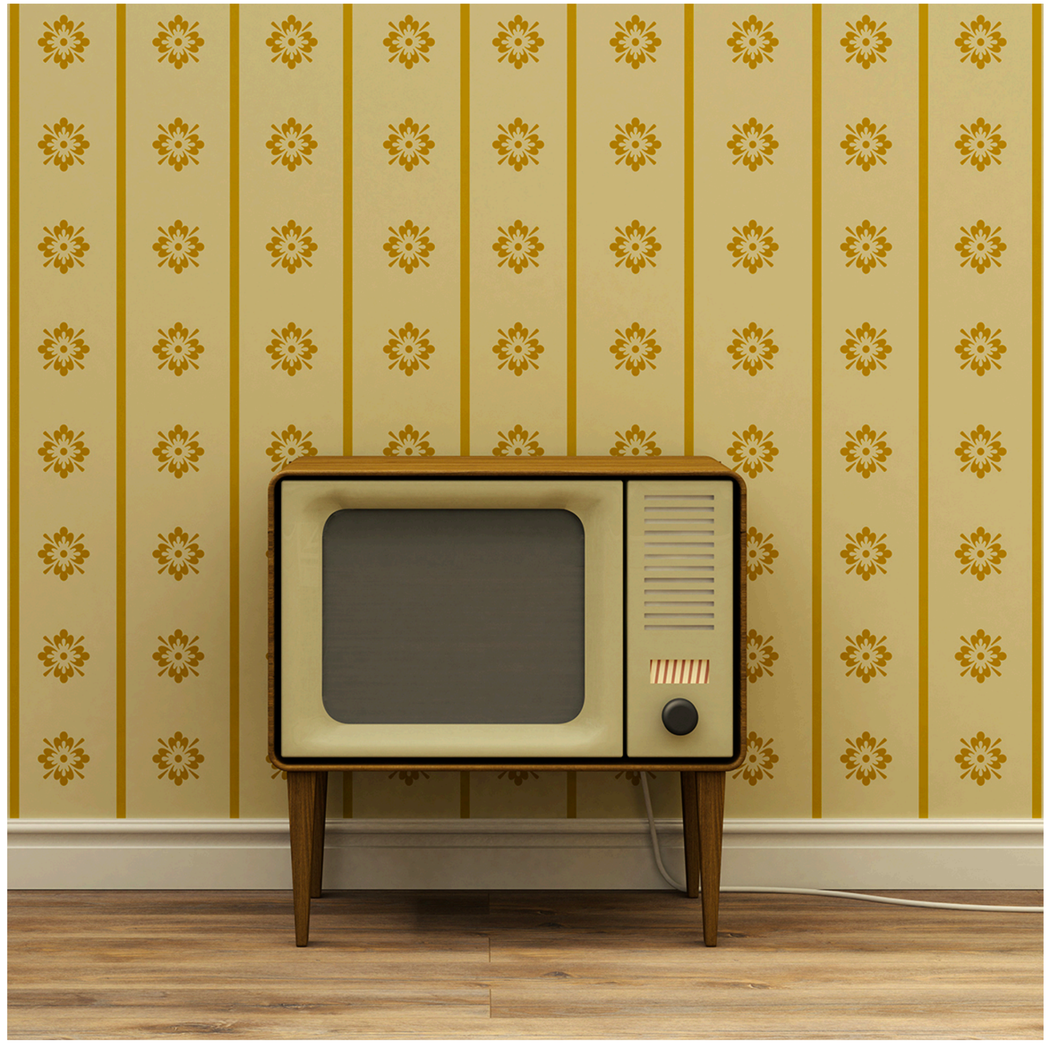
MACHINE LEARNING SESSION 1

Vaibhav Verdhan

Jan 8, 2021

AGENDA FOR THE SESSIONS

Session	Date	Agenda
1	Jan 8, 2021	Introduction to ML
2	Jan 15, 2021	Supervised Learning - 1
3	Jan 22, 2021	Supervised Learning - 2
4	Jan 29, 2021	Unsupervised Learning - 1
5	Feb 05, 201	Unsupervised Learning - 2

