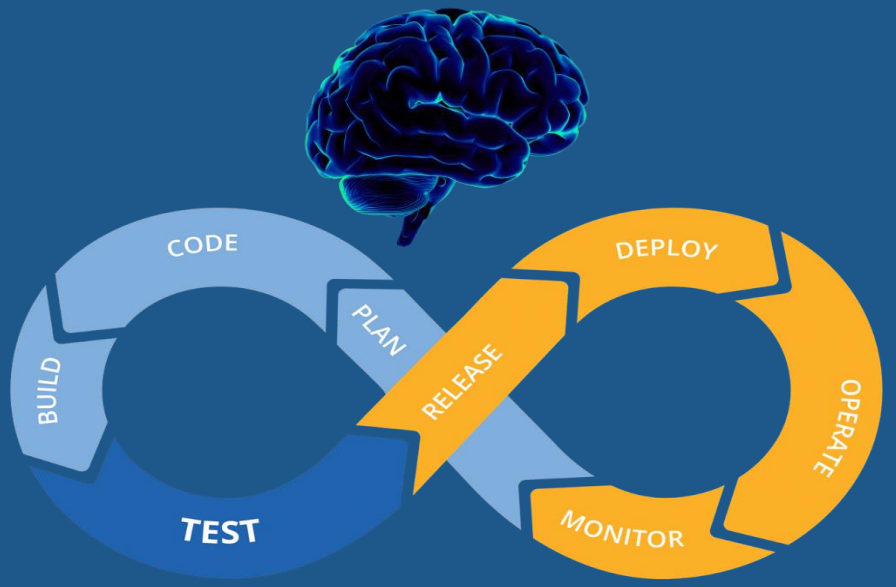
CSED!

"A Place to Invent and Learn"



Machine Learning Course

Course Outlines

Modules

1.Introduction

2.Python for Data Science

3.Data Preprocessing

4.Forecasting and Learning Theory

5.Validation

6.Performance measures

7.Project

8.Dimentionality Reduction

9.Supervised and Unsupervised Machine Learning

10.Ensamble Learning

11.Machine Learning Approach in NLP

12.Introduction to CNN

Data Science Skills Level - I

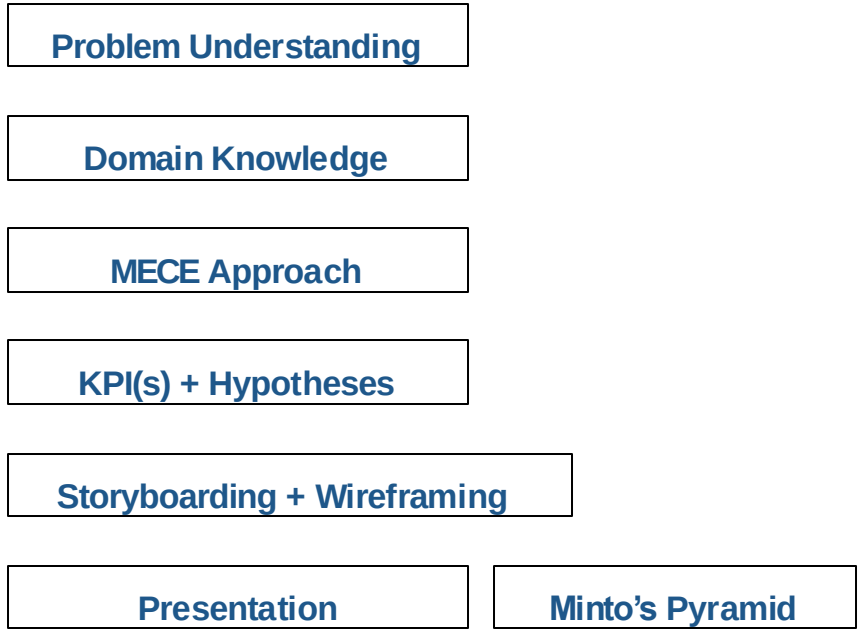
Problem Analysis

- Understanding the problem
- Identifying the industry domain
- Learning the MECE Approach
- Developing Hypotheses
- Developing KPIs
- Storyboarding the solution
- Developing wireframes to visualize the solution
- Build story flows on ppt/google slides/jupyter notebooks
- Learning the right communication approach - Minto's Pyramid

KPI Building

Storyboarding

Hypothesis Building



Data Science Skills Level - II

Ability to query data from a variety of sources	Ability to pre-process data	Ability to build simple to complex reports and visualizations
• Relational databases : MySQL, MS-SQL Server, Postgresql • NoSQL databases : MongoDB, Cassandra, DynamoDB etc. • Hadoop Distributed File System • Web API • Excel, Text, CSV, JSON, XML etc.	• Knowledge of simple to complex SQL queries • Brief idea of SQL functions and procedures • Extensive knowledge of data pre-processing in pandas • Missing value treatment • Outlier Treatment	• Knowledge of at least one BI & Visualization tools like Tableau/Power BI • Knowledge of data warehouse concepts ex. Facts & Dimensions • Visualizing data using Python (Matplotlib, Seaborn, Plotly etc.)

Data Science Skills Level - III

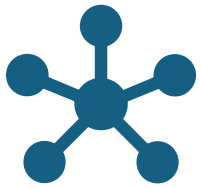
Ability to perform Statistical analysis on data

- Ability to summarize data using statistical properties
- Ability to perform exhaustive univariate analysis.
- Idea of statistical tests
- Ability to perform bivariate/multivariate analysis
- Understanding of hypothesis testing
- Good grasp over the idea of normal distribution
- Excellent skills at Exploratory Data Analysis

Ability to explain inner workings of Machine Learning models

- Knowledge of types of ML problems (Supervised, Unsupervised)
- Idea of the mathematics of ML models
- Knowing how to handle different types of data
- Knowing the steps to implement before the data is ready to be fed into ML models
- Ability to explain the output of models
- Ability to explain assumptions which are required for a model
- Ability to tune the model to arrive at the best results
- Ability to explain why a model should work in a given situation?

Data Science Skills Level - IV



Neural Networks

Brief idea of Multi-layer perceptrons

Brief idea of inner workings of a neural network. Understanding of backpropagation, activation functions, optimizers etc.

An understanding of when neural networks is NOT to be used

Computer Vision

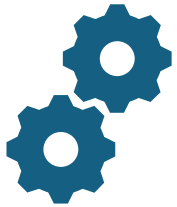
Knowledge of Image processing techniques

Brief idea of Convolutional neural networks

An idea of face detection and recognition techniques

Decent experience in implementation of CNNs both by scratch as well as by transfer learning

Good knowledge of computer vision libraries like PIL, OpenCV, dlib etc.



Natural Language Processing

Knowledge of text processing techniques

Understanding traditional NLP algorithms

Learning NLP the deep learning way : CBOW, Skip-grams

Learning about common Python packages for NLP : nltk, gensim, fasttext etc.

