Data Science Application

❖ Machine Learning Prediction

- Artificial intelligence known as "machine learning," algorithms and models are created that can learn from data and generate predictions.
- From banking and healthcare to transportation and marketing, machine learning prediction is now a crucial tool in many fields of study and industry.

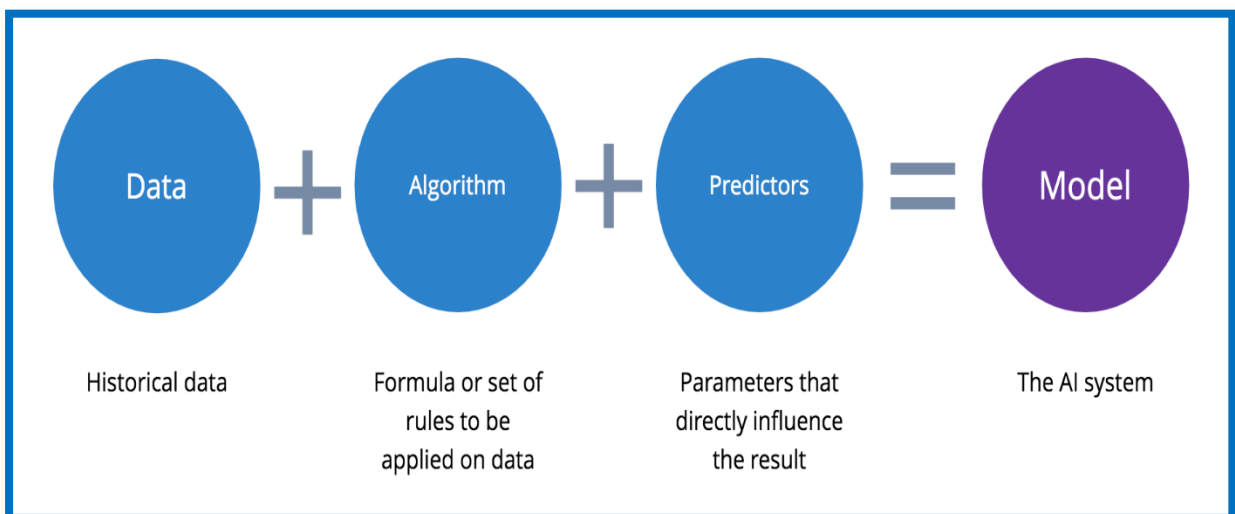## ❖ What is Machine Learning Prediction?

- The practise of using data to create predictions or foresee future events is known as **machine learning prediction.**
- These forecasts can be used to guide decisions, such as identifying the customers most likely to purchase a product, the individuals most likely to contract an illness, or the investments most likely to provide large returns.

Data Science Application

❖ <u>**How does Machine Learning Prediction Work**</u>

❖ **Machine learning prediction involves several steps:**

- Data gathering: Gathering pertinent data for the current issue is the first step in machine learning prediction. Many other sources, including sensors, surveys, and databases, can provide this data.

- Data preparation: After data has been gathered, it needs to be ready for analysis. This entails preparing the data for analysis by cleaning it, eliminating outliers or errors, and converting it to an appropriate format.

- Model development is the following action in the machine learning process. In order to do this, a suitable algorithm must be chosen, the input variables (also known as features) must be defined, and the model must be trained using the prepared data.

- Evaluation of the model: After the model has been trained, it is assessed using a different dataset known as the test set. This is done to make sure the model is not overfitting to the training data and is generalising well to new data.

- Once the model has been trained and assessed, predictions can be made using fresh, unused data.

Data	+	Algorithm	+	Predictors	=	Model
Historical data		Formula or set of rules to be applied on data		Parameters that directly influence the result		The AI system

Data Science Application

❖ Applications of Machine Learning Prediction

- Prediction using machine learning has several uses across various industries. Here are a few illustrations:

1. **Healthcare:** Doctors can concentrate their attention and resources by using machine learning prediction to identify which patients are most likely to contract a disease or experience complications.

2. **Finance:** Stock price predictions or profitable investment possibilities can be found using machine learning prediction.

3. **Marketing:** Businesses can customise their marketing strategies by using machine learning prediction to identify the customers who are most likely to purchase a product or respond to a marketing campaign.

4. **Transportation:** To help with route optimisation and wait time reduction, machine learning prediction can be used to forecast traffic patterns or demand for ride-sharing services.