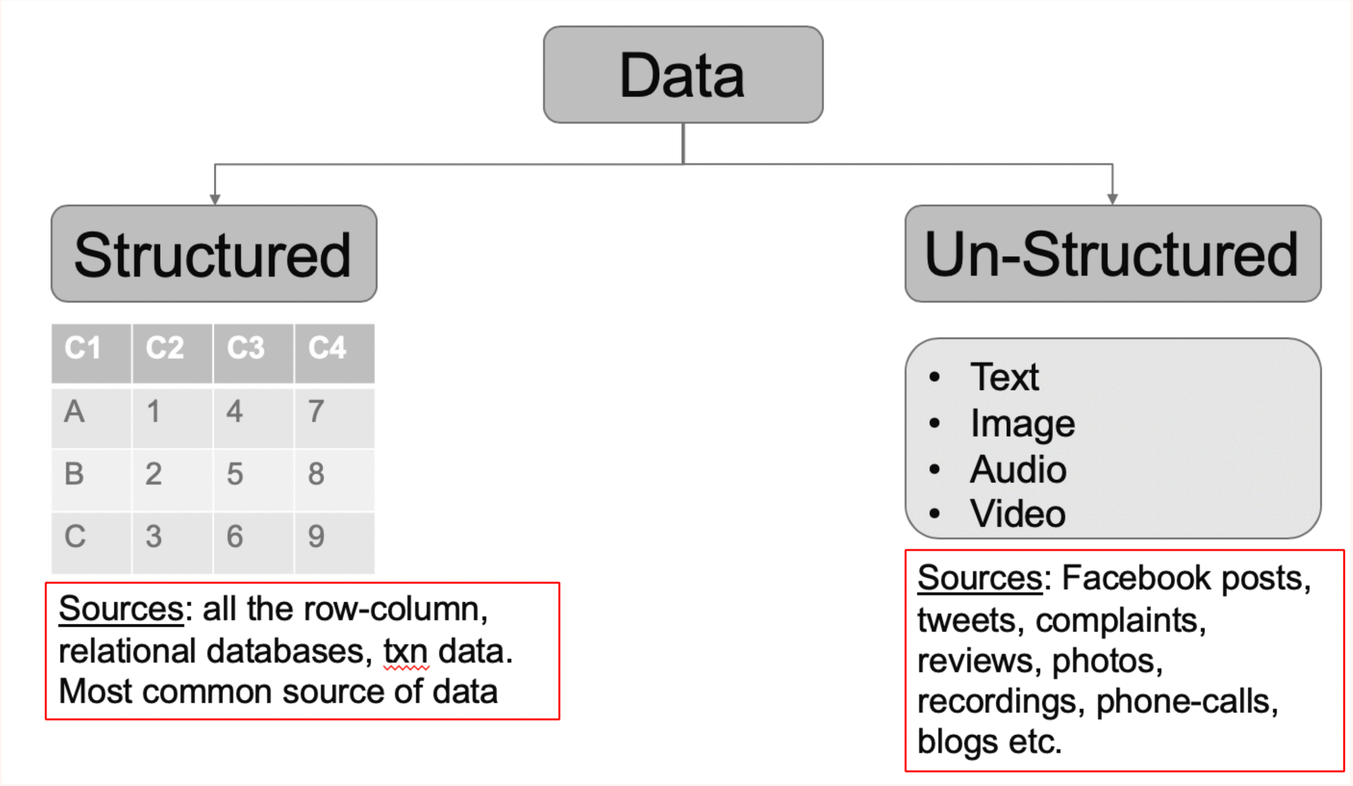


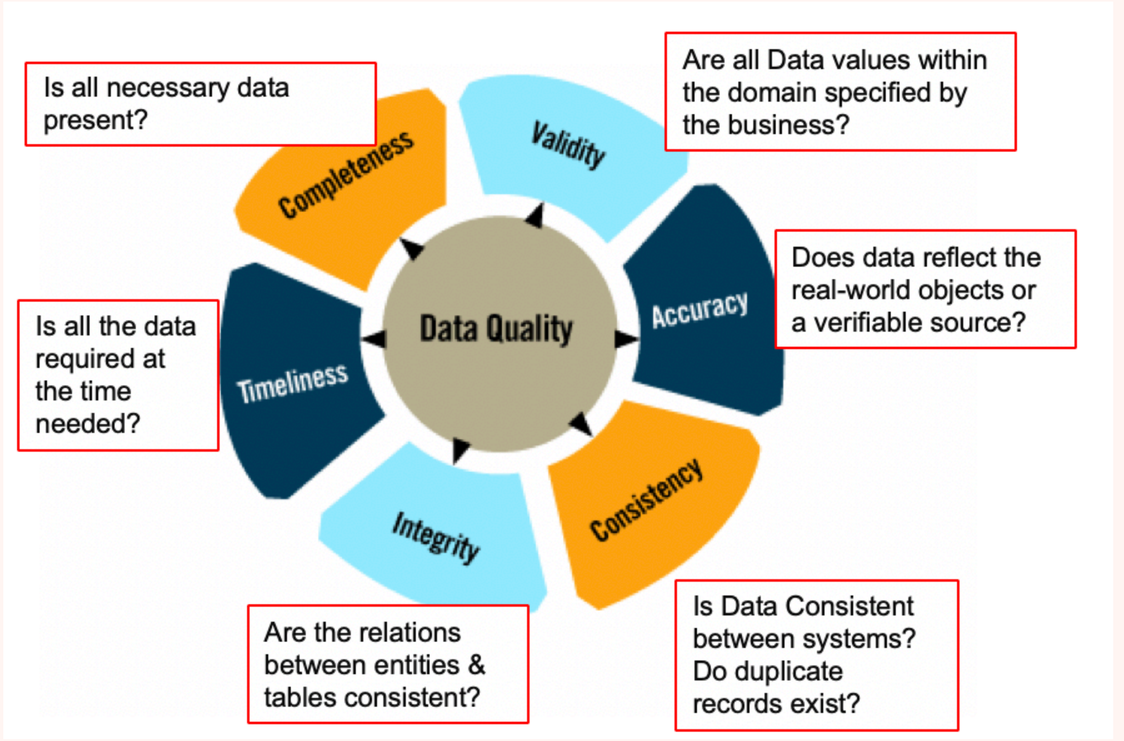
QUESTIONS WE WILL DISCUSS TODAY

- Machine Learning
- Difference between DA, DS, ML and AI
- Technical stack to learn
- Market Basket Analysis
- Recommendation system
- Python Implementations