Data Science Skills Level - V

Working with Cloud

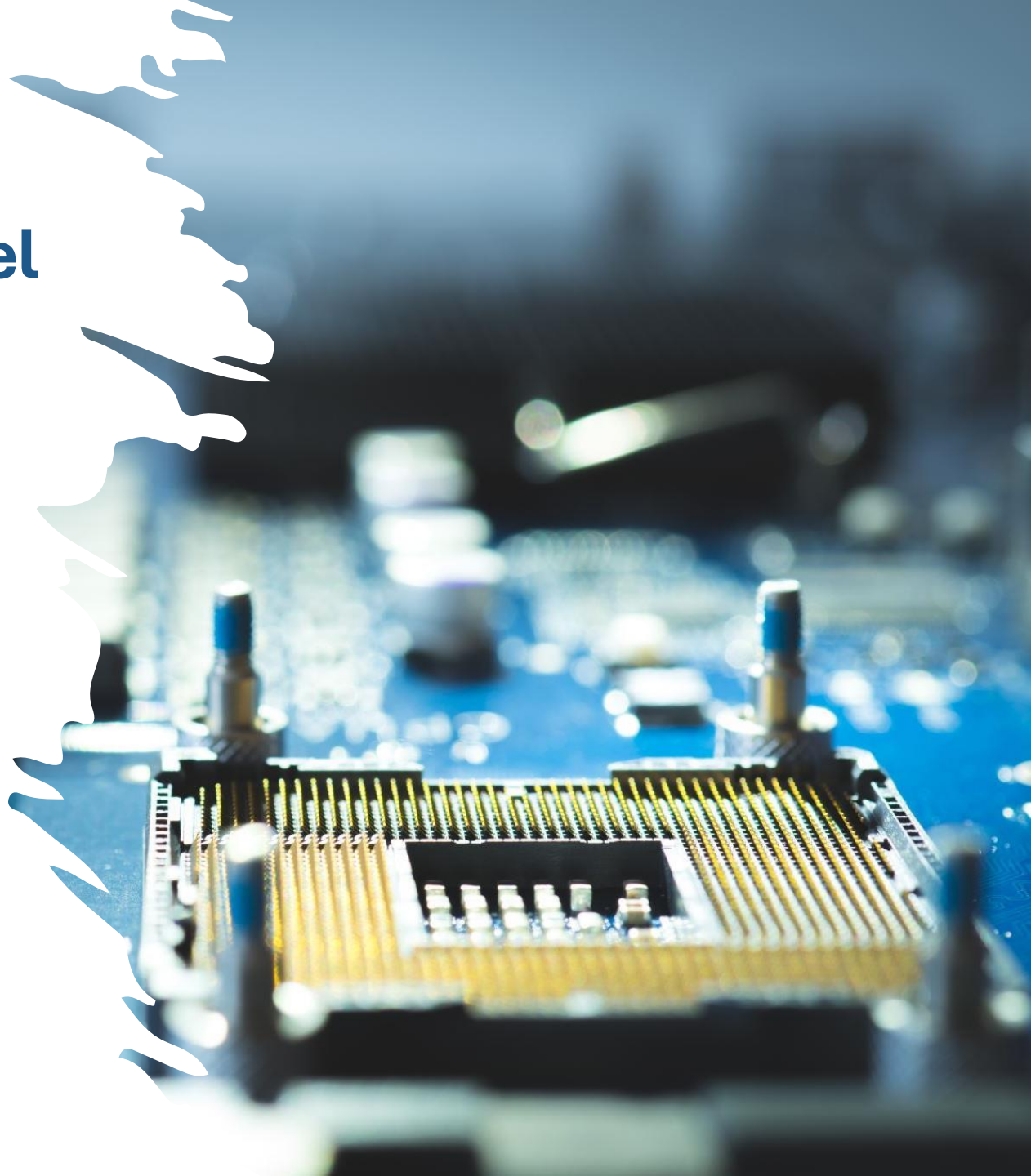
- Ability to perform all the tasks mentioned earlier on a Cloud platform (Azure, AWS, GCP)
- Ability to design a data science pipeline using Cloud components
- Deploy an end-to-end data science project in Production on Cloud

Technologies Required

- SQL for data mining
- Python for everything!
- Spark for data mining at Scale!
- Power BI for visualization

Positions can be offered according to skill level

- Level-1 : Strategy Consultant / Business Analyst
- Level-2 : Data Analyst / Data Engineer
- Level-3 : Data Scientist / Machine Learning Engineer
- Level-4 : Machine Learning Engineer / Computer Vision and NLP Specialist
- Level-5 : Cloud Data Scientist / Cloud Data Engineer / Solution Architect



Module 1: Introduction to Machine Learning

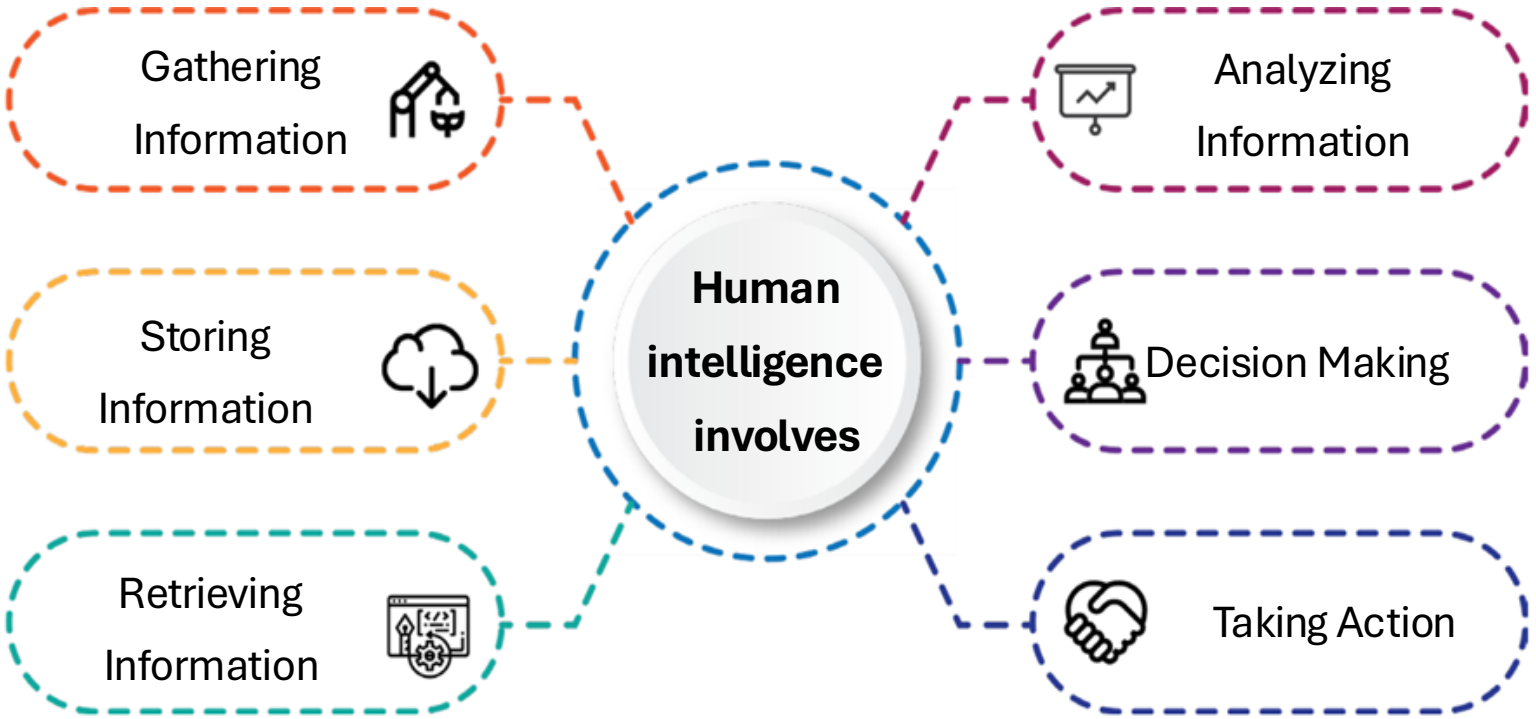
Topics

- What is Machine Learning
- Types of Machine Learning
- Components in Machine Learning
- Application of machine learning
- Machine learning Pipeline
- Machine learning solution Architecture design
- Tools for machine learning
- Machine learning algorithms
- What is Deep Learning
- Deep learning applications

