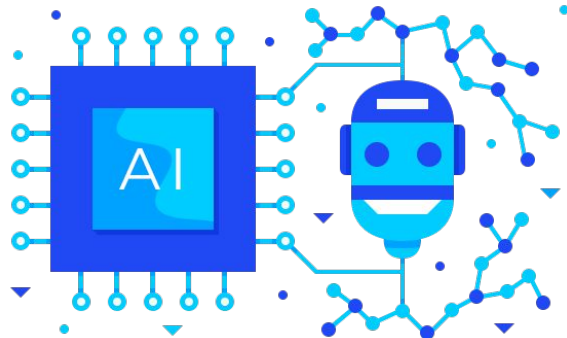# Data Science - The Backbone of AI

What is AI?

Is it the ability of machines to see?

Is it the ability of machines to make decisions?

Is it the ability of machines to understand our language?

Artificial Intelligence

"In computer science, artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and animals."

- Wikipedia

What is Data Science?

Data Science is using Data and combining algorithms to derive insights.

What comes under data science?

- Statistics
- Machine Learning (Algorithms that define patterns through learning data)
- Deep Learning
- Visualization

How is Data Science related to AI?

- The Data revolution started with internet being in our fingertips and the development of advanced computing.
- This forced the use of data in decision making, and helped develop algorithms based on data.
- Even though machine learning and deep learning have given cutting edge algorithms, classic statistics and data science techniques are still relevant today.
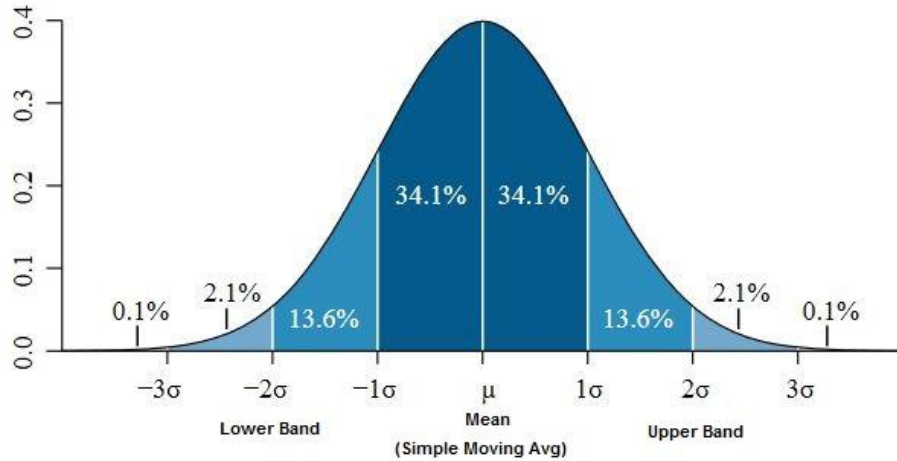
Statistics

Total Distribuition by survived or not
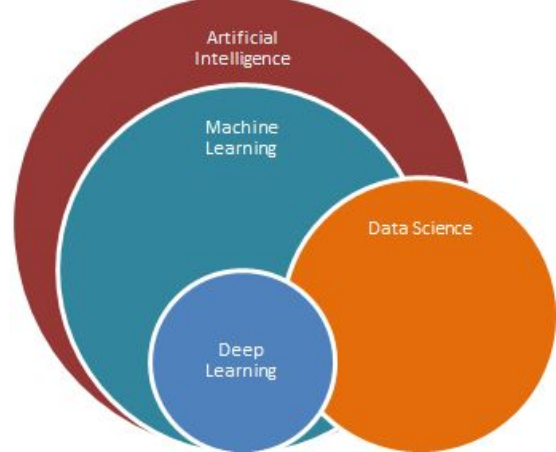
- Helps understand the data.
- Helps derive patterns in the data.
- Mean, median, standard deviation, are the most popular statistics measure for a data.
- Statistics are very useful when it comes to understanding a data. But statistics can be misleading too.
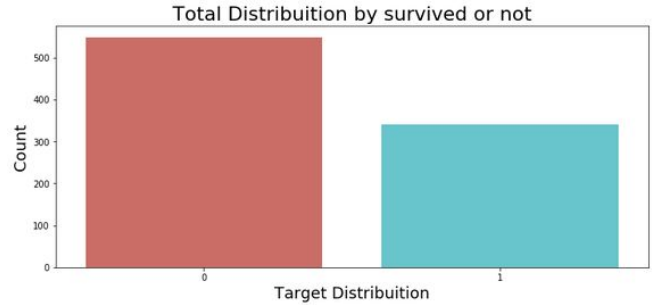
Imagine a college boasting an average placement offer of 7 lpa. What do you infer?

