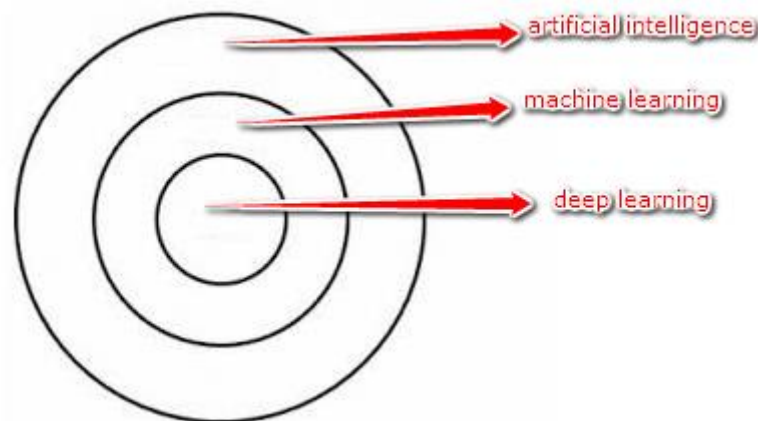# What is the difference between AI, ML and DL?

- In computer science, artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans. Leading AI textbooks define the field as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals. Colloquially, the term "artificial intelligence" is often used to describe machines (or computers) that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem solving". Modern machine capabilities generally classified as AI include successfully understanding human speech, competing at the highest level in strategic game systems (such as chess and Go), autonomously operating cars, intelligent routing in content delivery networks, and military simulations.

- Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop a conventional algorithm for effectively performing the task.

- Deep learning (also known as deep structured learning or differential programming) is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised or unsupervised. Deep learning architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases surpassing human expert performance. Artificial neural networks (ANNs) were

inspired by information processing and distributed communication nodes in biological systems. ANNs have various differences from biological brains. Specifically, neural networks tend to be static and symbolic, while the biological brain of most living organisms is dynamic (plastic) and analog.

What is Big Data, Data mining and Data Science?

- Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data. Data science is related to data mining and big data. Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyse actual phenomena" with data. It employs techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, and information science.

- Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use. Data mining is the analysis step of the "knowledge discovery in databases" process or KDD. Aside from the raw analysis step, it also involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. The term "data mining" is a misnomer, because the goal is the extraction of patterns and knowledge from large amounts of data, not the extraction (mining) of data itself.

- Big data is a field that treats ways to analyse, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software. Data with many cases (rows) offer greater statistical power, while data with higher complexity (more attributes or columns) may lead to a higher false discovery rate. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source. Big data was originally associated with three key concepts: volume, variety, and velocity. When we handle big data, we may not sample but simply observe and track what happens. Therefore, big data often includes data with sizes that exceed the capacity of traditional software to process within an acceptable time and value. Current usage of the term big data tends to refer to the use of predictive analytics, user behaviour analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set.

What is Business Intelligence and Product Intelligence?

- Business intelligence (BI) comprises the strategies and technologies used by enterprises for the data analysis of business information. BI technologies provide historical, current, and predictive views of business operations. Common functions of business intelligence technologies include reporting, online analytical processing, analytics, data mining, process mining, complex event processing, business performance management, benchmarking, text mining, predictive analytics, and prescriptive analytics. They aim to allow for the easy interpretation of these big data. Identifying new opportunities and implementing an effective strategy based on insights can provide businesses with a competitive market advantage and long-term stability. Business intelligence can be used by enterprises to support a wide range of business decisions ranging from operational to strategic. Basic operating decisions include product positioning or pricing. Strategic business decisions involve priorities, goals, and directions at the broadest level.

  Types of analytics:
  - Decision analytics: supports human decisions with visual analytics that the user models to reflect reasoning.
  - Descriptive analytics: gains insight from historical data with reporting, scorecards, clustering etc.
  - Predictive analytics: employs predictive modelling using statistical and machine learning techniques
  - Prescriptive analytics: recommends decisions using optimization, simulation, etc.

- Product intelligence is defined as an automated system for gathering and analysing intelligence about the performance of a product being designed and manufactured, such that this data is automatically fed back to the product managers and engineers designing the product, to assist them in the development of the next iteration or version of that product. The goal of product intelligence is to accelerate the rate of product innovation, thereby making the product and its owners more competitive and increasing customer satisfaction. Product intelligence is often applied to electronic products, but it is not necessarily limited to electronic products.
  Key points of this definition:
  - Product intelligence is a separate concept from artificial intelligence.
  - The focus is on the gathering of product performance, quality and test data.
  - There is an automated process for gathering information and converting it into intelligence.
  - The purpose of the exercise is to use the product intelligence to create the next, improved iteration or version of the product.