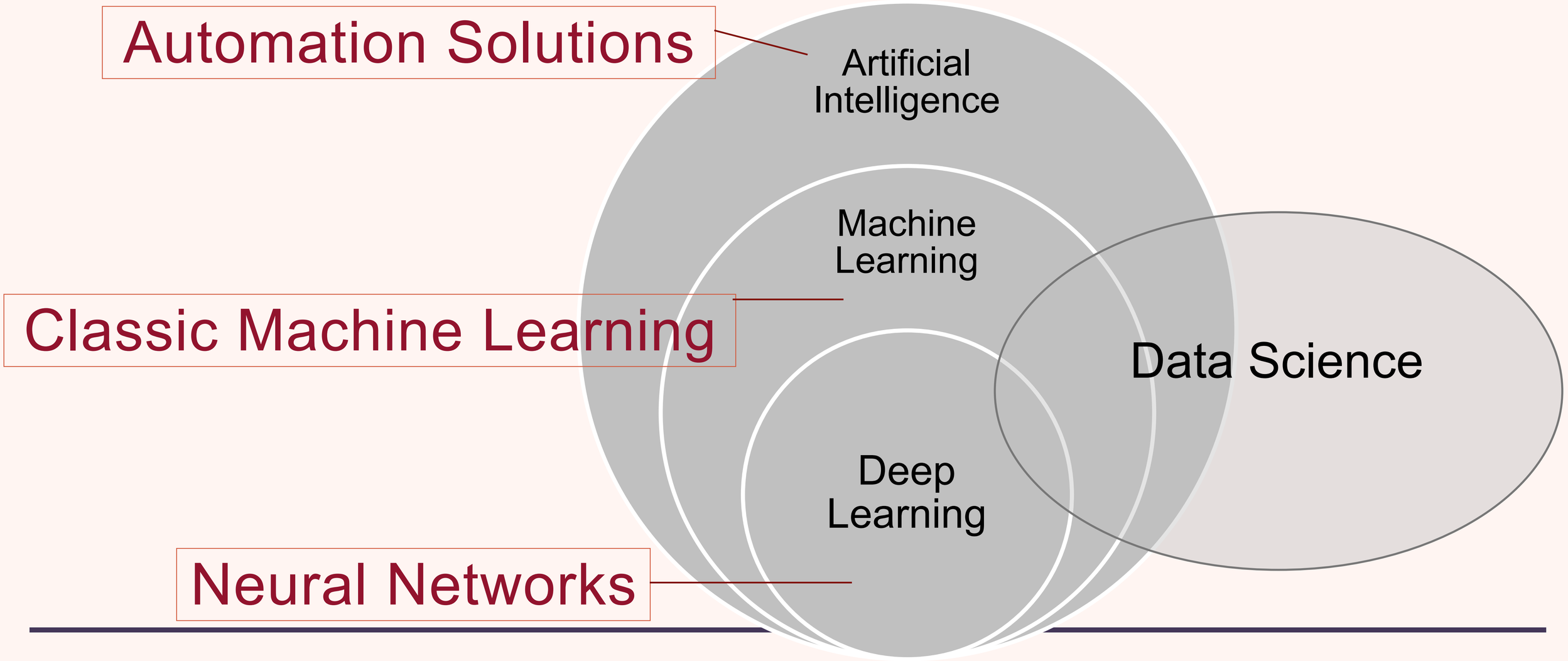
Data is the new oil in the current era & is shaping our world