Before that, what is mean, median and STD?

*Not all distribution of data are in this shape.
Some examples are: Tossing a coin, height of men,
IQ of people, etc.

1, 3, 3, **6**, 7, 8, 9

Median = 6

1, 2, 3, **4**, **5**, 6, 8, 9

Median = (4 + 5) ÷ 2

= 4.5

The reality of statistics

Student Name	Company	Salary Expectation (lpa)	Salary Actual (lpa)
Student 1	Company 1	7	3
Student 2	Company 1	7	3
Student 3	Company 1	7	3
Student 4	Company 1	7	3
Student 5	Company 1	7	4
Student 6	Company 1	7	4
Student 7	Company 1	7	4
Student 8	Company 1	7	4
Student 9	Company 2	6	9
Student 10	Company 2	6	9
Student 11	Company 2	6	9
Student 12	Company 3	8	15
Student 13	Company 4	9	20

Statistics of Expectation	
Average Salary:	7
Standard Deviation:	0.8164965809
Median Salary:	7
Maximum Salary:	9
Minimum Salary:	6

Statistics of Actual	
Average Salary:	~7
Standard Deviation:	~5.3
Median Salary:	4
Maximum Salary:	20
Minimum Salary:	4

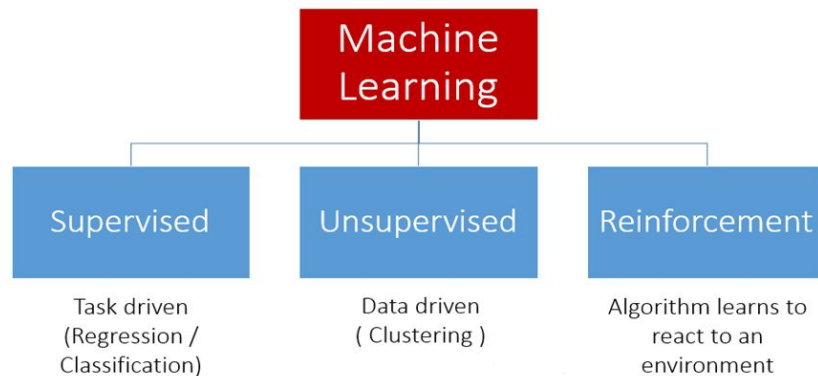# Advanced applications of statistics

- Finding correlations.
  - Exercise and heart diseases.
- Hypothesis testing.
  - Testing effectiveness of drugs.
- AB Testing.
  - Testing different designs
- Statistical Simulation.
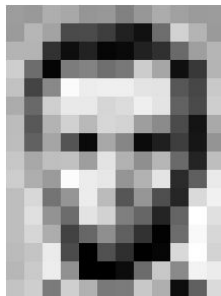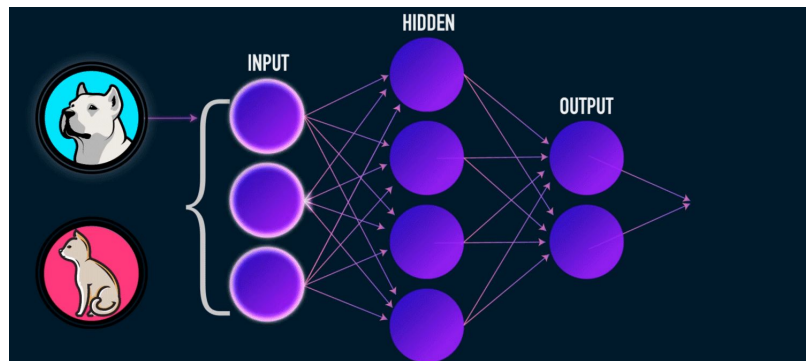- Machine learning.
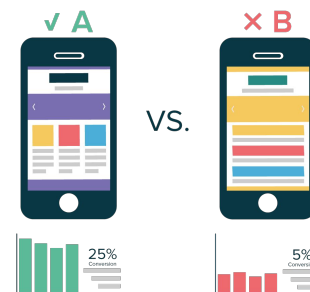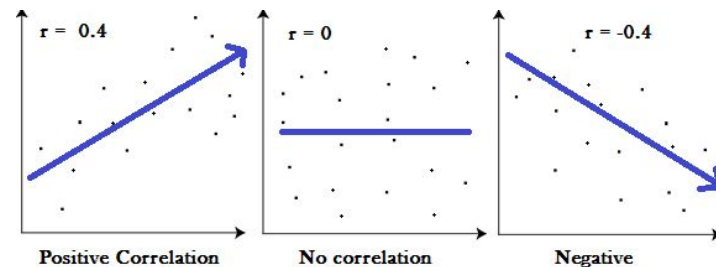
Machine Learning

