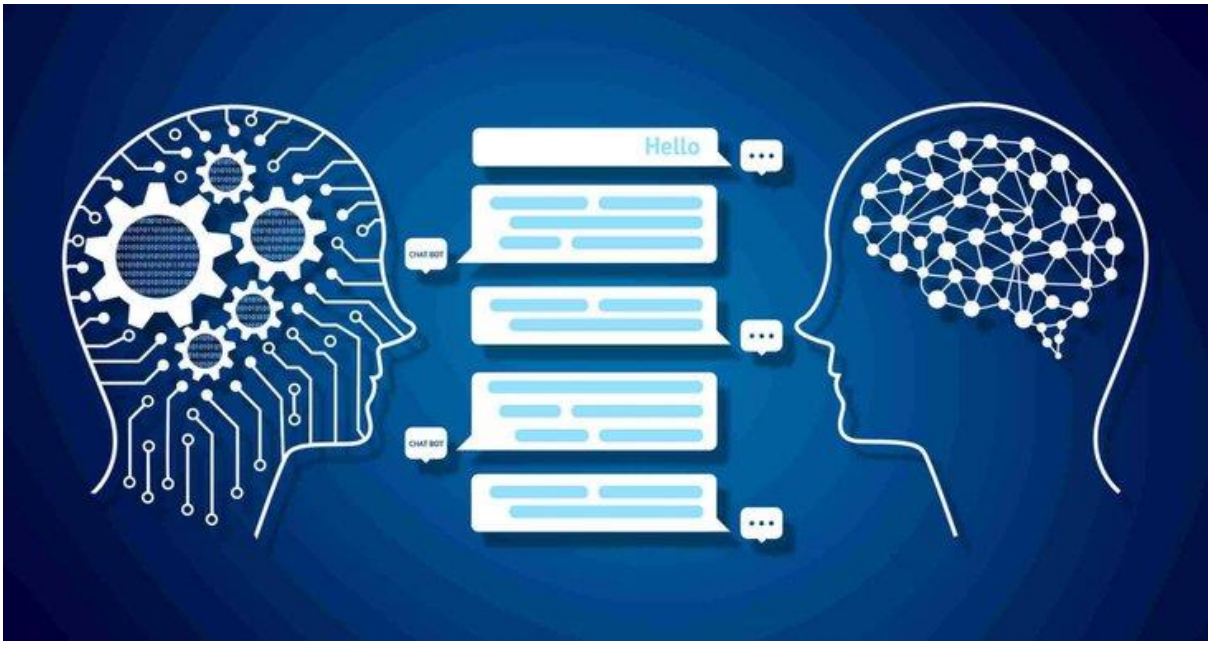
Artificial intelligence will have a major impact on cybersecurity in three key ways, according to Gartner. First, organizations will need to ensure they can protect AI-powered systems from being penetrated to prevent such systems from producing faulty decisions or leaking underlying data sets. According to the research firm, 30 percent of all AI cyberattacks will leverage training-data poisoning, AI model theft or adversarial samples to compromise AI systems. These organizations also will need to use AI to enhance their security capabilities, which is being made possible by next-generation security products that have started to enter the market. At the same time, they will need to be aware of how bad actors will use machine learning and other AI techniques to carry out new kinds of cyberattacks.

The growing and evolving cyber security risk facing global businesses can be stemmed by the integration of AI into security systems. Machine Learning techniques such as unsupervised learning and continuous retraining can keep us ahead of the cyber criminals. Machine learning and artificial intelligence (AI) are being applied more broadly across industries and applications than ever before as computing power, data collection and storage capabilities increase. This vast trove of data is valuable fodder for AI, which can process and analyse everything captured to understand new trends and details.

4. Conversational AI:

While efforts to bring conversational AI to the mass market began in earnest a few years ago, the technology is expected to have a big moment in 2020. Next year, for instance, Gartner expects 50 percent of analytical queries to come from search, natural language processing or voice queries, which will make analytics systems more accessible within organizations. In a separate report, the research firm said that 70 percent of white-collar workers will work with conversational platforms every day.

