

DATA SCIENCE

PREMANAND S

OVERALL MODULES

- Data Science Fundamentals
- Statistical Analysis and Business Application
- Numeric Computing with Python
- Data Manipulation, Collecting and Feature Engineering
- Scientific Computing with Python
- Encapsulating the data
- Data Visualization & Types of plots
- Tools and Techniques - Tableau

OVERALL MODULES

- Development of Model
- Supervised and Unsupervised Learning
- Dimensionality Reduction
- Evaluation Model
- Web Scarping and Natural Language Processing
- Recommender System
- Neural Nets and Deep Learning Overview
- Capstone Projects – Classification | Regression | NLP

Module 1: DATA Science Fundamentals

What we are gonna see in Module 1?!

- Technology and its existence
- Difference between AI – DS – ML – DL – NLP – RS
- Def & Purpose of AI – DS – ML – DL – NLP – RS
- Data Science life cycle
- What is DATA and its types?
- Data Science Importance
- Data Science – Platform
- Why Python is best?

What we are gonna see in Module 1?!

- Personalities to become Data Science
- Application of Data Science
- Skills needed to become a Data Scientist
- Linux for Data Science
- Version Control

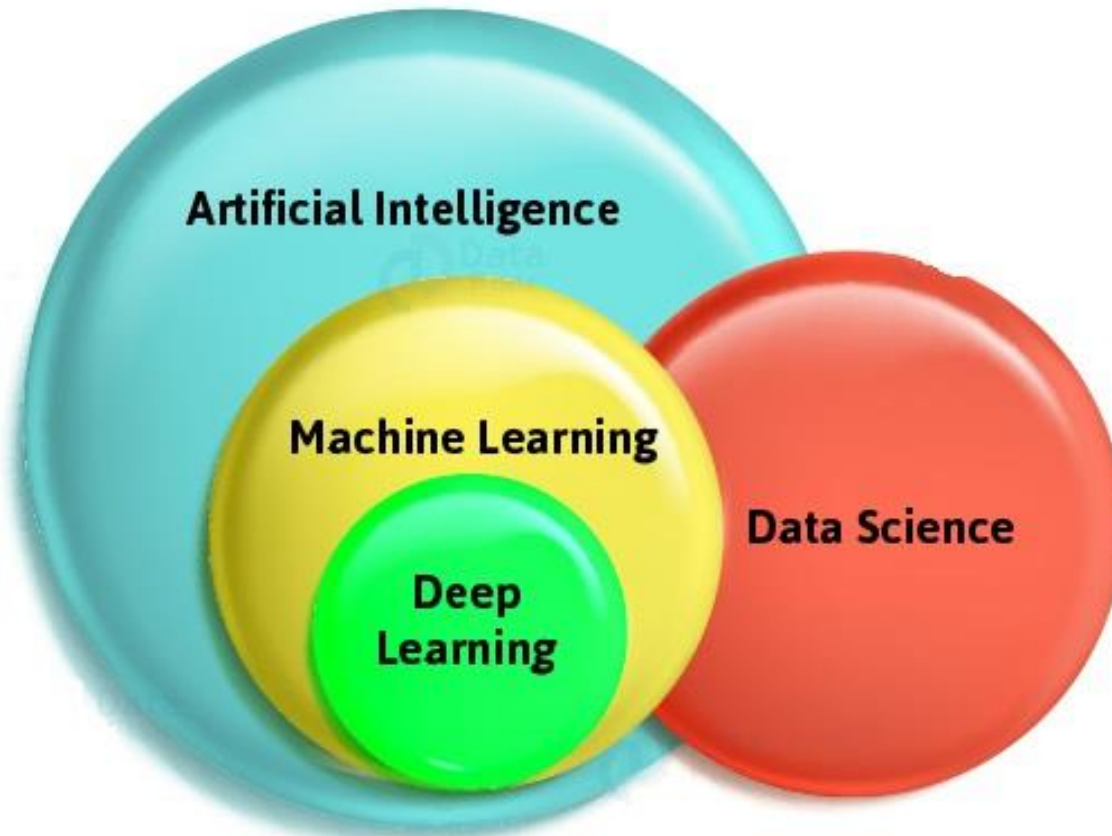
Technologies?