Product intelligence can also include two additional functions:

- An automated process for synchronizing all manufacturing locations involved in producing the product, so that production of the new version can begin immediately. This is typically accomplished by pushing product and test specification software to all the test stations in the manufacturing pipeline over a network or web connection. This accelerates the product's time-to-market.
- The automatic enforcement of quality manufacturing processes to ensure that products are manufactured correctly, according to the precise specifications of the product designers. This ensures quality and consistency across many manufacturing sites.

What is a GPU?

- A graphics processing unit (GPU) is a specialized electronic circuit designed to rapidly manipulate and alter memory to accelerate the creation of images in a frame buffer intended for output to a display device.

What is API?

- An application programming interface (API) is an interface or communication protocol between different parts of a computer program intended to simplify the implementation and maintenance of software. An API may be for a web-based system, operating system, database system, computer hardware, or software library. An API specification can take many forms, but often includes specifications for routines, data structures, object classes, variables, or remote calls. POSIX, Windows API and ASPI are examples of different forms of APIs. Documentation for the API usually is provided to facilitate usage and implementation. More recently, the term has been often used to refer to a specific kind of interface between a client and a server, which has been described as a "contract" between both - such that if the client makes a request in a specific format, it will always get a response in a specific format or initiate a defined action. This is a specialized form of API, sometimes defined as a Web API.

What is a Confusion matrix?

	Total population	True condition	
		Condition positive	**Condition negative**
Predicted condition	**Predicted condition positive**	**True positive**	**False positive**, Type I error
	Predicted condition negative	**False negative**, Type II error	**True negative**
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power $= \dfrac{\sum \text{True positive}}{\sum \text{Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \dfrac{\sum \text{False positive}}{\sum \text{Condition negative}}$
		False negative rate (FNR), Miss rate $= \dfrac{\sum \text{False negative}}{\sum \text{Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) $= \dfrac{\sum \text{True negative}}{\sum \text{Condition negative}}$

	Prevalence $= \dfrac{\sum \text{Condition positive}}{\sum \text{Total population}}$	Positive predictive value (PPV), Precision $= \dfrac{\sum \text{True positive}}{\sum \text{Predicted condition positive}}$	False omission rate (FOR) $= \dfrac{\sum \text{False negative}}{\sum \text{Predicted condition negative}}$	Positive likelihood ratio (LR+) $= \dfrac{\text{TPR}}{\text{FPR}}$	Negative likelihood ratio (LR−) $= \dfrac{\text{FNR}}{\text{TNR}}$
	Accuracy (ACC) $= \dfrac{\sum \text{True positive} + \sum \text{True negative}}{\sum \text{Total population}}$	False discovery rate (FDR) $= \dfrac{\sum \text{False positive}}{\sum \text{Predicted condition positive}}$	Negative predictive value (NPV) $= \dfrac{\sum \text{True negative}}{\sum \text{Predicted condition negative}}$	Diagnostic odds ratio (DOR) $= \dfrac{\text{LR+}}{\text{LR−}}$	F_1 score $= 2 \cdot \dfrac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$

HR Round

Tell me about yourself

Good morning/noon/evening,

Thanks for inviting me for the final interview.

My name is Navendu Shekhar. I am from district Sirsa, Haryana. I have completed my B.Tech. (Mechanical) from IIT Guwahati.

My father is a retired BSNL Employee, and my mother is a housewife. I have no siblings.

I love doing gym and running and have completed a half marathon. I was preparing for civil services examination, which helped in doing a lot of self-introspection. I have cleared SSC CGL twice and have an active interest in solving analytical problems. For the last 4-5 months, I am actively looking for a job as a business/product/data analyst.

Explain your Gap

Firstly, I was preparing for the civil services examination. UPSC has taught me the values of commitment and discipline. I kept myself committed despite having successive setbacks. Clearing UPSC prelims is as tough as clearing IIT-JEE and I cleared it thrice, most recently in 2019. I wanted to do my bit to serve the society at large. Secondly, I always imagined job as something more than an economic transaction. It has direct bearing on social and cultural aspect of my life like the place where I am going to live, people I going to meet with, topics on which I will keep myself occupied with. Finally, by my own mistake, I planned entry into a sector, without planning an exit. But returning to the job was my top most priority during last year and I did relevant certifications to make myself job ready especially in the field of analytics and data science.

Possible counter questions:

Have you done anything for the society till now?

- Though individual acts of social work must be appreciated, but I believe to sustainably influence the society, we need authority and resources. Civil services are one of the few options where it is a perfect combination of a job and a service.