5. **Weather forecasting:** Predictions based on machine learning can be used to forecast weather patterns and natural disasters, which can aid in emergency preparedness and response.
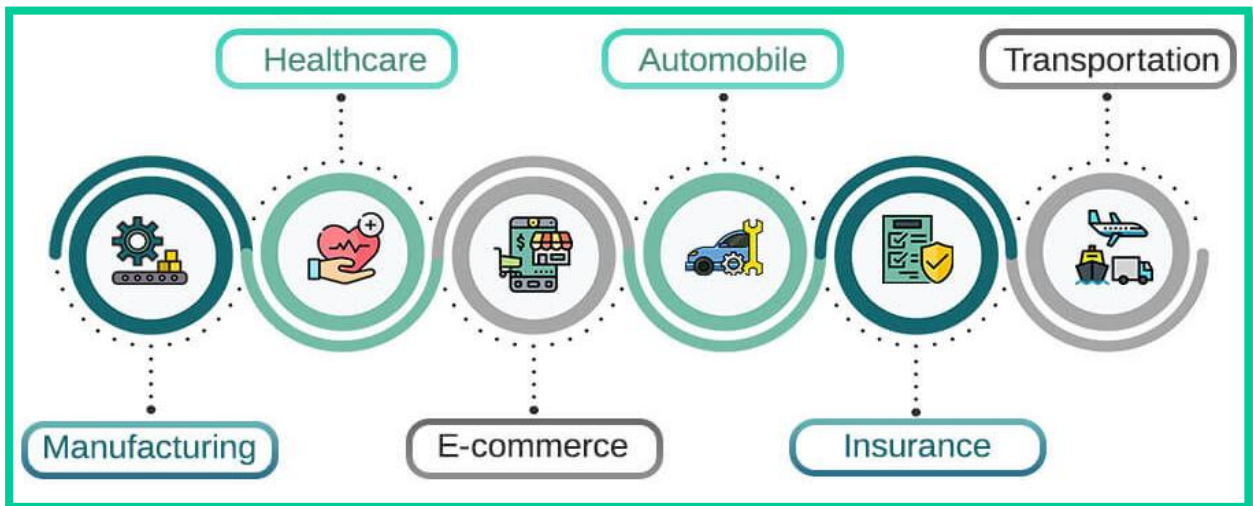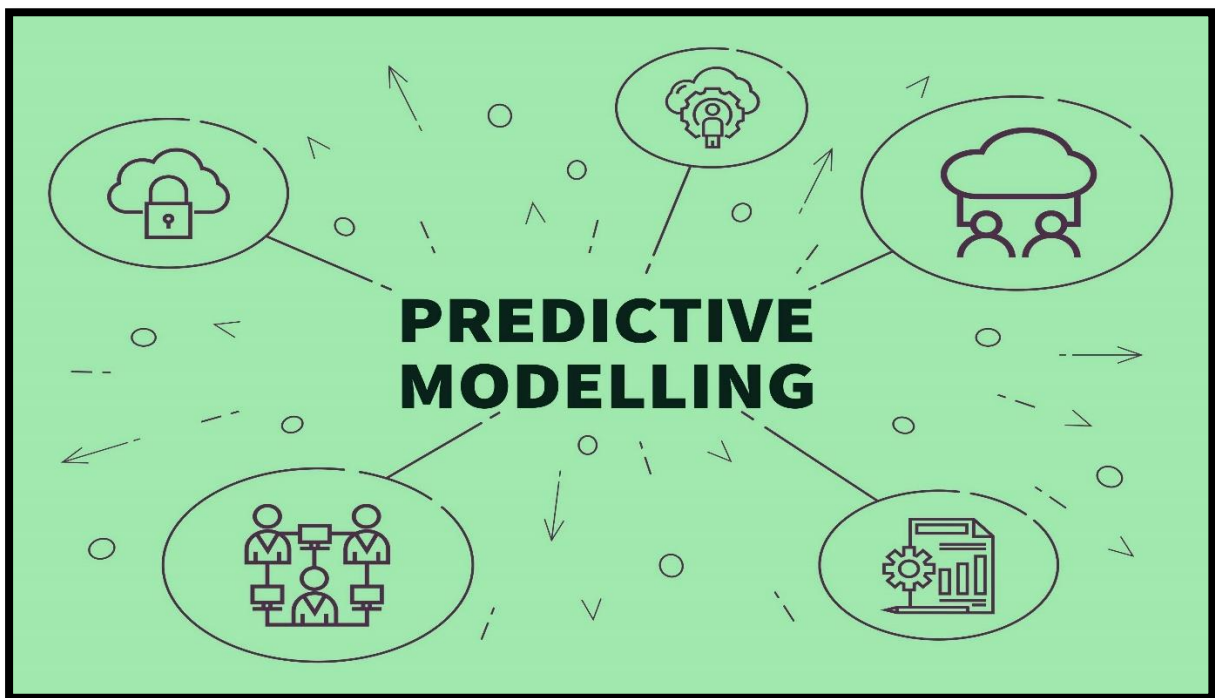
Data Science Application