- Artificial Intelligence
- Machine Learning
- Deep Learning
- Internet of Things
- Augmented Reality
- Cognitive Computing
- Blockchain
- Additive Manufacturing – 3D Printing
- Robotics & Automation
- Renewable Energy Technologies and still many more...

Why Technologies?

- Error-less when compared to human calculation
- Makes our life easy
- Improves efficiency
- Improves communication
- Easy access to information
- Connectivity and Communication
- Saves time
- Productivity

Technology difference – Better understanding!



Data Science - Def

Data science is a multidisciplinary field that combines scientific methods, algorithms, and systems to extract knowledge and insights from structured and unstructured data. It involves various techniques such as data mining, statistics, machine learning, and visualization to analyze data, identify patterns, make predictions, and support decision-making.

Artificial Intelligence - Def

Artificial intelligence refers to the development of computer systems that can perform tasks that typically require human intelligence. AI encompasses a wide range of techniques, algorithms, and methodologies that enable machines to understand, reason, learn, and adapt. It includes areas such as natural language processing, computer vision, expert systems, and robotics.

Machine Learning - Def

Machine learning is a subset of AI that focuses on developing algorithms and models that enable computers to learn and make predictions or decisions without being explicitly programmed. ML algorithms learn from data, identify patterns, and generalize from examples to improve performance on specific tasks. It includes techniques like supervised learning, unsupervised learning, and reinforcement learning.

