Assignment answers

1 Answer) History and developments in Artificial Intelligence:- The Idea of AI was existed from long time in the history of the humans as per modern science. We can get the evidence of existence of its idea, the idea of robots in the Greek mythologies where they had an idea of a machine to through up stones at the intruders by detecting the intruders. Even though the name for the thought of independently functioning machines was given very recently in the ara’s of 1940’s by Warren McCulloch. The first step could be taken as algorithms. The first algorithm is Euclid ’s algorithm to compute the greatest common divisor. Warren McCulloch drew fundamentals on three different sources.

There are a gradually rise in the advancements in the area of AI. From the field of gaming to the field of surgery, warfare, transportation there is a tremendous list of advancements.

The google AI has a game that trained it self against several game testcases while playing with it self and learning with each game in the game of GO, it has reached a level where the best of the best world champs could not defeat the computer. Now there are robots that can fire weapons and can do work on any terrains, so that can be considered as the advancement in rescue missions.

The Tesla is a better example of automated driving which is a progress in the transportation field. Another example is Baidu that has launched the Apollo Go Robotaxi service in the Chinese cities, this can be taken as advancement too. The mars rover which can identify and explore a complete new planet without any human interference where the rover exams the different climatic conditions, soil condition and soil fertility by identifying several different compositions of the soil, calculating the temperature and mapping different places on the mars, remembering identifying them and exploring other regions in the areas again. This can as well be taken as greater achievements.

When taken into account of the medical field we can see the research and development of the treatment, diagnose a disease from CT lung scan. The COVID-19 protein has been successfully decoded several times and future occurrences of the outbreak are being predicted along with the number of people affected which helps in the tackling of the covid situations.

In few city police takes help of AI to predict the future place where the possibility of crime to happen takes place, which helped them to prevent those from happening .So, The AI is helping the cops to secure the areas.

In Manufacture industry the robots works beside humans to a limited range of tasks such as in assembly line, stacking and analysis of sensors working that would help the manufacturing run smoothly and deliver more in limited time.

While the textbooks are being replaced with digital version, early-stage virtual tutors assist human instructors and facial analysis gauges the emotions of students to help determine who’s struggling or bored and better tailor the experience to their individual needs.

The AI helps in creating better future by predicting the occurrence and possible areas of impact due to any natural calamity, saving millions of life and property from damages. There are developments in the areas which helps the differently abled persons to cope up with the normally abled persons with technology.

Speech recognition is another application which helps in enhancing the security of personal devices like mobiles, cars, laptops, systems etc,.

Stock market prediction is another area of economy that helps the one who invest rich by accurately predicting the next move by capturing the sentiments of the market and big bull investers.

2 Answer) Data is what runs the industries in the present scenarios. The concept of big data is how to use the data that is generally in the market which is available in an effective manner and what can one understand from the data is given my the machine learning.

Data can be manipulations, as everything is controlled by data. I believe that one who can manipulate data can manipulate anything, he is in absolute control of anything and everything. Consider 3 Industries such as Food Industry, Manufacturing Industry, Banking Industry.

1)Food Industry:- Let us consider a simple factory, where the food is distributed all across the U.S. Machine learning technology in the food industry is an essential part of food production and delivery processes. The production and delivery is being done with the help of technology. To create the food in that heavy quantity is nearly impossible with humans alone but it is possible with the help of machine and human intelligence. The food quality is better with low cost. These new methods help in automating processes, saving money, and to reduce human errors. These Big data and ML can benefit bars and restaurants by summarizing the feedback, evaluating the feedback, which in turn helps those to offer great service, due to which it improves the economy of the U.S. They key for success for any restaurant would be preparing the dish. Knowing what goods to manufacture in large numbers or what dishes are the best choice to include on your restaurant menu is the key to increase revenue in that particular industry. As customers and industry expectations change so quickly, staying one step ahead of the competition has become more vital than ever. Food manufacturers and business owners should consider defining popular tastes and preferences as the most valuable resource. Large manufacturers and restaurants both need expensive and complex machinery to clean and process a large number of meals every day. This results in a large amount of contaminants passing through the cleansing equipment. The terminology for this system is SOCIP (Self-Optimizing-Clean-In-Place). A combination of ultrasonic detection and optical fluorescence imaging is used to assess food remnants and microbiological detritus in food processing equipment. However, this system has one drawback: it is operated automatically, that implies it is designed for the worst-case situation, which leads to overcleaning. Nonetheless, it is estimated that the method would save the US food sector approximately £100 million.