Conversational Artificial Intelligence or Conversational AI is a set of technologies that enable computers to simulate real conversations.



Businesses can use conversational AI to automate customer-facing touchpoints everywhere—on social media platforms like Facebook and Twitter, on their website, their app or even on voice assistants like Google Home.

Conversational AI systems offer a more straightforward and direct pipeline for customers sort problems out, address concerns and reach goals.

Both the terms 'Chatbot' and 'Conversational AI' have more or less the same meaning. ***'Conversational AI', however, is more inclusive of all the technology that falls under the bot umbrella like voice bots and voice + text assistants, whereas 'chatbots' have a more limited 'text-only' connotation.***

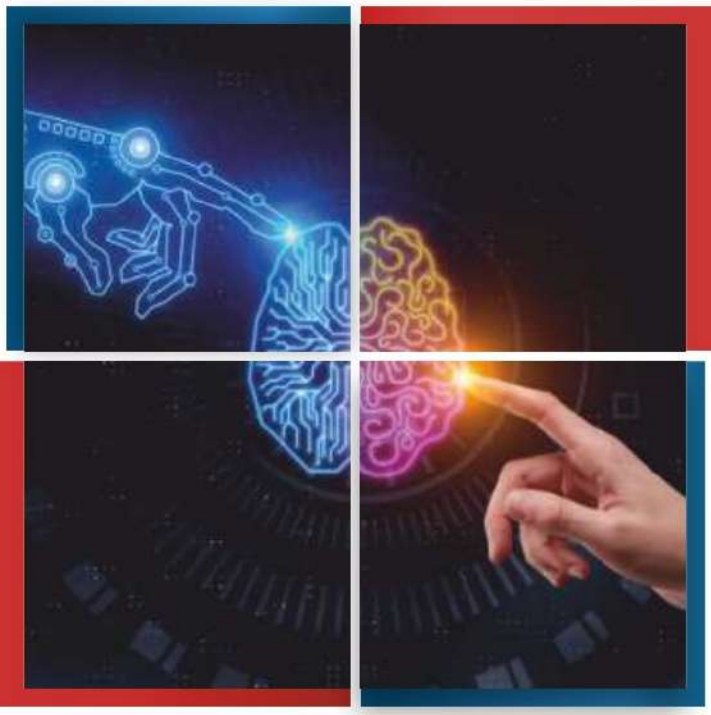
5. AI Infrastructure:

With artificial intelligence impacting a diverse set of workloads, it should be no surprise that AI itself will have an increasingly outsize role to play in future IT infrastructure investments. According to Gartner, AI will be one of the top drivers for infrastructure decisions through 2023. This is important because AI workloads require specialized hardware and software configurations as well as dedicated teams who can continuously manage AI models that help drive operational decisions.

Many AI and Deep Learning (DL) workloads run best on Graphical Processing Units (GPUs), but a switch to GPU based processing also requires optimization across filesystems and storage. An organization architecting for sustained AI success should look for storage solutions that are turn-key, pre-configured and provide scale-out capabilities for capacity and performance—which will ensure data is in the right place at the right time. AI storage infrastructure must be architected for all types of I/O patterns and data layouts handling any thread count, the toughest IO patterns, and dynamic data placement. In addition, they must be container aware, deliver direct GPU integration, multi-rail networking, and work with accelerated protocols for AI frameworks. These capabilities will ensure that the GPU compute engines will remain saturated and delivering full AI application acceleration.

Sai Ganesh.

Student (SY-BSc IT)



EDITORIAL TEAM :
Mr. Makarand Deshpande
Ms. Maitreyi Joglekar

VSIT | Vidyalankar School of Information Technology

Vidyalankar Educational Campus, Vidyalankar Marg, Wadala (East), Mumbai - 400 037 Maharashtra, India.
Telefax: +91 22 2416 1126 | Email: principal@vsit.edu.in