What did you learn in your UPSC journey?

- There are many lessons that I drew from my civil services journey. Some of them are:
    - Don't lose hope. Worst case scenario, is one of the many possibilities. If we keep on doing the right thing, that eventuality will never come. The fear of uncertainty is always greater than the actual pain.
    - Incremental change is a sure shot way of revolutionary change in the longer run.

- o Always be clear with your priorities. I took geography as an optional, which was one of the biggest mistakes of my UPSC journey.
- o Do some ground research before committing oneself for a longer duration or before allocating significant resources. But don't do this endlessly, have a timeframe or a sunset clause.

Anyplace where we have influenced the team decision:

Being from a remote area, it becomes really hard to separate the real coachings from the advertisement version of them. With no one to guide me or my father in my career path, my father took a loan to enrol me in a bogus JEE coaching (CLC, Sikar). After being there for 3 months, I realised that they were running a sham. I told my father to withdraw me from that coaching and that, I will repeat my 11th class in some reputed coaching in Kota. Everyone advised me to do the otherwise, but I persisted. I convinced my father that I will take the blame, if things didn't pan out the way we intend them to. My father took a leap of faith, that too against the wishes of other family members, and enrolled me Resonance, Kota. I cracked JEE 2008 in the first attempt itself with an AIR -2504.

Anyplace where we have shown commitment:

My weight loss journey is something where I convinced myself that the small incremental changes over a long period of time will add to big revolutionary change in the longer run. I realised that it is a lifestyle issue and this journey is something that is more mental than physical. I put a treat and punishment mechanism to guide me through the journey. It also gave me a clarity of thought in the personal health dimension of my life and helped me setting my priorities straight. I started to choose nutrition over taste in food. I started monitoring my small harmful habits. In the longer run, I shed 53 kgs (at my peak) in last 3 years. I hope to continue this journey

Leadership

Leaders are known for their ability to motivate, guide and will to take unpopular but right decisions. As a core member of event management team of the annual Technical-Management festival of IIT-Guwahati, Techniche, we were told to clear the classrooms of a huge block at very short notice, as the labourers backed out at the last instance. I with other members of the team, with the help of our juniors did clear the room at such short notice, thus saving our institute from bad name.

Why Quickr

The role that is being offered to me in Quickr will help me in:

- Getting a real-world experience where I get a chance to apply my knowledge to real-world problems, about consumer behaviour, business perspective of retail and brands.
- I will get a bird eyes view about business and how companies are being run.
- I will get an amazing opportunity to learn new technologies like data visualisation (Tableau)
- As a start-up it will provide me:
    - More flexibility to work and lively environment.
    - More responsibility, faster learning and faster career growth.
    - Bringing a positive change in the lives of thousands of retail merchants and helping them to digitalise.

Guesstimates data

	Delhi	Haryana	Assam	India	bangalore	Noida	Gurugram
Population (in lakhs)	190	283	350	1.35 billion	85	6.4	10
Area (in sq. km.)	1500	45000	80000	328000	709	203	732
Literacy	86.21	76.64		75			
Sex ratio	868	879		940			
Per capita income (in lakhs)	3.65						1.5
Median Salary (in lakhs)							8
Total GDP (in billion $)	110	110	52	3000			
Density /sq. km.	11500	573	397	405			
District	11						
Sub division average	3						5
Road length (in 1000 km)	32						
Registered vehicles (in mill)	11						
Metro length	311						
Metro stations	214						
Percent less than than 18	32						
Percent between 18 and 40	40						
Percent between 40 and 60	20						
Percent more than than 60	8						
Male/Female	52/48						

How to solve a guesstimate

- It is the approach that matters – not the exact numbers. However, you should cross-validate numbers once you have them.
- Always go top down to solve a problem. Draw neat segmentation & diagrams to illustrate your approach.
- Keep a few common starting points / proxies on your fingertips. Population of your country, population across the globe, the GDP of your country are a few good starting points you should definitely remember.
- Analyse all possible uses of the subject. E.g. You should consider B2B & B2C markets, if you are asked to estimate market of tablets or smartphones.
- Call out assumptions and possible blind spots.
- Case study is all about illustrating 3 things – Structure, structure and structure! Focus on putting framework to the problem provided, and you will be safe. Try deviating from it and you'll find yourself in trouble.
- For example, when asked how can you increase Profits for a product company, you should not jump to conclusions like "I'll improve marketing or I'll cut costs". You should say Profits = Revenues – Costs. In order to increase profits, we can either increase Revenues or reduce costs. Revenues can be increased by increasing Sales or increasing the price. Costs can be reduced by doing ….
- Keeping a structure will not only help the interviewer understand you better, it will also help you make sure that you have not missed out anything.
- Call out assumptions, whenever you are making them. These could be assumptions about business or the sector in discussion.
- Lay out things neatly on paper, such that, they can be re-used later. Most of the times, case studies evolve over time. You will be asked to do similar questions, multiple times under multiple scenario. Keeping them handy can reduce calculation time!
- Think out loud – it is the thinking process, which matters. If you are not sure – ask the interviewer rather than staying quiet!
- Communicate crisply and clearly – if you are not clear about your thoughts, take 2 minutes from the interviewer to arrange your thoughts and then communicate them nicely

