

Applied AI – A Data Science Application Perspective



a presentation by

B.Sivaselvan

Associate Professor, CSE
IIITD&M Kancheepuram

In God we trust; all others must bring Data -- W.E Deming

AI vs ML vs DS

Artificial Intelligence

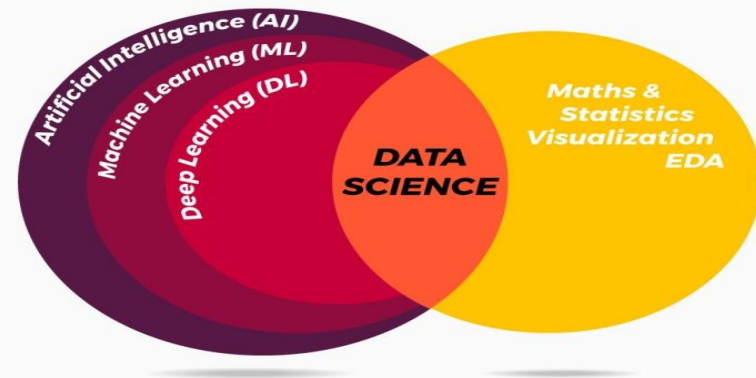
- Learn from data and make decisions
 - train computers to do things which human does better! (mimic human intelligence)

Machine Learning uses experience learnt for pattern search in test scenarios – need for big data – subset of AI

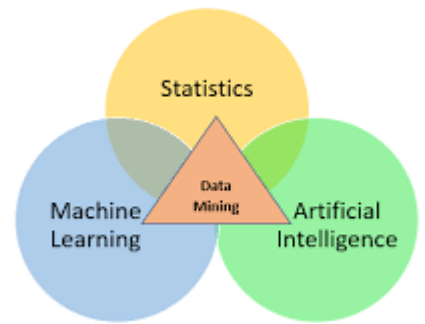
- AI uses experience to learn skills and to apply in new environments (optimally)

[knowledge behind patterns learnt]

Data Science – study of data – capture, maintain, process, analyse, communicate



Applied Flair of AI!



- **Data Mining** - Non-trivial process of extraction of hidden, previously unknown and potentially useful information from large databases.
- v/s DBMS ; {Statistics, Artificial Intelligence, Machine Learning, Databases & others.}
- **Data Analytics** – Process , Perform Stats Analysis of Data – how to use data for conclusions,
- **Big Data** – Huge Data Volumes – Data Capture, Storage, Sharing, Querying

“We are drowning in Data, but starving for Information”

- **Data Scientists** – Predicts future based on Patterns
 - Dev. Of New Artificial Methods, ML Models
- **Big Data Proff.** – Analyse System Bottlenecks
 - Large Scale Data Processing,
- **Data Analysts** – Acquire, Process, Summarize Data, package data for insights, design / create data report

Search Engines v/s **Finance** v/s **Healthcare** (non disjoint)

Skills – {**Python,ML,..**} {**Java,Scala,MongoDB,..**} {**R,Python, ..**}

MVP ! 123K v/s 88 v/s 61

Modern Apps & Analytics - Inseparable

- GMAIL – 2004 – overtook then successful Yahoo/Hotmail - monopoly – AI
 - Idea of Conversation
 - Spam Identification - classification
 - Primary, Social, Promotions
 - Priority Inbox - frequent patterns
 - Important / Unimportant – classification
 - Subject / Message Text is mined!
-
- ✓ "what times of the day do you send most of your emails" and "how long does it take you to reply to a particular person"