❖ <u>Challenges in Machine Learning Prediction</u>

- Although machine learning prediction offers numerous advantages, it also has a number of drawbacks:

1. <mark>Data calibre:</mark> The data that machine learning models are trained on must be of high calibre. The forecasts will be erroneous if the data is unreliable, inaccurate, or biassed.

2. <mark>Model complexity:</mark> With several parameters and potential interdependencies between variables, machine learning models can be complicated. Due of its complexity, the model may be susceptible to overfitting and create difficult to interpret predictions.

3. <mark>interpretability:</mark> Deep learning-based machine learning models can be particularly challenging to understand. This can make it challenging to spot faults or comprehend why the model is making particular predictions.

4. <mark>Ethics:</mark> Prediction using machine learning can lead to ethical issues, particularly if it uses sensitive data or decision-making. For example, discrimination against particular demographics may result from the use of machine learning to anticipate criminal behaviour.

Data Science Application

❖ What is predictive modelling?

- Predictive modeling is <mark>a commonly used statistical technique to predict future behavior</mark>. Predictive modeling solutions are a form of data-mining technology that works by analyzing historical and current data and generating a model to help predict future outcomes.

- Predictive modelling is a mathematical process used to predict future events or outcomes by analysing patterns in a given set of input data. It is a crucial component of <u>predictive analytics</u>, a type of data analytics which uses current and historical data to forecast activity, behaviour and trends.

- Financial institutions use predictive modeling <mark>to evaluate loan risks, insurance claims, and potentially fraudulent activities.</mark> By predicting which transactions are likely to be fraudulent, businesses can reduce their exposure to financial losses.

- "Predictive modelling is a form of data mining that analyses historical data with the goal of identifying trends or patterns and then using those insights to predict future outcomes,"

Data Science Application

❖ **Top types of predictive models**

- There are many ways of classifying predictive models and in practice multiple types of models may be combined for best results.

- The most salient distinction is between <u>unsupervised versus supervised models</u>.

1. Unsupervised models use traditional statistics to classify the data directly, using techniques like <u>logistic regression</u>, time series analysis and decision trees.

2. Supervised models use newer machine learning techniques such as neural networks to identify patterns buried in data that has already been labeled.

- The biggest difference between these approaches is that with supervised models more care must be taken to properly label data sets upfront.

❖ **Top types of predictive models**

Data Science Application

❖ Some of the most popular methods include the following:

- **Decision trees.** Decision tree algorithms take data (mined, open source, internal) and graph it out in branches to display the possible outcomes of various decisions. Decision trees classify response variables and predict response variables based on past decisions, can be used with incomplete data sets and are easily explainable and accessible for novice data scientists.

- **Time series analysis.** This is a technique for the prediction of events through a sequence of time. You can predict future events by analyzing past trends and extrapolating from there.

- **Logistic regression.** This method is a statistical analysis method that aids in data preparation. As more data is brought in, the algorithm's ability to sort and classify it improves and therefore predictions can be made.

- **Neural networks.** This technique reviews large volumes of labeled data in search of correlations between variables in the data. Neural networks form the basis of many of today's examples of artificial intelligence (AI), including <u>image recognition</u>, smart assistants and natural language generation.

- **Predictive modeling algorithms include logistic regression, time series analysis and decision trees.**

❖ Common algorithms for predictive modeling

- **Random Forest.** This algorithm combines unrelated decision trees and uses classification and regression to organize and label vast amounts of data.

- **Gradient boosted model.** Similar to Random Forest, this algorithm uses several decision trees, but in this method, each tree corrects the flaws of the previous one and builds a more accurate picture.

- **K-Means.** This algorithm groups data points in a similar fashion as clustering models and is popular in devising personalized retail offers. It create personalized offers by seeking out similarities among large groups of customers.

- **Prophet.** A forecasting procedure, this algorithm is especially effective when dealing with capacity planning. This algorithm deals with time series data and is relatively flexible.