1. Find the right proxy: This is by far the most important technique. The proxy is a parameter which behaves in a similar manner as the dependent parameter. In the cigarette estimation problem, the population of India is a good proxy for the number of cigarettes consumed monthly in India. If the population of India increases, it can be safely said that cigarette consumption will increase proportionally. Other proxies used is the growth in population, growth in demand of a newly introduced technology, average number of planes parked at major airports etc.
2. Segment till you can find differentiated clusters: Estimating parameters on a segment level is far more accurate than making guesses on the overall population. In the cigarette

estimation problem, population below 16 years can safely be ignored for cigarette consumption and female population is expected to have a lower average cigarette consumption than male population. This is how segmentation helps making accurate assumptions.

3. Do smart calculations and number round off: Speed is very critical in such problems and one needs to maintain a balance between accuracy and time consumption. Say you need to fin 2999/3. It is much easier to calculate 3000/3 than 2999/3. In such cases right the answer as 1000 (-). This indicates the number is slightly lesser than 1000 and can be compensated in further calculations.

4. Validate number magnitude: It is always a good idea to keep on validating intermediate numbers using your experience and sense checks.

Some ground rules to be followed while doing a guesstimate

1. Following are some factors one should keep in mind while solving a guess estimate problem:

2. Analyse all possible uses of the subject. For example, while estimating the number of tennis balls in India, one should consider balls being used in tennis, cricket and all other sports which are potential users of tennis balls.

3. Keep population of your country, state and city on finger tips. As population is the most common proxy for many case studies, such numbers give a good starting point.

4. Have a look on some key parameters for airline management: Many of guess estimate problems are related to airlines. A sense on the number of flights which normally stays in major airports, time lag between flight take off etc. helps.

5. Draw neat diagrams to show the segmentation. This not only helps do calculations quickly but also makes it easier to redo the calculations on the segment level if required.

6. Don't do round off in the same direction. Such round off magnifies the error term. Putting a sign in front of rounded off number helps

Step-by step-approach for solving a guesstimate problem

Case 1: Estimate the number of cigarettes consumed monthly in India

Solution: A good proxy in such a problem is the population of India, i.e., 1.2 billion. Following is an effective way to segment this population:

	Population : 1.2 Bn (100%)								
Segment level I	Age above 22 yrs (60%)				Age between 16 & 22 yrs(10%)				Age <16yrs (30%)
Segment level II	Urban (20%)		Rural (40%)		Urban (3%)		Rural (7%)		
Segment level III	Male (11%)	Female (9%)	Male (25%)	Female (15%)	Male (1.5%)	Female (1.5%)	Male (4%)	Female (3%)	
Avg. cigarettes PM	30	15	5	2	20	10	2	1	0
Population	132000000	108000000	300000000	180000000	18000000	18000000	48000000	36000000	360000000
# cigarettes PM	3960000000	1620000000	1500000000	360000000	360000000	180000000	96000000	36000000	0
Total cigarettes	8.1 Trillion								

Guess

Following were the key considerations in building the segmentation and the intermediate guesses:

1. The rural population consumes far lesser cigarettes than urban because of the purchasing power difference.
2. Male consume more cigarettes than female in both urban and rural populations.
3. Children below 16 years consume a negligible number of cigarettes.
4. Male to Female ratio in Urban is closer to 1 than that of Rural.
5. Male to Female ratio in younger generations is closer to 1 than that of older. This is because of the increase in awareness level.
6. Bulk of population start smoking after getting into a job and hence the average number cigarettes are higher in older groups.
7. Total number of cigarettes from the supply side also come to around 10 Trillion, which gives a good sense check on the final number.