What is AI

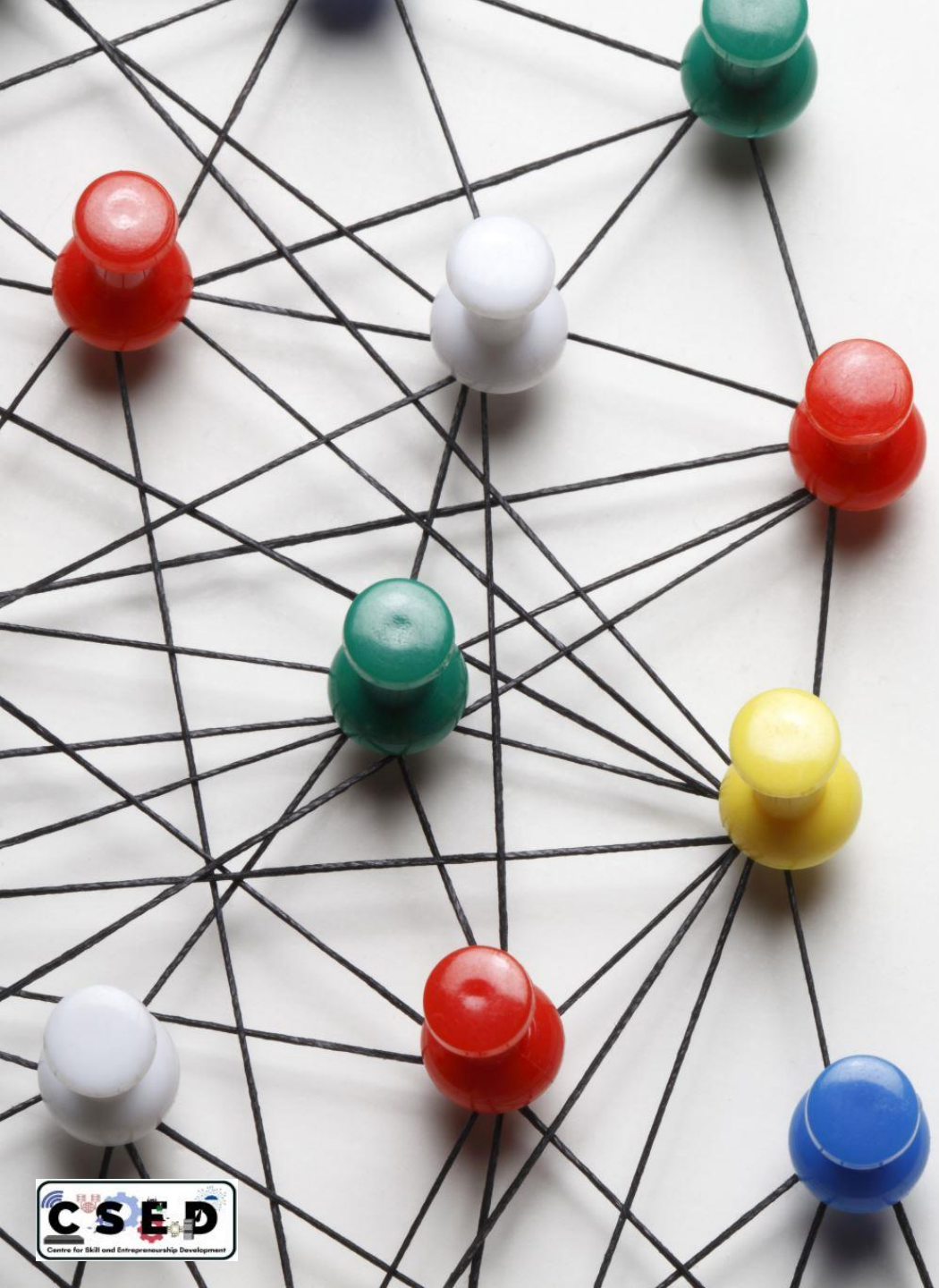
Artificial Intelligence (AI) is a field of computer science that focuses on creating systems capable of performing tasks that typically require human intelligence. These tasks include learning, reasoning, problem-solving, understanding natural language, and perception. AI systems are designed to mimic cognitive functions such as recognizing patterns, making decisions, and adapting to new information.

Human Intelligence



NEED OF AI

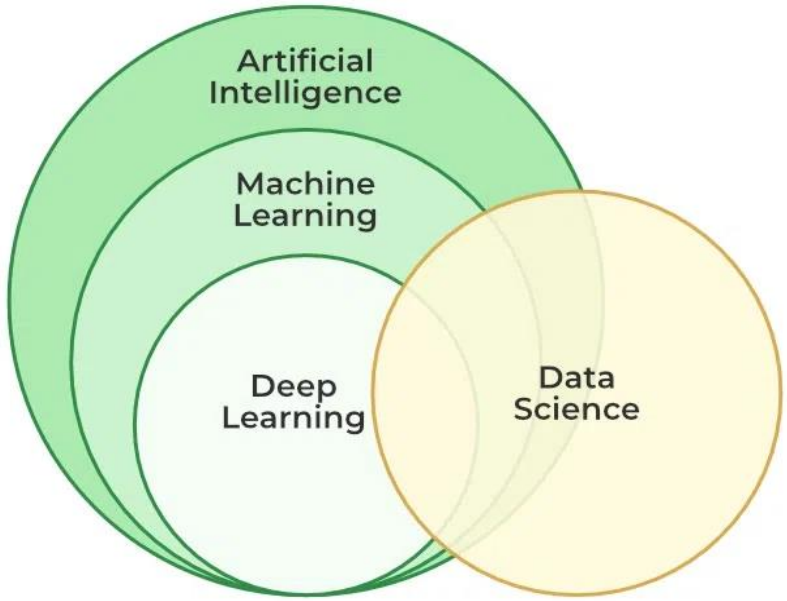
- To do tasks that humans want to avoid because of the risks involved to do things faster
- To do things that require more power
- To be more accurate
- To overcome human inefficiency
- To achieve consistency
- To have machines as companions
- To understand how humans' function and have evolved



Necessary abilities for AI

- To understand and respond to spoken or written commands
- To recognize you and your friends
- To see and differentiate between objects, animals, and people
- To go from one place to another by themselves
- To play games
- To do creative work