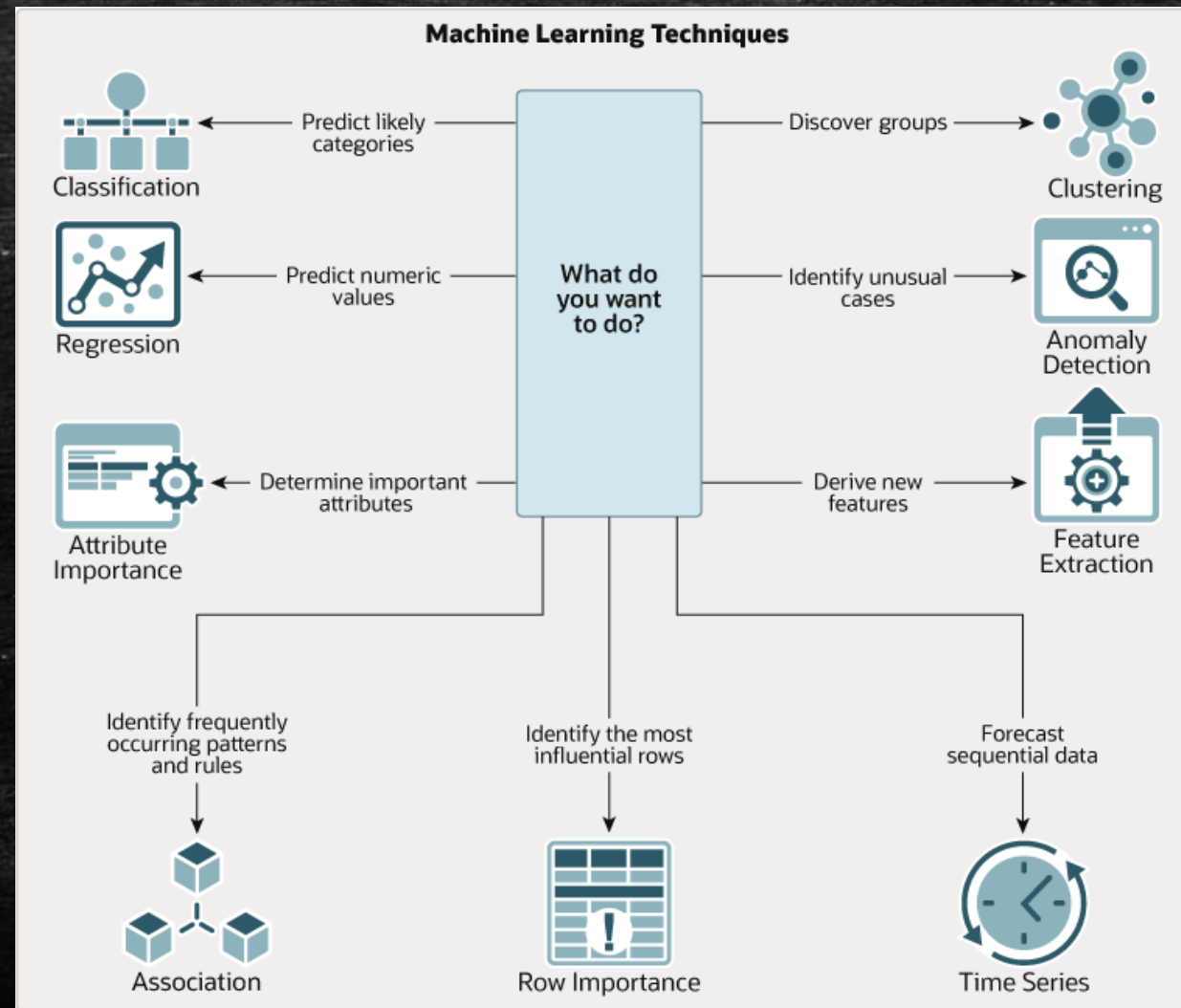
Deep Learning - Def

Deep learning is a specialized subfield of machine learning that uses artificial neural networks inspired by the structure and function of the human brain. Deep learning models, also known as deep neural networks, consist of multiple layers of interconnected nodes (neurons). These models can automatically learn hierarchical representations of data and perform complex tasks such as image and speech recognition. Deep learning has been particularly successful in areas like computer vision and natural language processing.

What Artificial Intelligence will do?

- Automation
- Autonomous System
- Natural Language Processing
- Image & Video Recognition
- Recommender System
- Predictive analysis
- Robotics
- Cybersecurity
- Creative Applications
- Environmental Monitoring
- Healthcare Diagnosis and Treatment
- Data Analysis and Insights and many more...

What Machine Learning will do?



What Deep Learning will do?

- Image Classification
- Object detection
- Natural Language Processing
- Speech Recognition
- Speech Synthesis
- Generative Models
- Recommender System
- Time Series analysis
- Medical Diagnosis
- Autonomous Vehicles
- Biometric and Facial Recogn
- Natural Language understanding
- Robotics
- Language generation
- Virtual Reality & Augmented Reality and many more...

What Data Science will do?

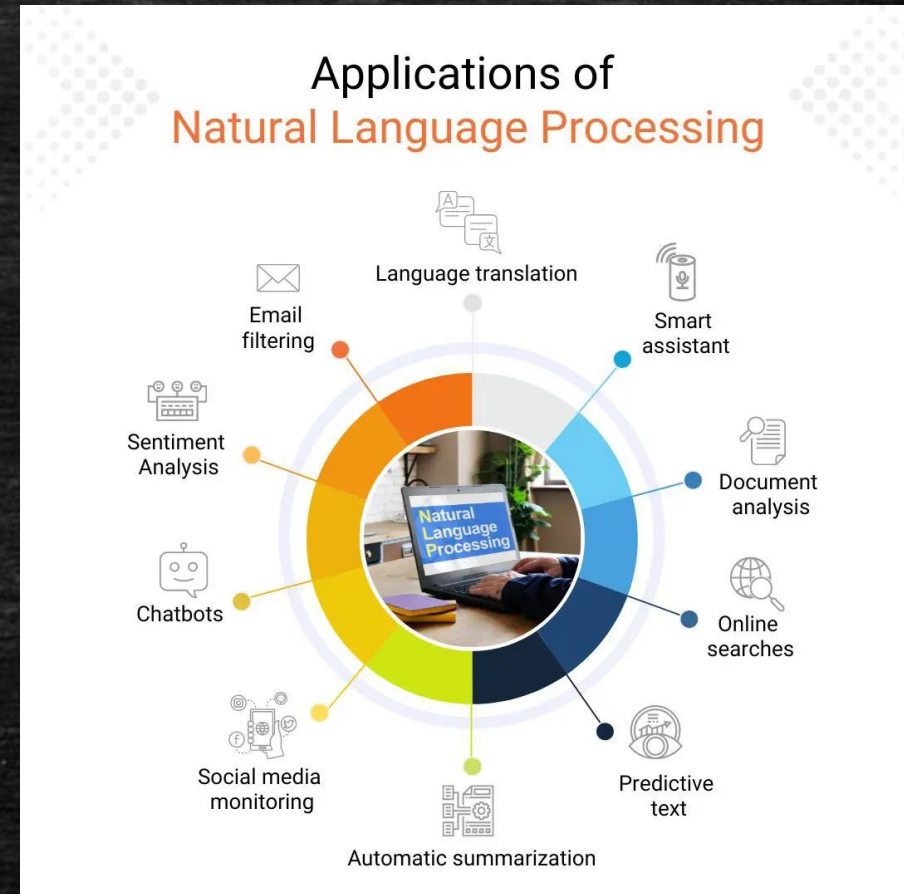
- Data Exploration and Visualization
- Classification and Regression
- Data Pre-processing and Cleaning
- Data-driven Decision Making (Stat + Domain Kno + Business)
- Data Visualization and Communication
- Data Governance and Privacy
- And many like AI, ML, DL...