2) Manufacturing Industry :- To manufacture heavy machine we definitely need heavy machine to create them. To reduce the accidents and errors while dealing with such huge tasks with machines could use some extra muscle and extra intelligence.

With the data the manufacturers can create the models which are more useful and more beneficial to the costumers. Due to increase in productivity there would be a good rise in the imports of oil and exports of manufactured vehicles would take place which builds good relations with the neighboring countries and also the economy.

3)Banking Industry:- This industry is the one of the most important part of the economy. There are 3 main areas in this industry are fraud detection, marketing and credit risk management. Bank are able to use ML to combat the fraud by using the data from the end users. Analyze the performance and return on investment can be calculated different algorithms used in the ML. These can also be used in decision making to manage the probability of loss by analyzing the end user behavior and granting credits based on the results.

3Answer)

SUPERVISED LEARNING:

Supervised machine learning is a category of machine learning where the machine’s input is being labelled. The data which is being present is already labelled so the machine just need to get the output from the given data. This can be of 2 types regression and classification. Regression gives a single output this can be desired using many techniques. Unlike regression classification gives set of outputs on the basis of the type. It can do these using techniques like random forest, Svm etc. This model has both the output and input so designing the model will be simple comparatively. As the data is labelled the algorithms is trained with it. As the model is trained with both the input and output the prediction of output is highly accurate. The only challenge with this is that it cannot deal with large volume of data.

UNSUPERVISED LEARNING:

Unlike supervised learning the input here is not labelled so the prediction of the output becomes difficult. Designing this model is very difficult as the data is not labelled. This model will only get input. This will have clustering and association. The model tries and identifies similar kind of items and makes clusters of the similar items. One can even set the number of clusters that model need to forms. Association is used to gain a relation between the inputs. The algorithms for unsupervised learning are Cluster algorithm, knn(k means clustering), Hierarchical clustering etc.

SEMI SUPERVISED LEARNING:

This is a category of machine learning which will be in between supervised and unsupervised learning. This will be having labelled inputs of small amount and rest of that will be unlabeled. The category has continuity assumption, cluster assumption and manifold assumption. Continuity assumption is a type where the data is close to the label that is most likely matching. Cluster assumption is where the input data which is not labeled will be sharing different label. And Manifold is where it has both labeled and unlabeled data in it together. Methods for this are generative models, low-density separation, Laplacian regularization, etc.

4 Answer) Data Consolidation enables organizations to break down information barriers and data silos in order to make information more accessible. Businesses may have better control over their records assets by consolidating information in one integrated source. Because the dataset is already in a pre-processed state, the ETL technique utilized for Data Consolidation ensures high-velocity analyses. The availability of fantastic Data Analytics not only aids in accelerated decision-making but also saves operating expenditures.

Data Audit Even with procedures in place to preserve records in excellent condition, errors are unavoidable. Conducting quarterly data audits helps you to discover errors before they skew your decision-making and uncover fundamental causes of errors before they become systematic issues. Take a random sample of your facts. It should be large enough to gather data from your database at some point. Schedule an interview with key parties to assess the veracity of the data.

Data cleaning is available to regulate the data in an organized fashion, data repository is minimized and secured. This is obtained by utilizing actions such as completely removing duplicates from raw records, special characters, empty fields, misspelled terms, adding or removing columns and rows as necessary, as well as using third-party equipment and information to assist with records curation.

When one comes across errors, repair them and compare them to comparable data and information supplied by the same person to determine whether the faults are individual or routine. Examine each fact point to confirm that it is valid, up to current, has all necessary information, and that the records are consistent throughout.

5 Answer) Velocity:- Velocity is the rate at which huge data is being gathered. This velocity tends to rise year after year as community generation and technology get more powerful, letting firms to collect more data points at the same time.