Data Science Application

❖ **where predictive models are used include the following:**

    I.   capacity planning
    II.   change management
    III.   disaster recovery
    IV.   engineering
    V.   physical and digital security management
    VI.   city planning

❖ **Benefits of predictive modeling**

- **Prioritizing resources.** Predictive modeling is used to identify sales lead conversion and send the best leads to inside sales teams; predict whether a customer service case will be escalated and triage and route it appropriately; and predict whether a customer will pay their invoice on time and optimize accounts receivable workflows.

- **Improving profit margins.** Predictive modeling is used to forecast inventory, create pricing strategies, predict the number of customers and configure store layouts to maximize sales.

- **Optimizing marketing campaigns.** Predictive modeling is used to unearth new customer insights and predict behaviors based on inputs, allowing organizations to tailor marketing strategies, retain valuable customers and take advantage of cross-sell opportunities.

- **Reducing risk.** Predictive analytics can detect activities that are out of the ordinary such as fraudulent transactions, corporate spying or cyber attacks to reduce reaction time and negative consequences.

❖ **The future of predictive modeling**

There are three key trends that will drive the future of data modeling.

1. First, data modeling capabilities are being baked into more business applications and citizen data science tools. These capabilities can provide the appropriate guardrails and templates for business users to work with predictive modeling.

2. Second, the tools and frameworks for low-code predictive modeling are making it easier for data science experts to quickly cleanse data, create models and vet the results.

3. Third, better tools are coming to automate many of the data engineering tasks required to push predictive models into production. Carroll predicts this will allow more organizations to shift from simply building models to deploying them in ways that deliver on their potential value.

Data Science Application

❖ Challenges of predictive modeling

Here are some of the <u>challenges related to predictive modeling</u>.

A. **Data preparation.**
One of the most frequently overlooked challenges of predictive modeling is acquiring the correct amount of data and sorting out the right data to use when developing algorithms. By some estimates, data scientists spend about 80% of their time on this step. Data collection is important but limited in usefulness if this data is not properly managed and cleaned.Once the data has been sorted, organizations must be careful to avoid overfitting.

B. **Technical and cultural barriers.**
While predictive modeling is often considered to be primarily a mathematical problem, users must plan for the technical and organizational barriers that might prevent them from getting the data they need. Often, systems that store useful data are not connected directly to centralized <u>data warehouses</u>.

C. **Choosing the right business case.**
Another potential obstacle for predictive modeling initiatives is making sure projects address significant business challenges. Sometimes, data scientists discover correlations that seem interesting at the time and build algorithms to investigate the correlation further.

D. **Bias.**
"One of the more pressing problems everyone is talking about, but few have addressed effectively, is the challenge of bias," Bias is naturally introduced into the system through historical data since past outcomes reflect existing bias.
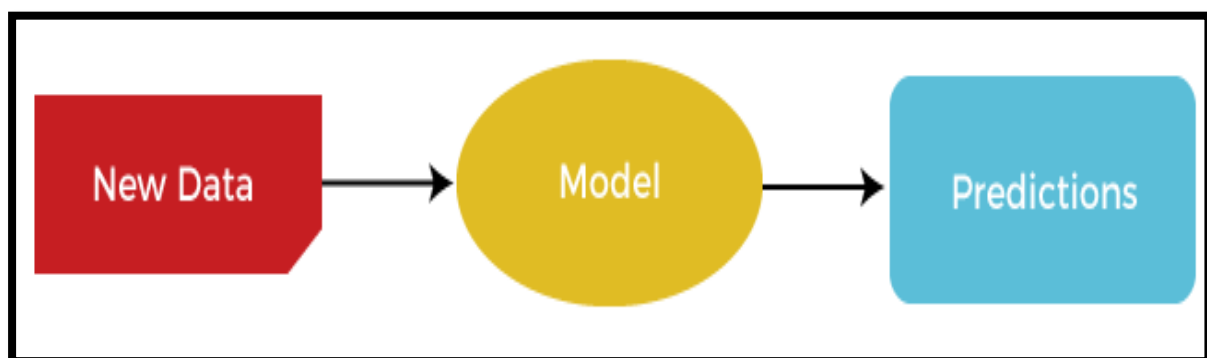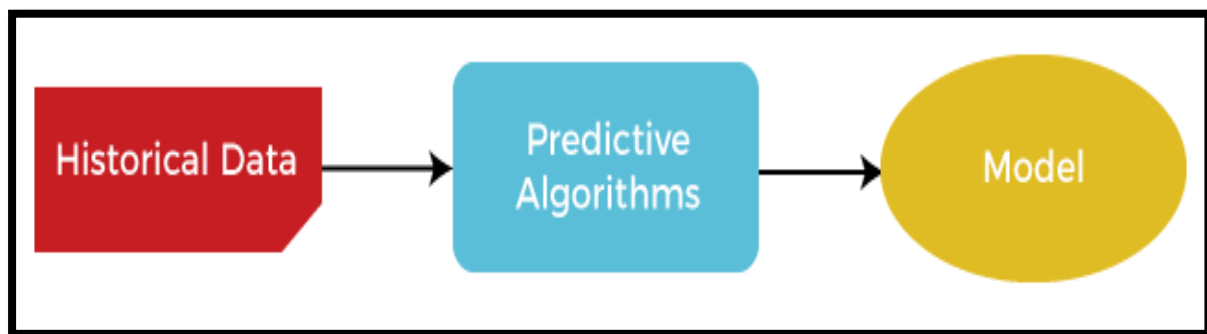
Data Science Application