Inference – DS Vs AI Vs ML Vs DL

- DS focuses on extracting insights from data
- AI aims to create intelligent systems
- ML enables computers to learn from data and make predictions
- DL is a subset of ML that uses neural networks to learn hierarchical representations.
- Above fields are interconnected and often work together to solve complex problems and drive advancements in technology and innovation

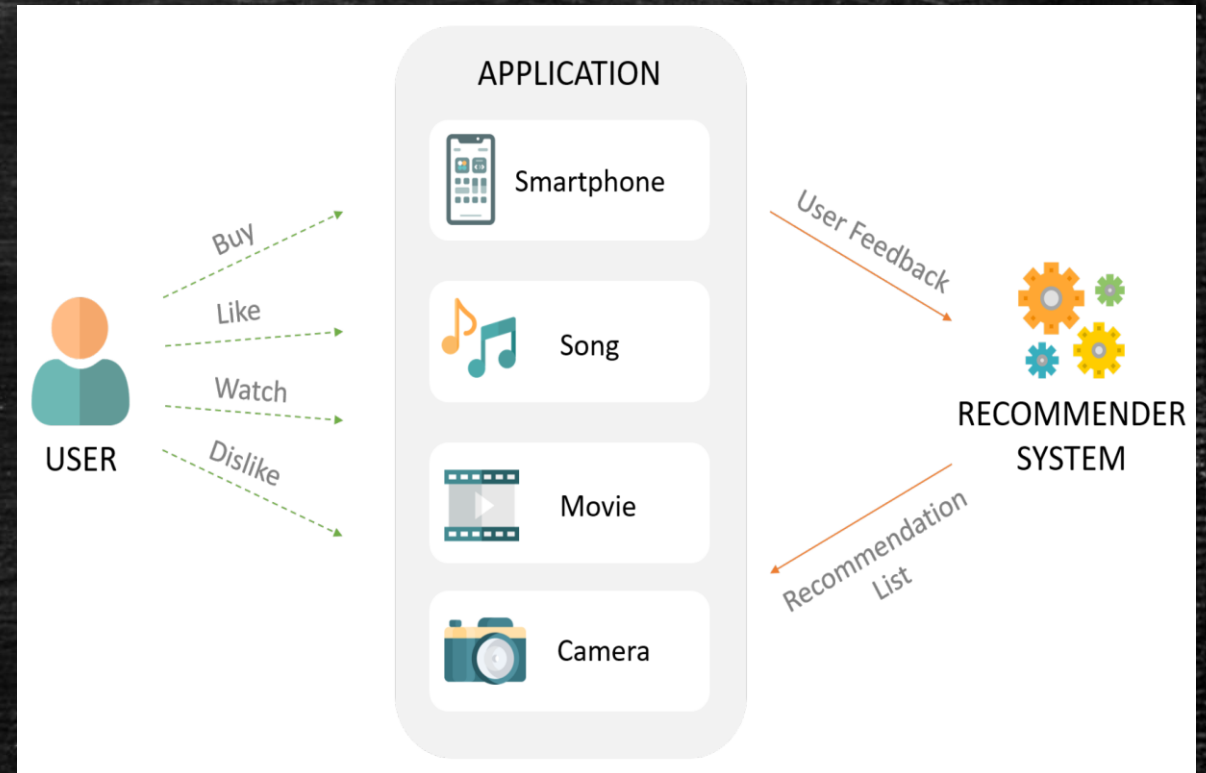
Then What about NLP?!