What is Machine Learning

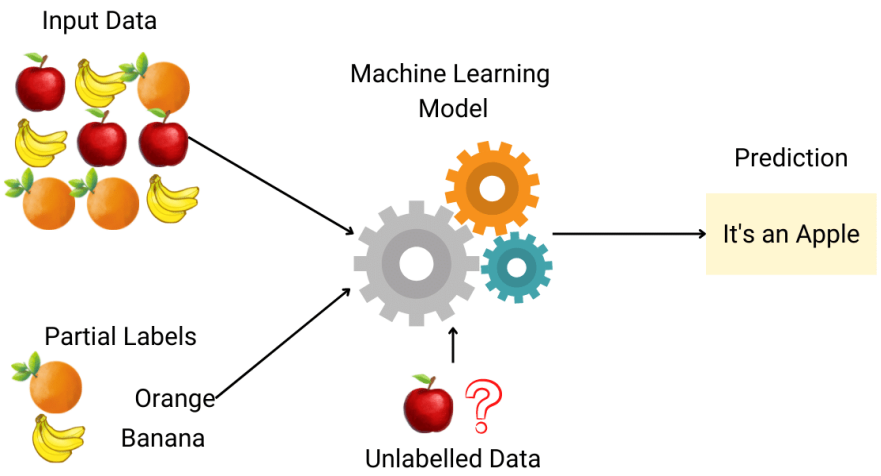


Machine learning is a subset of artificial intelligence (AI) that focuses on building systems that can learn from and make decisions based on data. Unlike traditional programming, where explicit instructions are provided to achieve a specific task, machine learning involves creating algorithms that allow computers to identify patterns, make predictions, and improve their performance over time based on experience.

Types of Machine Learning

- Supervised
- Semi-Supervised
- Unsupervised
- Reinforcement

Supervised Machine Learning



In supervised learning, the algorithm is trained on **labeled data**, meaning the input data comes with corresponding output labels. The goal is to learn a mapping from **inputs to outputs**.

Examples:

- **Classification:** Predicting a discrete label. (e.g., spam detection, image recognition)
- **Regression:** Predicting a continuous value. (e.g., edicting house prices, stock market trends)

Supervised Machine Learning Algorithm

- Logistic Regression
- k-Nearest Neighbors (k-NN)
- Support Vector Machines (SVM)
- Decision Trees
- Random Forest
- Gradient Boosting Machines (GBM)
- XGBoost
- Naive Bayes
- Artificial Neural Networks (ANN)
- Convolutional Neural Networks (CNN)
- Recurrent Neural Networks (RNN)
- Multilayer Perceptron (MLP)
- Linear Regression

Unsupervised Machine Learning

Unsupervised machine learning is a branch of machine learning where the algorithm is trained on data that does not have labeled responses. In simpler terms, the machine is given a bunch of data and is asked to find patterns, structures, or relationships in that data without being told what to look for.

Clustering: Grouping similar data points together. Imagine you have a collection of mixed fruit images, and you want to group them into categories like apples, bananas, and oranges without knowing the categories beforehand.

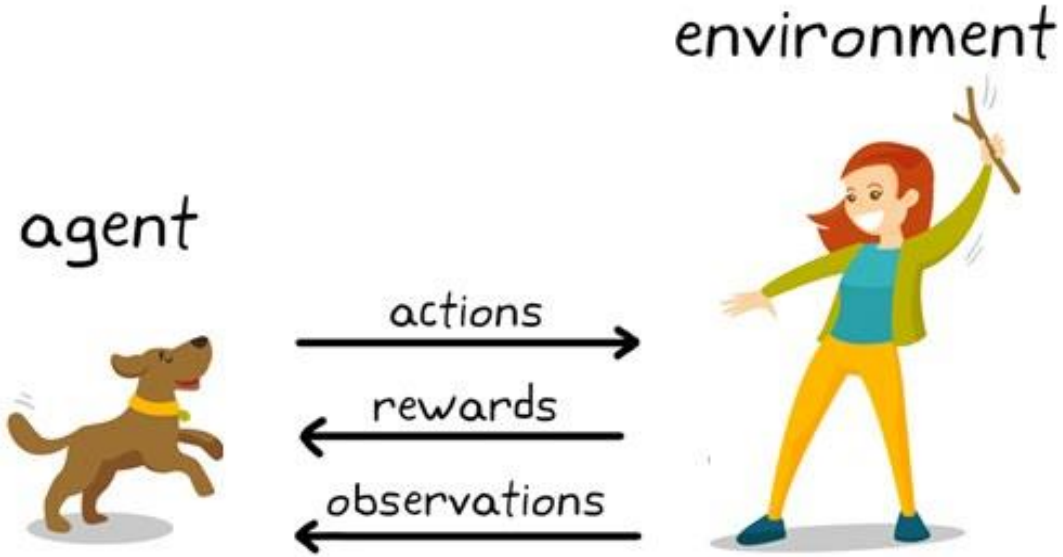
Dimensionality Reduction: Simplifying the data while keeping its essential features. Think of it like summarizing a long article into a few key points.

Unsupervised machine learning Algorithms

- K-Means
- Hierarchical Clustering
- DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
- Mean Shift
- Gaussian Mixture Models (GMM)
- Agglomerative Clustering
- Spectral Clustering
- Principal Component Analysis (PCA)
- Linear Discriminant Analysis (LDA)
- Independent Component Analysis (ICA)
- Self-Organizing Maps (SOM)
- Restricted Boltzmann Machines (RBM)
- Generative Adversarial Networks (GANs)

Reinforcement Machine Learning

Reinforcement learning (RL) is a type of machine learning where an agent learns to make decisions by taking actions in an environment to maximize some notion of cumulative reward. Unlike supervised learning, where the model learns from a labeled dataset, reinforcement learning focuses on learning from the consequences of actions, typically through trial and error.



Reinforcement machine algorithms

- Q-Learning
- Deep Q-Network (DQN)
- Double Q-Learning
- Dueling DQN
- Prioritized Experience Replay

- REINFORCE
- Deep Deterministic Policy Gradient (DDPG)
- Trust Region Policy Optimization (TRPO)
- Proximal Policy Optimization (PPO)
- Soft Actor-Critic (SAC)

Key Concepts in Reinforcement Learning

AGENT: THE LEARNER OR DECISION-MAKER THAT INTERACTS WITH THE ENVIRONMENT.

ENVIRONMENT: EVERYTHING THE AGENT INTERACTS WITH; IT PROVIDES FEEDBACK IN THE FORM OF REWARDS OR PENALTIES BASED ON THE ACTIONS TAKEN BY THE AGENT.