Modern Apps & Analytics - Inseparable

- Banking – Loan Risk Classification, Customer Retention
- Fraudulent Card Analysis – Outlier analysis
- Amazon – “They know what you want before you buy it”
 - ✓ Anticipatory Package Shipping (2012 Patent awarded)
 - ✓ shipping a package of one or more items “to the destination geographical area without completely specifying the delivery address at time of shipment,” with the final destination defined en route
 - ✓ Famous earlier books example – People who bought C++ also bought Java, etc.
- Netflix – Online Movies Portal
 - ✓ When you pause, rewind, or fast forward
 - ✓ When you pause and leave content (and if you ever come back)
 - ✓ Your ratings for shows ; Resulting Knowledge ;

Those who watched the British version “*House of Cards*” also watched Kevin Spacey films and/or films directed by David Fincher

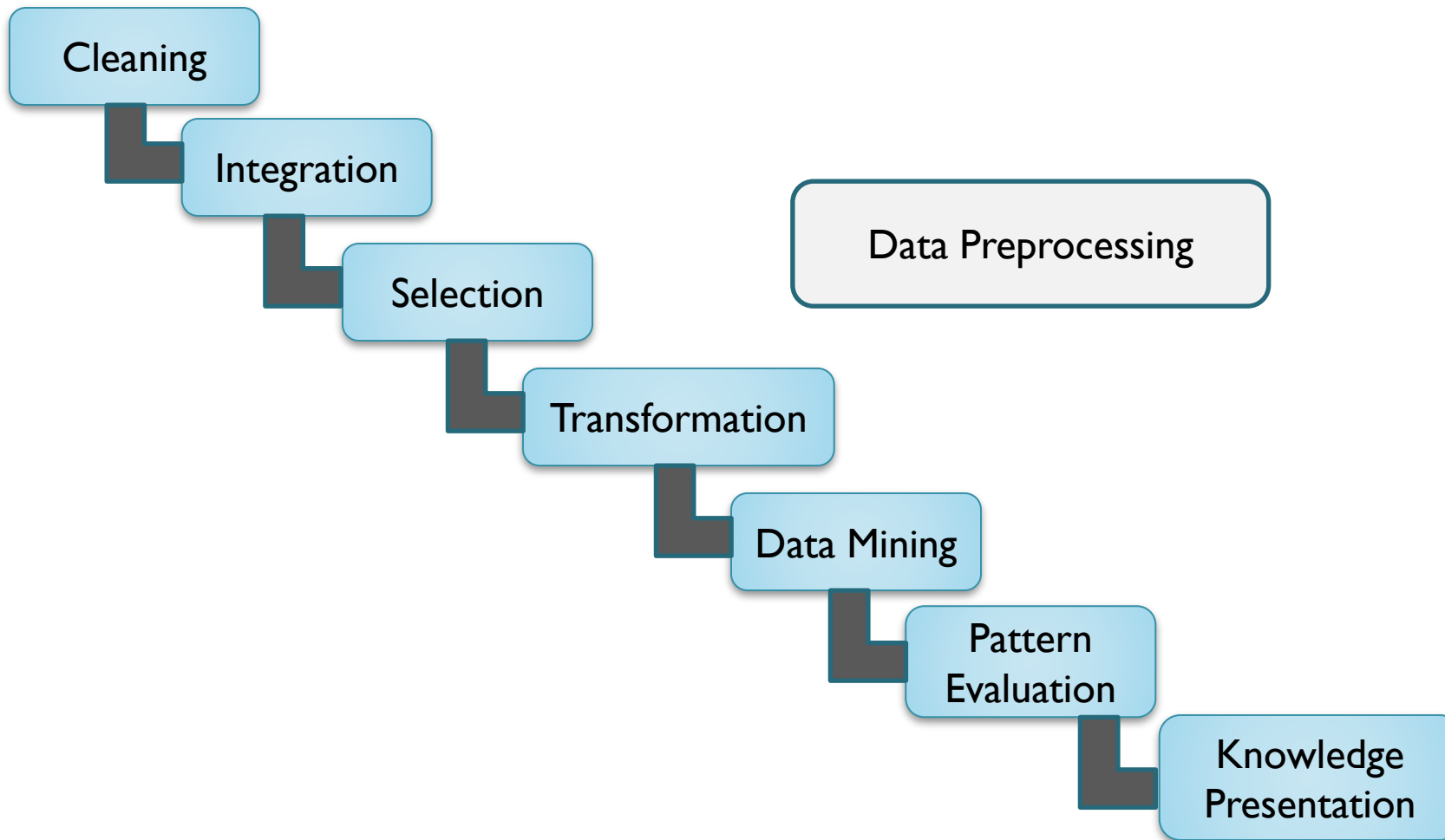
Modern Apps & Analytics - Inseparable

“Connect everybody in the world”

- People You May Know feature
- Social media analytics – search relevance, relevant advertisement, targeted ads (ur cell phone offers!), user profiling
- Uber - data mining / car service company ?
- -:“This is Your Digital Life” – does it need intro?
- Healthcare – Measuring Treatment Effectiveness
- Pharmacovigilance (PV) – ADR / Drug Safety Surveillance
- Best Mining Product – Human Brain!





Data Mining





- Terminology – Misnomer!
- Data Mining Phases






Data Mining Techniques

- Data Classification
- Association Rule Mining
- Clustering
- Prediction
- Outlier Mining
- Multimedia Data Mining – Text / Audio / Video

Name	Image	Affiliation	Major Contributions
Dr Jeffrey D Ullman		Stanford Prof	DIC Algorithm, Mining Massive Data Sets Textbook
Dr Rajeev Motwani (late)		Stanford Prof	Google Founders (Page and Brin's Mentor) "Today, <u>whenever you use a piece of technology, there is a good chance a little bit of Rajeev Motwani is behind it</u> "
Dr Jiawei Han		UIUC Prof	FPGrowth Algorithm, Data Mining Book,