The quality of data is going to define the insights you bring

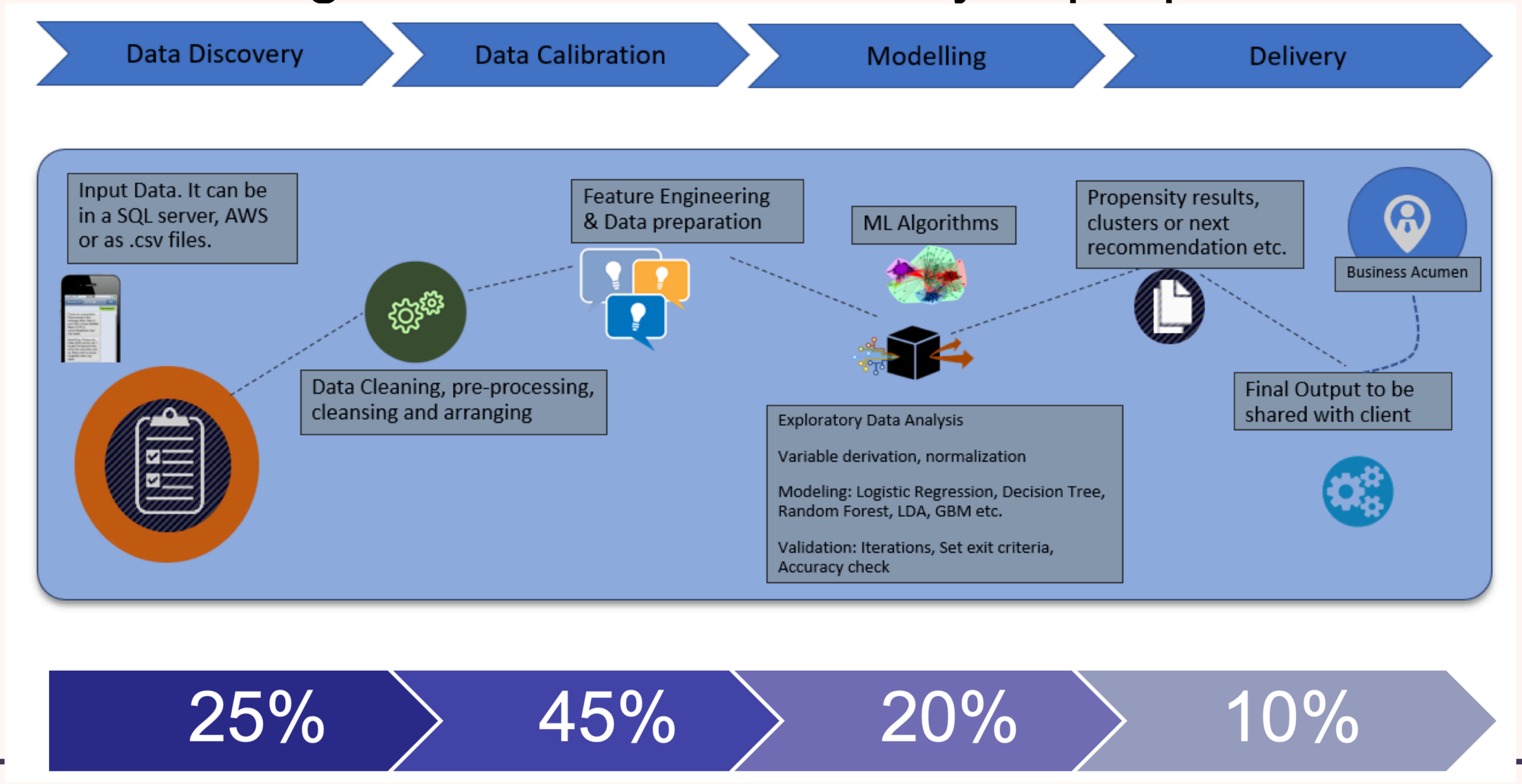




Data Science, Machine Learning & Artificial Intelligence



70% of time goes to Data discovery & preparation