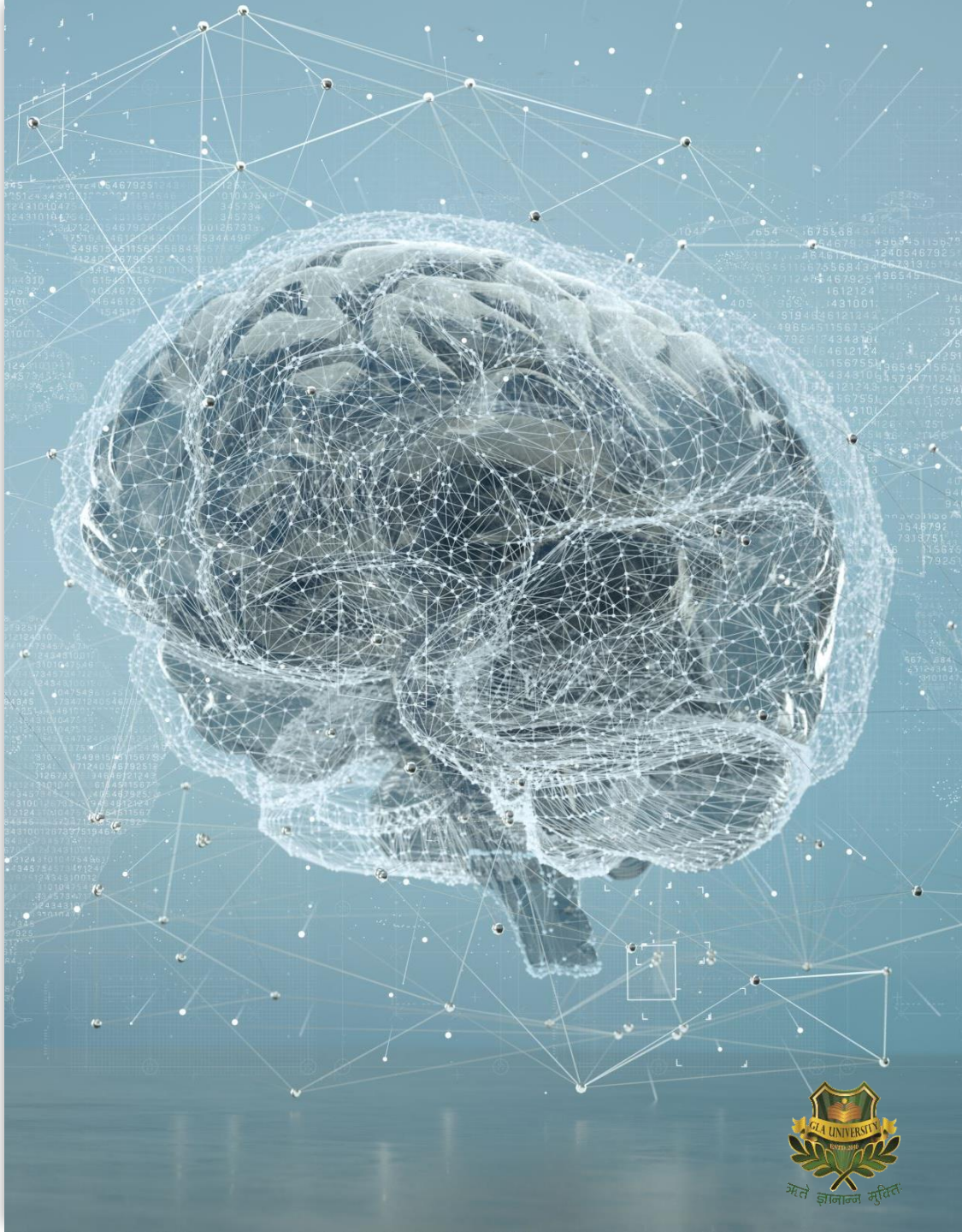
STATE (S): A REPRESENTATION OF THE CURRENT SITUATION OR CONFIGURATION OF THE ENVIRONMENT.

ACTION (A): ANY MOVE THE AGENT CAN MAKE IN THE ENVIRONMENT.

REWARD (R): FEEDBACK FROM THE ENVIRONMENT BASED ON THE ACTION TAKEN; IT CAN BE POSITIVE OR

What is deep learning

Deep learning is a subset of machine learning that uses multilayered neural networks, called deep neural networks, to simulate the complex decision-making power of the human brain. Some form of deep learning powers most of the artificial intelligence (AI) applications in our lives today.



Deep Learning Architecture

Feedforward Neural Networks (FNN)

Convolutional Neural Networks (CNNs)

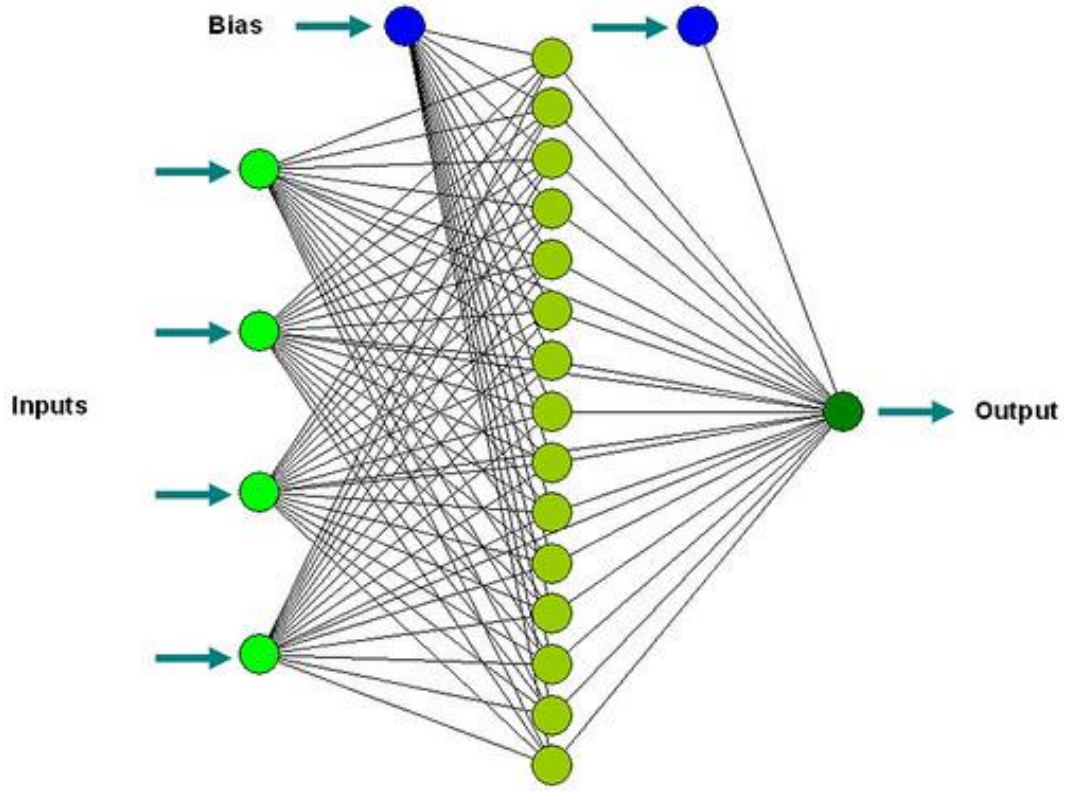
Recurrent Neural Networks (RNNs)

Long Short-Term Memory Networks (LSTMs)

Gated Recurrent Units (GRUs)

Autoencoders

Generative Adversarial Networks (GANs)



Pre-Trained Deep learning Architecture

Natural Language Processing

- BERT (Bidirectional Encoder Representations from Transformers)
- GPT (Generative Pretrained Transformer
- XLNet
- DistilBERT
- ALBERT (A Lite BERT)
- T5 (Text-To-Text Transfer Transformer)
- DeepSpeech
- Wav2Vec

Image classification

VGG (Visual Geometry Group)

ResNet (Residual Networks)

Inception (GoogLeNet)

MobileNet

DenseNet

EfficientNet

NASNet (Neural Architecture Search Network)

Object Detection and Localization

- Faster R-CNN
- YOLO (You Only Look Once)
- SSD (Single Shot MultiBox Detector)