- Programming machines to **learn** from the data.
- Supervised Learning.
  - Predicting Rainfall
  - Classifying cats and dogs
- Unsupervised Learning.
  - Grouping the customers in a shop.
- Reinforcement Learning.
  - Teaching machines to play a game.

Types of Machine Learning

Machine Learning

| Supervised | Unsupervised | Reinforcement |

Task driven (Regression / Classification)

Data driven (Clustering)

Algorithm learns to react to an environment

Deep Learning

- Models according to neurons.
- Each node will be representing a function and the layers of function will help us in defining complex problems.
- Breakthrough in computer vision was possible by combining this with Convolutional Neural Networks.
- Requires high computational power.

Skills and expectations

Data Scientist / AI Researcher

- Responsible for researching, developing, optimizing algorithms for a particular problem.
- Good mathematical intuition about algorithms.
- Thorough research knowledge.
- Experience expectations can be Masters/ PhD.
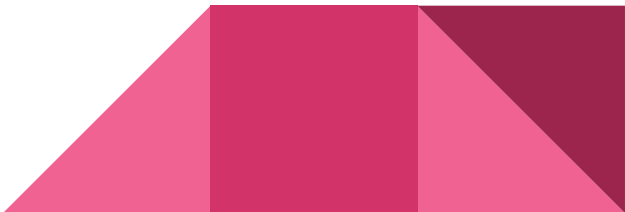- Adequate coding experience.

Data Engineer / AI Engineer

- Responsible for coding, deploying, existing algorithms to use in business.
- Also responsible for maintaining robust data pipelines and architectures.
- Good understanding of computer science fundamentals and an overview of data science concepts.
- Experience expected are Bachelors/Masters with adequate coding experience.

Tools used in Data Science and AI

How is machine learning and AI used in industry?

- In the age of rapid advancements, it is as important as electricity and internet.
- Hence the **Data Revolution**.
- The misconception of **Technology companies**.
- **The business problem should be thoroughly understood.**
- So adequate data collection infrastructure should be maintained.
- Data gathering, preprocessing and cleaning should be done thoroughly.
- Proper exploratory data analysis should be performed.
- Proper validation should be conducted.
- Proper deployment should be ensured.

Challenges for using AI and Data Science

- Availability of Data.
- Quality of Data.
  - Misleading patterns in the data.
  - Outliers in the data.
  - Missing Data.
- Overfitting.
- Unnecessary hype.
- Complicating simple problems.
- Computational complexity.
  - Deploying complex models on the edge.