<u>**Analytics**</u>

- primary categories of analytics software: <mark>predictive analytics and data mining.</mark> Predictive analytics refers to the use of both new and historical data, statistical algorithms, and machine learning techniques to forecast future activity, patterns, and trends.

- The primary objective is to go beyond knowing what has happened to assess better what will happen in the future. On the other hand, data mining refers to the computational technique of discovering patterns in huge data sets involving methods at the intersection of AI.

❖ **What are Predictive Analytics?**

- 'Predictive' means to predict something, so predictive analytics is the analysis done to predict the future event using the previous data.
- It is the process of extracting information from existing sets of data to find useful information, trends and forecast future events.
- Predictive analytics <mark>does not tell the exact thing</mark> that will happen in the future. It predicts what might happen in the future.

Data Science Application

❖ <u>**Predictive analytics with the help of an example;**</u>

- consider a hotel chain that wants to predict how many customers will stay in the specific location this weekday to ensure they have enough hotel staff and resources to manage the demand of the customers.

❖ <u>**Advantages of Predictive Analytics in business**</u>

1. Predictive Analytics Increase production efficiency.
2. It minimizes business risks.
3. It helps in decision-making purposes in any business organization.
4. It drives a competitive environment.

❖ <u>**Strategies of Predictive Analytics**</u>

- Based on the information gathered from predictive analytics, many companies have increased their turnover, achieved goals, and increased revenues by applying these strategies.

1. Synchronizing supply with demand.
2. Fraud prevention.
3. Creating lasting inventories.
4. Customer satisfaction.
5. Setting prices appropriately in order to increase profit.

Data Science Application

❖ **What is Data Mining?**

- Data mining refers to a process of analysing data from different contexts and summarizing it into useful information.
- The information gathered from data mining could include customer patterns, purchase patterns, transaction times, customer demand, and the relationship between the sold items.
- It is a powerful technology with great potential to assist companies in targeting the most significant information in the data set they have gathered about the customer behaviours and potential of the customers.

❖ **These are the given steps involved in the process of data mining**

1. Business Understandings
2. Data Selection
3. Data Preparation
4. Modelling
5. Evaluation
6. Deployment

❖ **Application of data mining**

1. Financial Analysis
2. Biological Data Analysis
3. Market Analysis
4. Retail Industry
5. Manufacturing Engineering
6. Criminal investigation

Data Science Application

❖ Predictive modeling versus predictive analytics

- Predictive modeling is but one aspect in the larger predictive analytics process cycle. This includes collecting, transforming, cleaning and modeling data using independent variables, and then reiterating if the model does not quite fit the problem to be addressed.

- "Once data has been gathered, transformed and cleansed, then predictive modeling is performed on the data," said Terri Sage, chief technology officer at 1010data, an analytics consultancy.

- Collecting data, transforming and cleaning are processes used for other types of analytic development.

- "The difference with predictive analytics is the inclusion and discarding of variables during the iterative modeling process," Sage explained.

- This will differ across various industries and use cases, as there will be diverse data used and different variables discovered during the modeling iterations.

- For example, in healthcare, predictive models may ingest a tremendous amount of data pertaining to a patient and forecast a patient's response to certain treatments and prognosis. Data may include the patient's specific medical history, environment, social risk factors, genetics -- all which vary from person to person. The use of predictive modeling in healthcare marks a shift from treating patients based on averages to treating patients as individuals.

- Similarly, with marketing analytics, predictive models might use data sets based on a consumer's salary, spending habits and demographics. Different data and modeling will be used for banking and insurance to help determine credit ratings and identify fraudulent activities.