Components of machine learning

1.Data

2. Models

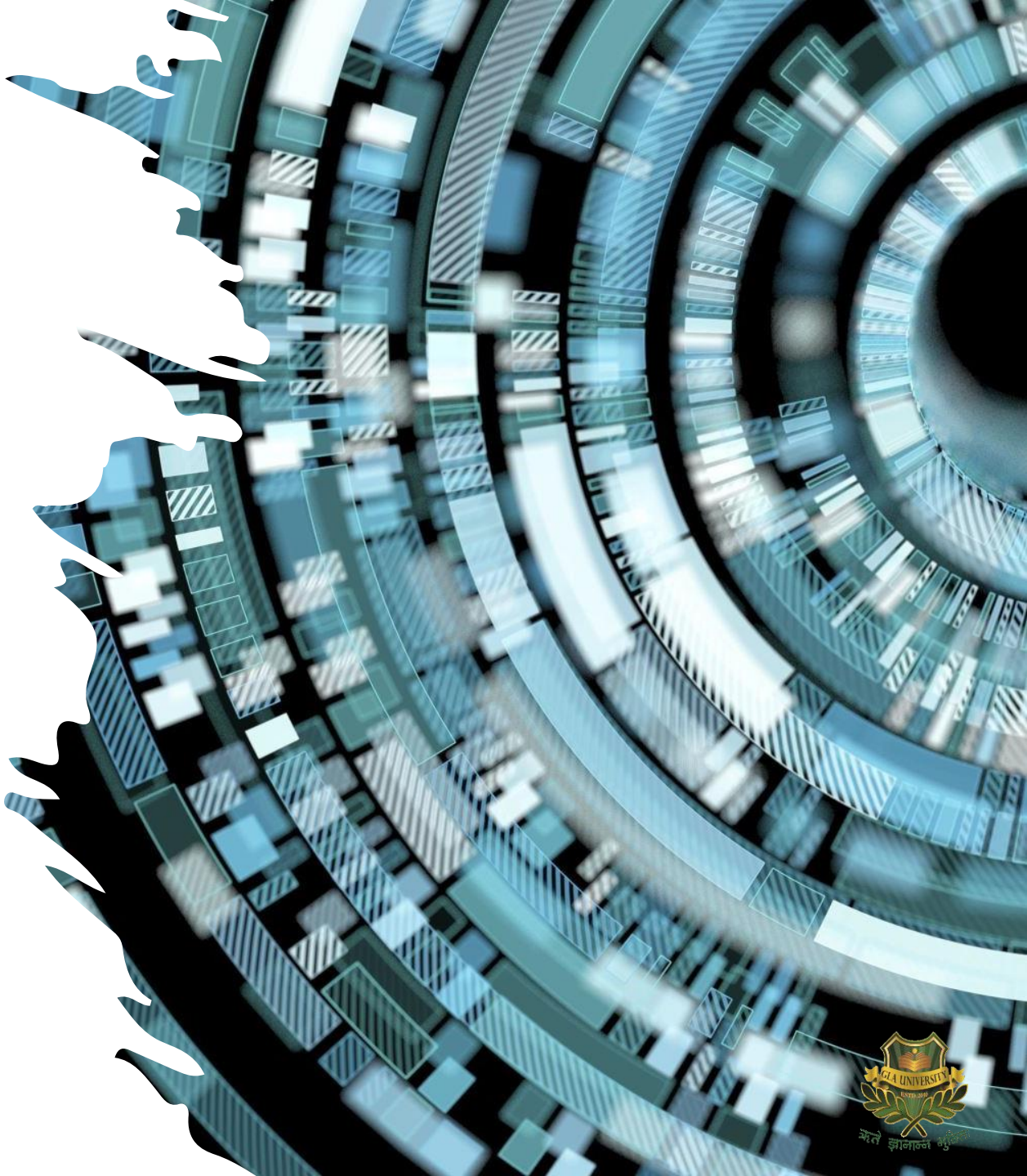
3.Algorithms

4. Error Functions

5. Evalutation Metrics

Application of machine learning

Machine learning finds applications across numerous fields and industries, transforming how businesses operate, automate tasks, and make decisions. Here are some key applications of machine learning



Healthcare

- **Disease Diagnosis:** ML models can analyze medical data and images to diagnose diseases such as cancer, diabetes, and heart disease.
- **Personalized Medicine:** Tailoring treatment plans based on individual patient data and genetic information.
- **Predictive Analytics:** Predicting patient outcomes, hospital readmissions, and potential health risks.

Finance

- **Fraud Detection:** Identifying fraudulent transactions by analyzing patterns and anomalies in financial data.
- **Algorithmic Trading:** Developing trading strategies that automatically execute trades based on market data and trends.
- **Credit Scoring:** Assessing the creditworthiness of individuals or businesses by analyzing financial histories and behaviors.

Retail

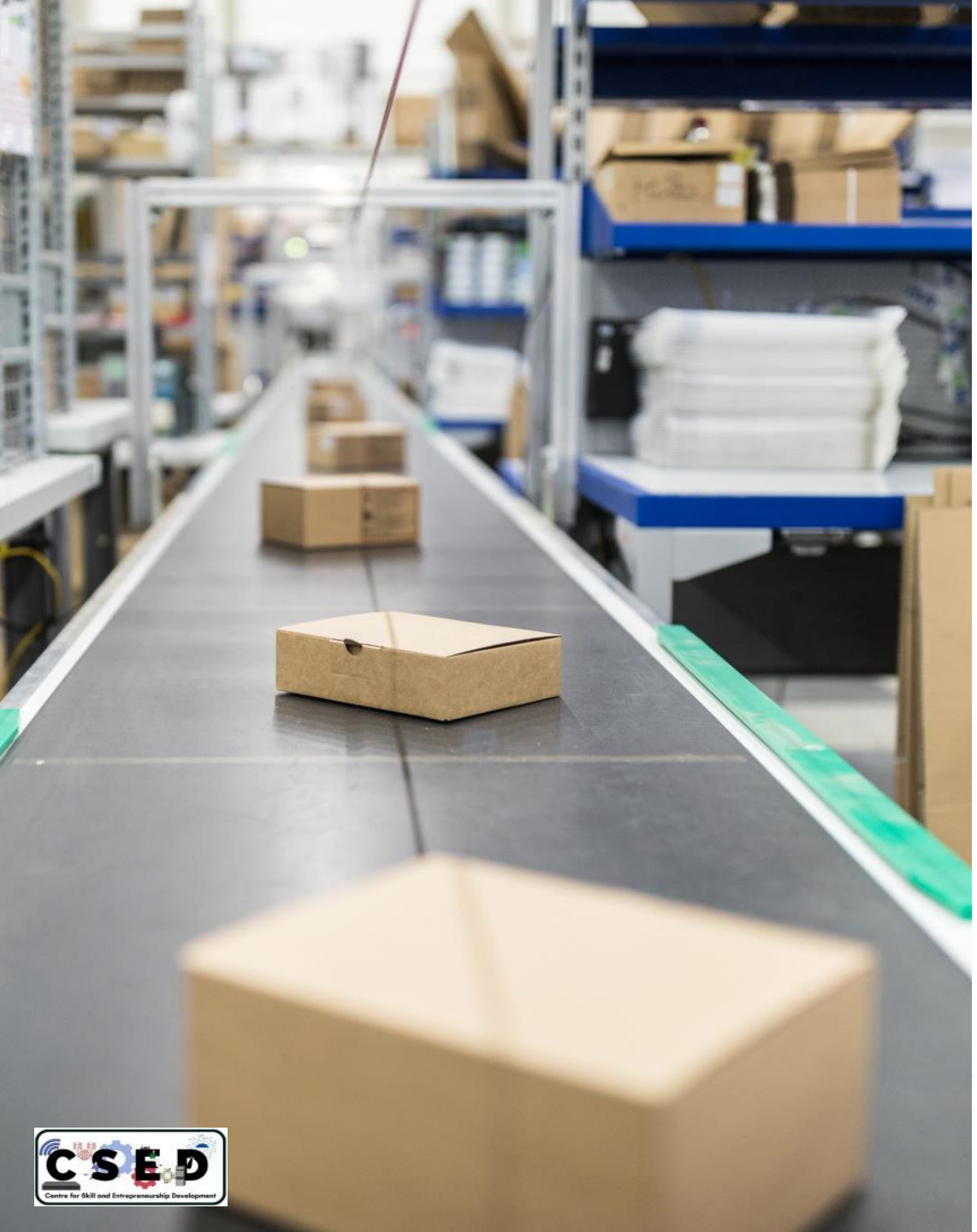
- **Recommendation Systems:** Suggesting products to customers based on their browsing and purchase history (e.g., Amazon, Netflix).
- **Inventory Management:** Predicting demand to optimize inventory levels and reduce stockouts or overstock situations.
- **Customer Segmentation:** Grouping customers into segments based on behavior and preferences for targeted marketing.