Example:- The wiki receives more than a million visits per minute every single day.

Volume:- The volume of data refers to the size of the data sets that must be examined and processed, which are now typically in the terabyte and petabyte range. The sheer volume of data necessarily requires processing techniques that are distinguishable from standard storage and processing capabilities.

Example:- Amazon has more than 95 million subscribers all around the world and receives more than 150 million order per day.

Value:- The price of the records being collected is referred to as their value. In terms of making choices or enhancing operations, some of the Big Data acquired by a commercial enterprise agency may be quite low-cost. Due to regulatory obligations, a company may be required to gather and retain enormous volumes of records at no expense. However, in order to obtain Big Data voluntarily, a company must first determine what data are being collected and how they may be used to the organization. It may be desirable to forego acquiring information if there is no cost for doing so now or in the near future. Free records may sometimes serve as both a diversion and a hindrance.

Variety:- The many types of information that may be gathered are referred to as variety. This might be interdependent records such as a first name or a last name or an e-mail address. It may also be unorganized as a group with a product comparison. In one of these cases, the facts should be processed in order to be analyzed. In the case of a product evaluation, this might involve acting out a sentiment evaluation to evaluate whether the evaluation is good or bad. From there, a final outcome of % of excessively positive ratings might be obtained.

Veracity: Veracity can be termed as the first-rate or trustworthiness of the information. There is no use in acquiring Big Data if you aren't certain that the subsequent analysis can be trusted.

6Answer) CRISP DM FRAMEWORK:

There are 6 steps in this framework and it is mentioned below

1) Business understanding

2) Data understanding

3) Data preparation

4) Modeling

5) Evaluation

6) Deployment

Business understanding: the 1st stage in the framework. This is used to understand the business. This helps to examine the business goals and requirements of a organization.

Data understanding: 2nd step in framework. This is used to gather the data that client wants. like the requirements of the customer may be a cell number or name of a person etc. so with this step we can get what data is needed and what data is redundant.

Data preparation: we now have the data that’s required so we need to process the data and then we need to store it. So, in this step we process the data. This step uses extract, transform and load software. We first prepare the data.

Modeling:

This is the step where the machine learning model is being trained. We now got proper dataset and now using various tools we need to model the data. We can train the model in many ways and it can be supervised or unsupervised or semi-supervised. To get better efficacy we shouldn’t totally rely on machine model.

Evaluation:

In this step we test the model being developed and with the results we decide the next step. We will analyze the model here. If the result is as expected then we go to deployment step else we go to the first step and start from scratch again.

Deployment:

If the output of the model is as predicted then we are just left with deployment of the designed model. This is the step to deploy the model. After considering all the aspects the model has to be deployed. This process may take time as in real world we may not be satisfied with the output at the very first iteration.

7Answer) Python has several libraries as it is open source. Few of them are below.

OpenCV is a popular open source library that would be used for a variety of computer vision applications such as video analysis, CCTV footage analysis, and picture analysis. The collection contains over 2500 optimized algorithms, including a complete range of traditional and cutting-edge computer vision and machine learning techniques. These algorithms can also be used to identify and detect faces, identify objects, classify human behavior in videos, track camera motion, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high image resolution of an entire segment, identify similar images from an image database.

TensorFlow is a complete open source machine learning platform. It includes a rich, adaptable wide range of tools, modules, frameworks, and community resources that enable researchers to push the boundaries of ML and programmers to rapidly develop and execute ML-powered projects. Build and train ML models quickly and simply using intuitive high-level APIs like Keras and eager execution, which allows for instantaneous model iteration and uncomplicated debugging.

Numpy (Numerical Python) is another easy and complete library and is an excellent tool for doing simple and sophisticated array operations.  Many beneficial features are given by the library when performing operations on n-arrays and matrices in Python. It facilitates the processing of arrays that contain values of the same data type and makes math operations on arrays (and their vectorization) easier. In reality, vectorizing mathematical operations on the NumPy array type improves efficiency and shortens execution time.

Biopython :-An worldwide team of developers created Biopython, a collection of publicly accessible tools for biological computing written in Python. It is a global collaborative effort to create Python modules and apps to meet the demands of present and future bioinformatics work.