NLP is a branch of AI that focuses on enabling computers to understand, interpret, and generate human language. It involves the development of algorithms and models to process and analyze text data, extract meaning, and derive insights. NLP techniques include tasks such as sentiment analysis, text classification, named entity recognition, language translation, text summarization, and question answering.



Then What about RS?!

Recommender systems, also known as recommendation systems, are AI systems that provide personalized recommendations to users based on their preferences, behavior, and historical data. Recommender systems use algorithms to analyze user data and generate suggestions for items, products, movies, music, or content that are likely to be of interest to the user.



Data Science life cycle

- Problem definition
- Data Collection
- Data Cleaning & Pre-Processing
- Exploratory Data Analysis
- Feature Engineering
- Model Selection & Training
- Model Evaluation
- Model Deployment
- Communication & Visualization
- Iteration and Improvement

DATA



Signal – Useful information

Types – Numerical, Image,
Video, Audio, Text & many...

Different forms of DATA

- Structured data – Rows and Columns
- Unstructured data – text, emails, social media posts, images, audio, videos & sensor ...
- Semi-structured data – XML, JSON, log, HTML...
- Time series data – Forecasting, ICU
- Spatial data – Mapping, location based services
- Categorical data – Yes/No, Male/Female, types
- Numerical data – Qualitative – Continuous / Discrete
- Textual data – emails, social media, pdf, txt, doc & Graph data

Data Science Importance

- Extracting Insights from Data
- Data-Driven Decision-Making
- Improved Business Performance
- Personalization and Customer Experience
- Fraud Detection and Risk Management
- Forecasting and Predictive Analytics
- Process Optimization and Efficiency
- Innovation and New Product Development
- Healthcare and Research Advancements
- Social Impact and Public Policy

Data Science - Platform

- Python
- MATLAB
- R
- Julia
- Java
- Scala
- SQL and lot more...

Python @ BEST

- Easy to Learn and Readability
- Large and Active Community
- Rich Ecosystem of Libraries and Frameworks
- Versatility and Cross-Platform Compatibility
- Data Science and Machine Learning
- Automation and Scripting
- Integration and Extensibility
- Rapid Prototyping and Development
- Job Market and Career Opportunities
- Open Source and Free

Personalities to become Data Science

- Who can crack complex data oriented problem with their strong expertise in scientific disciplines
- Strong Business minded
- Mathematical reasoning
- Machine learning | Deep Learning
- Coding skills

Application of Data Science

- Business Analytics
- Healthcare
- Finance and Banking
- E-commerce and Retail
- Internet of Things (IoT)
- Social Media and Digital Marketing
- Energy and Utilities
- Transportation and Logistics
- Manufacturing and Quality Control

Skills needed for Data Scientist

- Statistics
- Python programming
- EDA
- ML/DL
- Problem solving skills
- Business communication
- Data Visualization
- SQL

Linux in Data Science

- Powerful Command-Line Tools
- Seamless Integration with Data Science Libraries
- Accessibility to Open-Source Software
- Computing Performance and Scalability
- Stability and Reliability
- Community and Documentation Support
- Compatibility with Cloud Computing and Containers

Version Control

- Version control, also known as source control or revision control
- It is a system that enables the tracking and management of changes made to a set of files or a project over time
- It is a critical component of software development and collaborative work environments.
- Version control systems (VCS) help teams and individuals keep track of changes, collaborate effectively, and maintain a history of project modifications

Key aspect of version control

- Tracking Changes
- Collaboration and Teamwork
- Reverting and Rollback
- Branching and Parallel Development
- Code Review and Collaboration
- Traceability and Auditing
- Backup and Disaster Recovery
- Reproducibility and Experimentation