Skills matter in long term. Tools are means to an end

Data Engineering: Spark, Hadoop, Spark, SQL, Redshift, Kafka, Java, C++

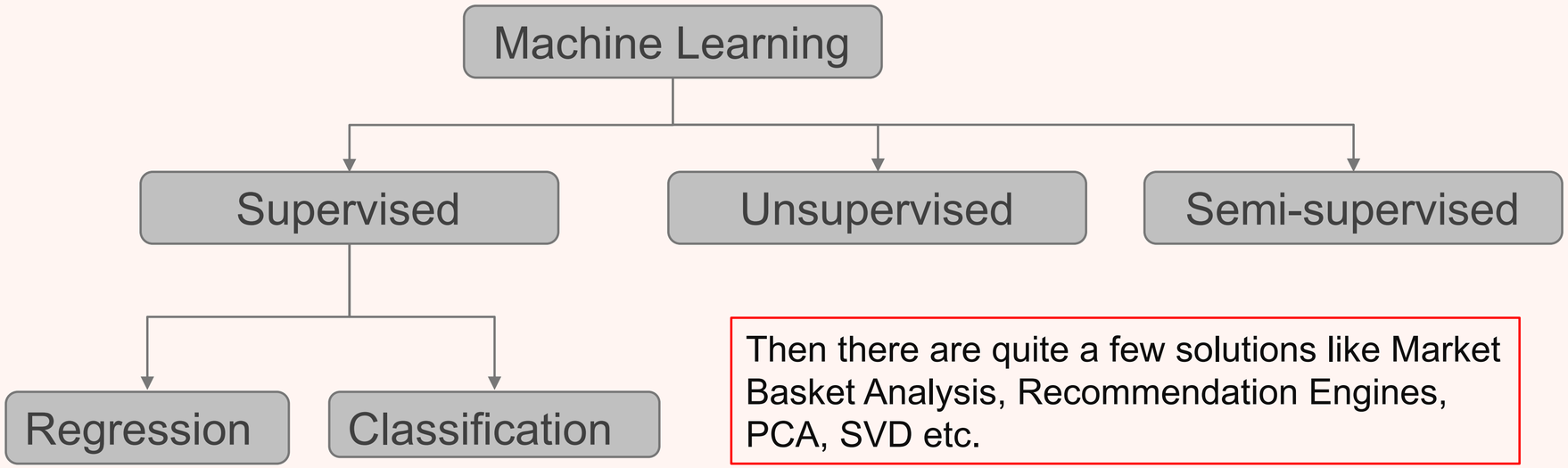
Data Analysis: Excel, SQL, postgres, mySQL, noSQL, R, Python