Transportation

- **Autonomous Vehicles:** Enabling self-driving cars to navigate and make decisions in real-time using sensor data and machine learning algorithms.
- **Route Optimization:** Finding the most efficient routes for delivery services and ride-sharing apps.
- **Predictive Maintenance:** Predicting when vehicle components will fail to schedule maintenance before breakdowns occur.

Marketing

- **Customer Insights:** Analyzing customer data to understand preferences and improve targeting of marketing campaigns.
- **Sentiment Analysis:** Monitoring social media and online reviews to gauge public opinion about products or brands.
- **Churn Prediction:** Identifying customers at risk of leaving and taking proactive measures to retain them.

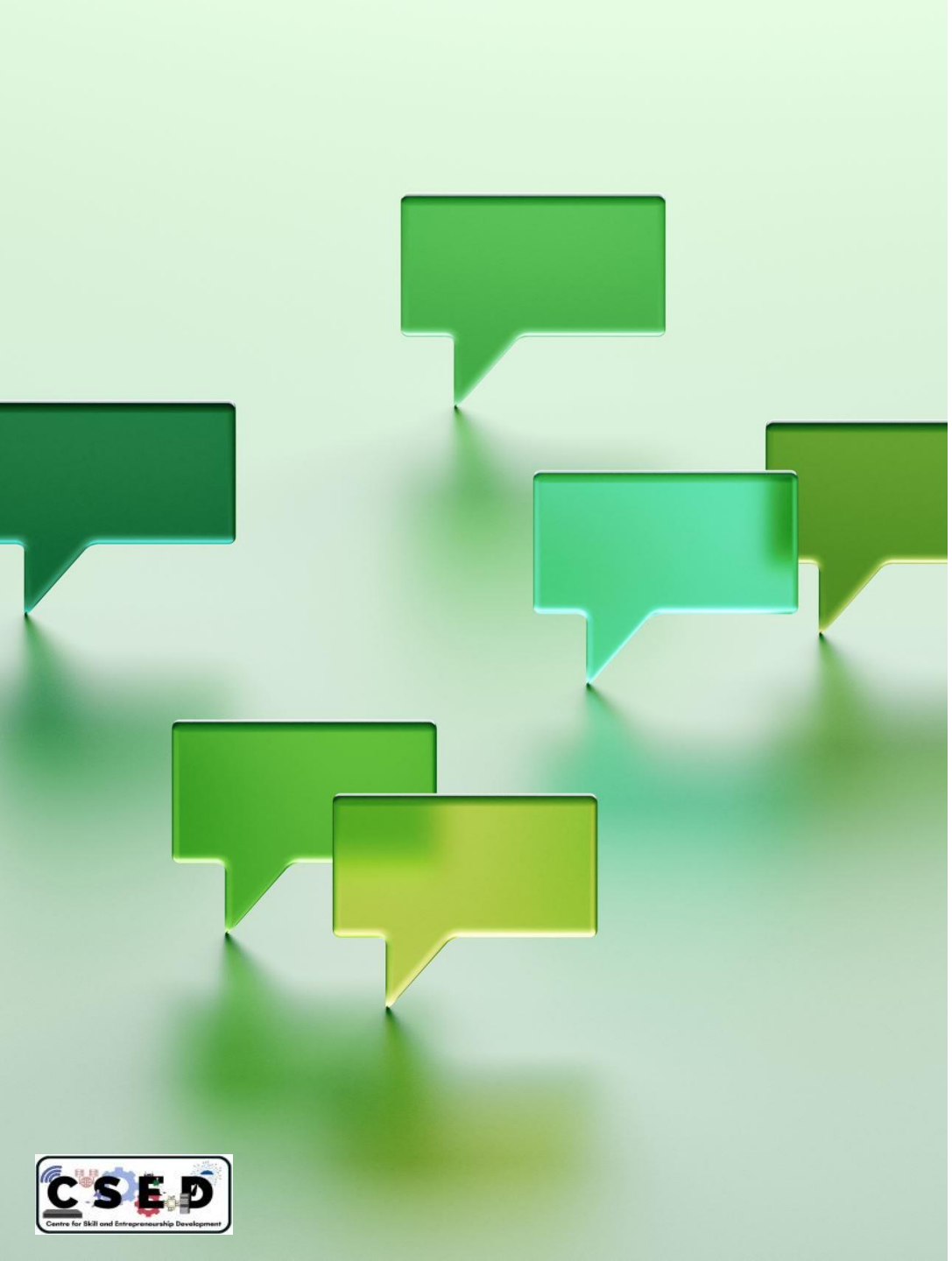


Manufacturing

- **Quality Control:** Using ML to detect defects in products by analyzing images or sensor data from production lines.
- **Predictive Maintenance:** Predicting equipment failures to perform maintenance proactively, reducing downtime and costs.
- **Supply Chain Optimization:** Optimizing the entire supply chain process from raw material procurement to delivery of finished goods.

Agriculture

- **Crop Monitoring:** Analyzing satellite images and sensor data to monitor crop health, predict yields, and detect pests or diseases.
- **Precision Farming:** Optimizing the use of resources like water, fertilizers, and pesticides based on data-driven insights.
- **Automated Machinery:** Developing autonomous farming equipment for planting, harvesting, and soil analysis.



Natural Language Processing (NLP)

- **Language Translation:** Translating text or speech from one language to another (e.g., Google Translate).
- **Chatbots:** Providing customer support and information through automated chat systems.
- **Text Analysis:** Extracting meaningful information from unstructured text data for insights and decision-making.