Under- and Over-fitting examples
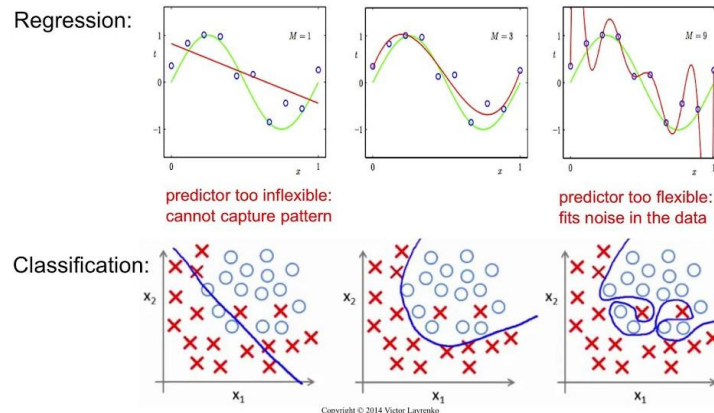
Expectation vs Reality
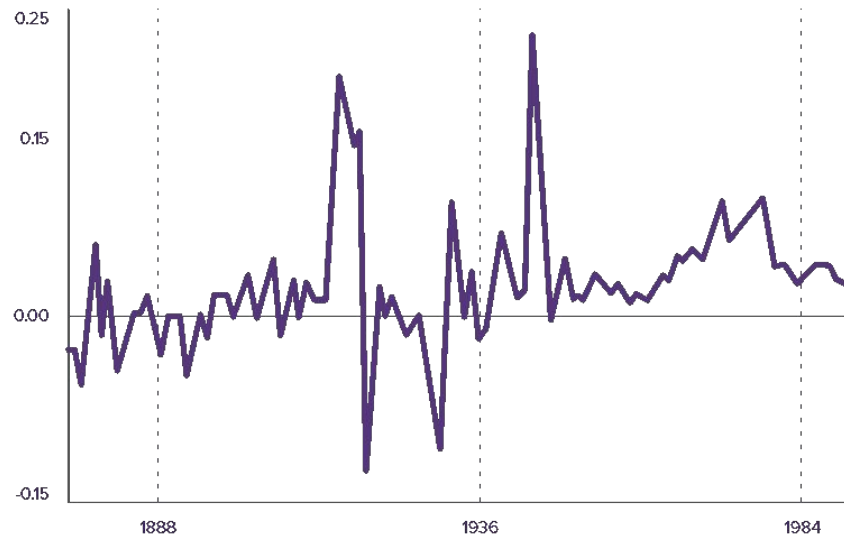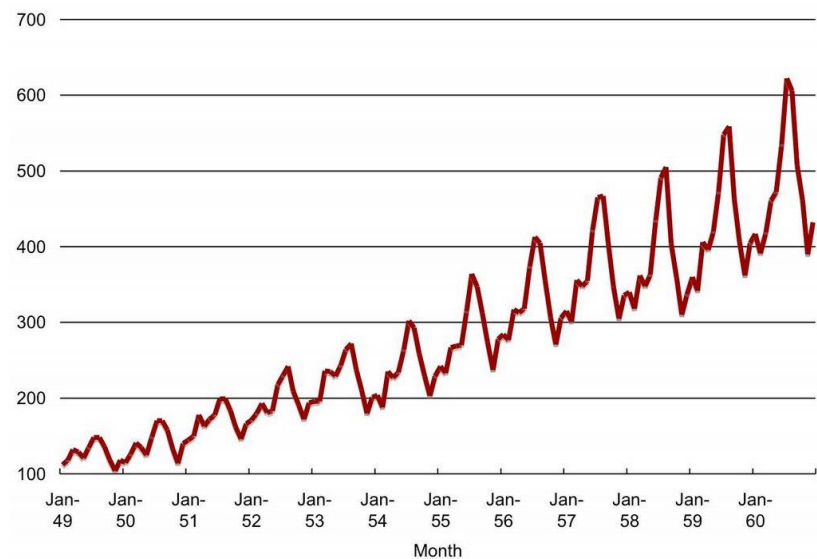
Dangers of AI

Dangers of AI

We expect a self aware robot, aiming for the extinction of mankind. But in reality, they are just men dressed in suits and controlling big corporations.

- Facebook-Cambridge Analytica data scandal.
- Targeted ads.
- Persuasive AI.

Protect yourselves by switching to platforms that protect your privacy like, Duckduckgo, protonmail, telegram/signal, etc.

Applications of Data Science in AI

- Analytics in various departments
- Demand forecasting and Supply chain management
- Medical drug discovery
- Medical Diagnostics
- Anomaly detection (internet security, social media, Visual surveillance)
- Stock market trading bots.
- Self Driving Cars.

Further resources for learning

- Kaggle
- Open courses on youtube
- Articles and research papers on Medium and other blogs
- Github for various code examples
- Communities

Doubts and Discussions

Further discussions

https://github.com/nevinbaiju/data_science_backbone_of_ai

Contact me at nevinbaiju@gmail.com

Thank you