Machine Learning: SAS, R, Python, Weka, SPSS, Julia

Visualizations: Tableau, Power BI, Cognos, Qlik

Cloud Services: Microsoft Azure, Amazon Web Services, Google Cloud Platform

Stick to the principals and focus on the basics of ML



Then there are quite a few solutions like Market Basket Analysis, Recommendation Engines, PCA, SVD etc.

Algorithms generally used in Machine Learning:

Regression (Linear/Logistic)
Decision Tree
Random Forest
Gradient Boosting
Support Vector Machine
Neural Networks

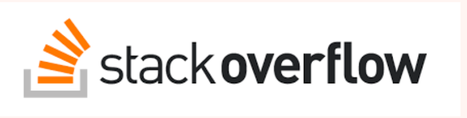
Clustering
Market Basket Analysis
Recommendation Engine
Principal Component Analysis

Tip for the interviews

- *Prepare at least 1 supervised classification project*
- *Prepare at least 1 supervised regression project*
- *Prepare at least 1 unsupervised learning project*
- *Review all the accuracy measurement parameters*

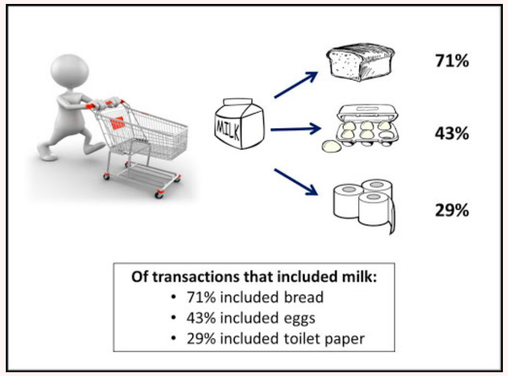
Story-telling blended with technical, pragmatic & soft skills

Technical skills	Practical skills	Soft skills
Programming	Self-starter	Communication
Data Analysis	Inquisitive	Gather domain knowledge
ML concepts	Accountable	Collaborative
Database/SQL	Ask questions	Inspirational



Make a very strong digital profile to get attention

Market Basket Analysis is used to get the relations



$$Rule:\ X \Rightarrow Y$$

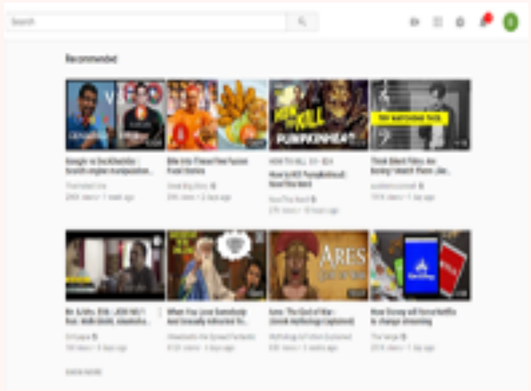
$$Support = \frac{frq(X,Y)}{N}$$

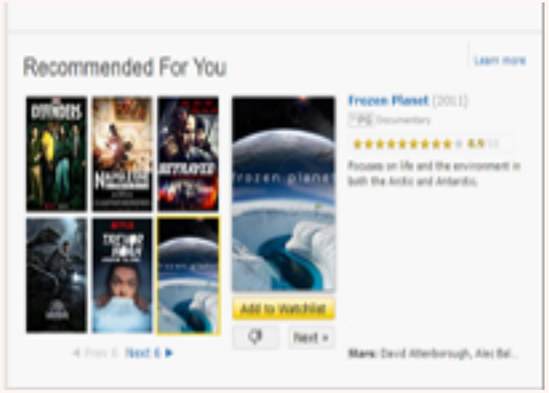
$$Confidence = \frac{frq(X,Y)}{frq(X)}$$

$$Lift = \frac{Support}{Supp(X) \times Supp(Y)}$$

Support Confidence and Lift are the three KPI which are used to assess the rules we derive