Image and Video Analysis

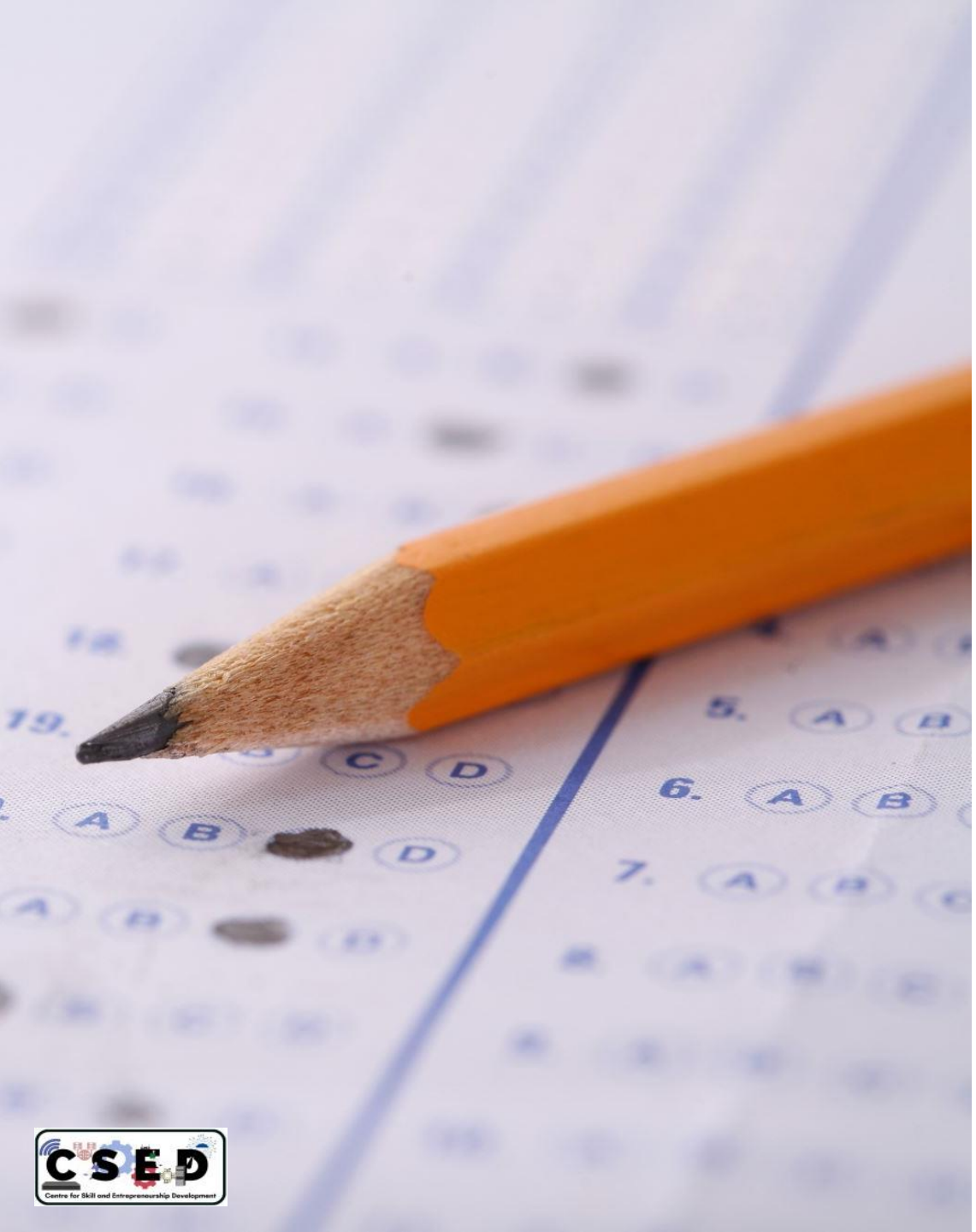
- **Facial Recognition:** Identifying individuals in images or videos for security and authentication purposes.
- **Object Detection:** Detecting and classifying objects within images or videos for applications like surveillance and autonomous vehicles.
- **Medical Imaging:** Analyzing medical images to detect and diagnose conditions such as tumors and fractures.

Energy

- **Smart Grids:** Optimizing the distribution of electricity by predicting demand and managing supply.
- **Energy Consumption:** Analyzing usage patterns to reduce energy consumption and increase efficiency.
- **Renewable Energy:** Predicting the availability of renewable energy sources like solar and wind.

Entertainment

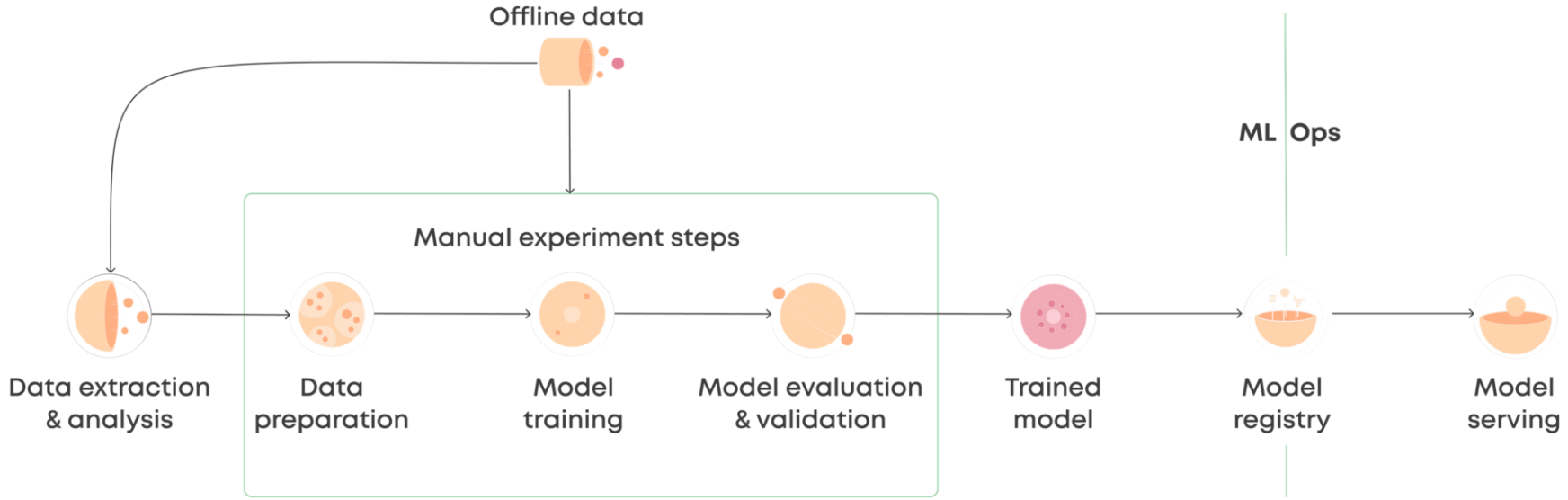
- **Content Recommendation:** Suggesting music, movies, and other content based on user preferences and behavior.
- **Content Creation:** Using ML to generate music, art, and even news articles.
- **Audience Analysis:** Understanding audience preferences and behaviors to tailor content and marketing strategies.



Education

- **Personalized Learning:** Adapting educational content and pacing based on individual student performance and learning styles.
- **Grading and Assessment:** Automating the grading of assignments and exams using ML algorithms.
- **Student Retention:** Predicting which students are at risk of dropping out and intervening with support.

Machine Learning Pipeline



Machine Learning Solution Architecture design

- Define the Problem and Objectives
- Data Collection and Preparation
- Exploratory Data Analysis (EDA)
- Feature Engineering
- Model Selection and Training
- Model Evaluation
- Deployment
- Maintenance and Continuous Improvement

Tools and Technologies

DATA PROCESSING: APACHE SPARK, PANDAS, HADOOP.

MODEL TRAINING: TENSORFLOW, PYTORCH, SCIKIT-LEARN.

DEPLOYMENT: DOCKER, KUBERNETES, AWS SAGEMAKER, GOOGLE AI PLATFORM, AZURE ML.

MONITORING: PROMETHEUS, GRAFANA, ELK STACK.

Thanks