Case 2: Estimate the number of WhatsApp Android applications installed

Solution: A good proxy in this problem is the world population, i.e., ~7.2 Billion. Following is a possible approach to this problem:

Parameters	World Population : 7.2 Bn							
Segment Level I	Developed countries (80%)				Developing countries (20%)			
Segment Level II	Age < 10 yrs	Age btw 10 & 20	Age btw 20 & 40	Age >40	Age < 10 yrs	Age btw 10 & 20	Age btw 20 & 40	Age >40
% population	15%	15%	30%	20%	4%	4%	8%	4%
%Population using Android phone	0%	3%	10%	2%	0%	15%	20%	10%
Population with android phone	10,800	32,400,000	216,000,000	28,800,000	288,000	43,200,000	115,200,000	28,800,000
#mobiles/watsapp installed	5	2	4	20	3	2	4	10
#watsapp installed	2,160	16,200,000	54,000,000	1,440,000	96,000	21,600,000	28,800,000	2,880,000
Total watsapp installed	125 Million							

The actual number of Whatsapp installed on Android phone is slightly more than 100 Million. As can be seen from this example that guess estimates can be fairly accurate if we choose good segments and approximations.

Case 3: Estimate the number of tennis balls bought in India per month

Solution: A good proxy in this problem is the number of cities in India i.e. ~1700. The catch in this problem is to analyze where all can we use tennis balls. Once we have the number of tennis balls used monthly, we can easily find the number of tennis ball bought in a month using the lifetime of tennis balls.

Following is an effective way to segment this population:

Parameters	Possible Tennis ball usage							
Segment Level I	Tennis				Cricket			
Segment Level II	Urban				Urban			
Segment Level III	Metro	Tier-2	Small towns	Rural	Metro	Tier-2	Small towns	Rural
#cities	5	60	1600	5000	5	60	1600	5000
# sectors/cities	100	50	30	10	100	50	30	10
# grounds/sectors	5	3	2	0	50	40	30	10
# daily balls consumed	5	3	2		2	2	2	2
Total daily balls consumed	12500	27000	192000	0	50000	240000	2880000	1000000
Monthly ball consumption	4.4 Million							

Following were the key considerations in building the segmentation and the intermediate guesses:
1. Rural areas have negligible number of tennis courts.
2. Metro cities have the highest number of sectors.
3. For each sector in metro cities, the number of grounds for both tennis and cricket is higher. This is both because of the bigger area and the higher buying capacity in metros.
4. Number of balls consumed in metros per ground is higher because of the higher engagement in metros.

How many Pizzas are sold in Delhi on a particular day?

To start with let us divide the type to pizza stores
1. Branded one which includes Pizza Hut, Dominos, Papa jones
2. Small junk stores and good restaurants who serve pizzas too along with other cousins.

There would be around 40 pizza hut outlets in Delhi (excluding NCR) around 50 Dominos and 15 papa Jones. There could be around 200 small junk stores and some 100 restaurants severing pizza.

Now let's divide the days as holiday and working day.

on normal days at any store of pizza hut or dominos or papa Jones there would be around 15 people and each spending around 40 mins there and eats half a medium pizza so in 40 mins they approximately sell 8 pizzas: 5 mins/pizza (fair guess) these shops run from 10 AM to 10 PM and we can consider around 8 actual working hours when customers are present. So I a day they sell around 96 pizzas. For home delivery say there are 10 delivery boy and each takes on avg 40 mins to deliver a pizza and come back. We can take avg number of pizza delivered in one order be 1 and half medium size, then I whole day they can at max deliver 180 pizzas, considering the fact that every time not all delivery boys will be delivering lets assume 7.5 to be the approximation. Total delivered pizza in a normal day would be 135. Total sum is 240. Since papa Jones has less market penetration let's assume their sum to be half of this 115.

Total share from them: 40*230 + 50 *230 + 10*115.

On holidays and weekends pizza sell from restaurant will be 1.5 times the normal as demand is higher on these days.

It will come out to be 1.5*(40*230 + 50 *230 + 10*115)

Now from small food stall and multi cousin restaurant: there would be some restaurant which are famous for pizza (say 30) in locality where sell may go to 50 - 60 pizzas on normal days and 80 on holidays and others 20 pizzas on normal days and 30 on holidays (170 restaurants)

Total pizza sells on weekdays: 26750 nearly 27000 pieces

On holidays: it will come at around 39000-40000 pizzas

Alternate ways population of Delhi (except NCR) 1cr, out of then only 15% of the population will be interested in pizzas (mostly people between age 15 to 30 are regular pizza eaters which will constitute 30% population and out of them half doesn't like eating pizza), and avg person eats pizza once in 20 days and that too half of medium size so it stand at 37500.

(there would be people eating pizza every week and may be every alternate day but there would be people who eat pizza once in a month or two)

NOTE: above mentioned assumptions are just approximations and may not match with real data. Please ignore any grammatical mistake (if any).