Recommendation Systems are quite powerful

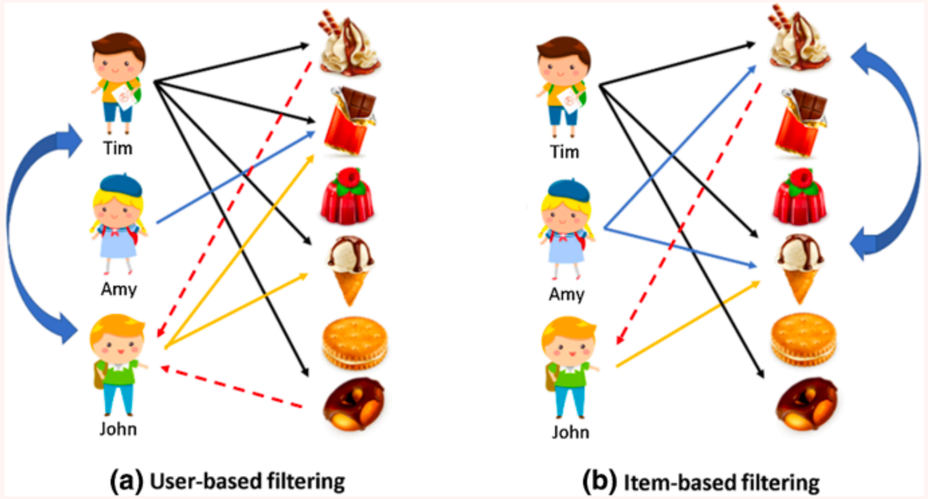




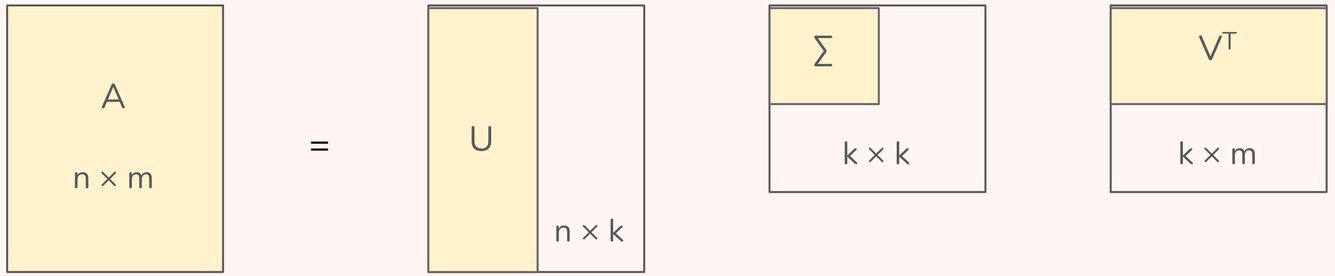
Recommendation systems are used to predict users interests and recommend products/services they are most likely to purchase and are interesting for them

There are various types of recommendation models

- Popularity based recommendation models
- Collaborative Filtering models
- Matrix Factorisation Methods of recommendation



Collaborative Filtering



Matrix Factorisation Method

Different Methods to measure the similarity

	Cosine Similarity	Pearson Correlation	Euclidean Distance
Formula	$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$	$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$	$d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$
Range	0 to 1	-1 to 1	>=0
When to use	When the data is sparse (many ratings are undefined)	When there is user-bias/ different ratings scales of users in the data	When the data is not sparse and the magnitude of the feature values is significant

Time to hit the code!

QUESTIONS PLEASE

THANKS!