Dr Sanguthevar Rajaseakaran		Univ of Connecticut Prof	Trans Map Algorithm for FPM, Famous data structures text with Sahni

Larry Page and Sergey Brin	 A photograph of Larry Page and Sergey Brin, the founders of Google, smiling and standing together. The text "Google Founders" is visible at the bottom of the image.	GOOGLE	This table details was searched in Google! Google is a word in the Oxford Dictionary
Prof. Krithi Ramamritham	 A black and white portrait of Prof. Krithi Ramamritham, a man with a mustache wearing a suit and tie.	Major Bhagat Singh Rekhi Chair Professor IIT Bombay	VLDB Association,
Prof Rakesh Agrawal	 A color portrait of Prof. Rakesh Agrawal, a man with glasses wearing a light-colored shirt, against a blue background.	Wisconsin Madison Prof	Apriori – Landmark contribution, IBM Almaden Research Labs,
Prof Raghu Ramakrishnan	 A color portrait of Prof. Raghu Ramakrishnan, a man with glasses and a mustache, smiling, wearing a light-colored shirt.	Wis Mad Prof	Famous book on Databases, many contributions in data mining

Jeff Bezos		Amazon.com	Online Retail Giant, American Tech Entrepreneur
Jan Koum		WhatsApp Founder	WhatsApp Founder Was with Facebook
Dr Osmar Zaiane		Albrerta Univ Prof	Han's Scholar, landmark contributions in multimedia data mining,

This is definitely not an exhaustive list but my recall of some landmark names
in Data Mining! Plz do google for more!

Some of the projects our team pursues

- Perspectives of Data Mining - Induction, Compression, Querying, Approximation, Search
- - search, compression and approx pursued in depth
- - Novel Text and Image Compression Algorithms – FPM based [Graduated Ph.D Thesis]
- Data Mining over Pharma data – Adverse Drug Reaction Prediction Models
- Emotion Recognition Framework for Suicidal Tendency Detection
- Unsupervised approaches for Video Summarisation – Educational Videos focus
- Online Course Moderation based on ML models – Karam
- Fake News Detection in OSN's – social engagement and visual content based