Ethnotech Academic Solutions Pvt. Ltd.

IT SPECIALIST
ARTIFICIAL INTELLIGENCE

Pearson

ARTIFICIAL INTELLIGENCE

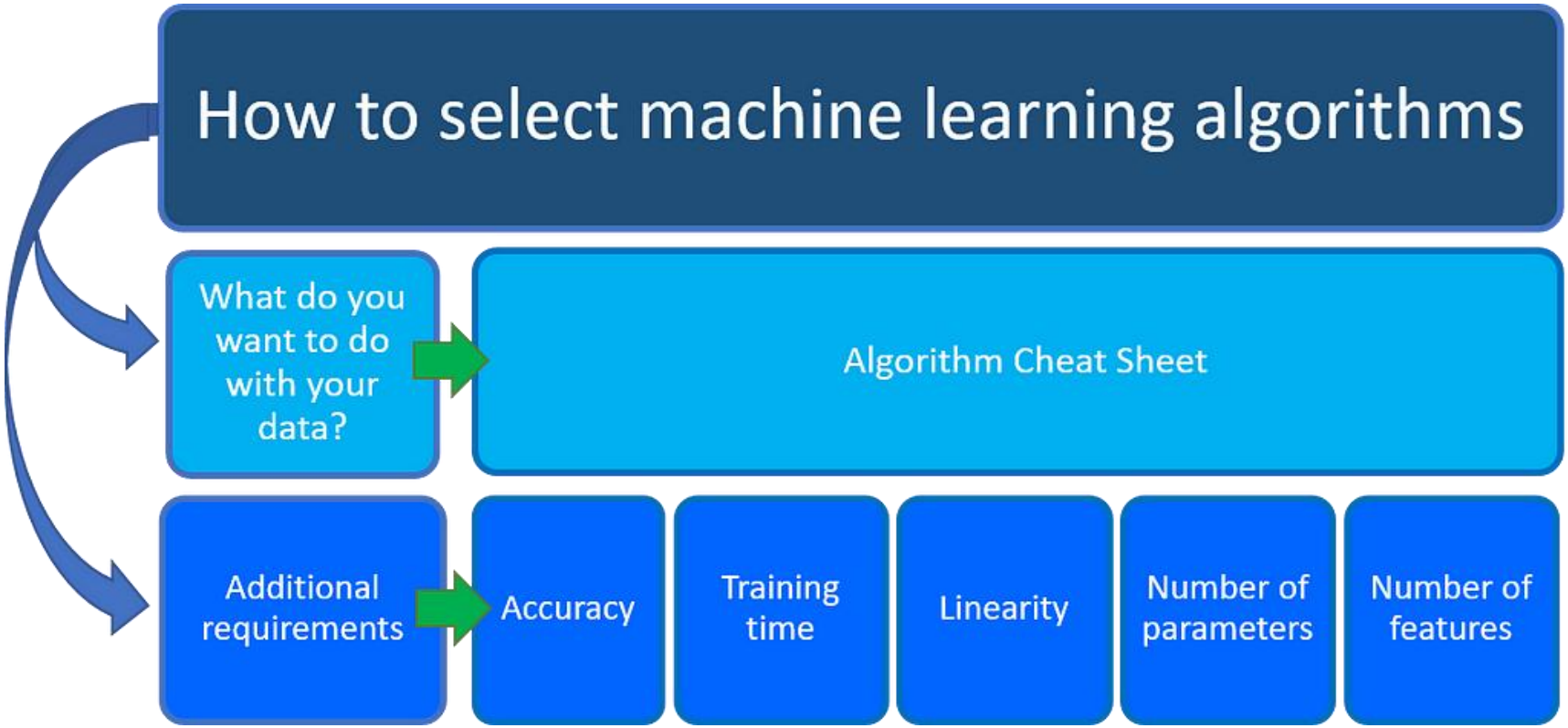
Session Content

1	Building an AI model that solves the problem
2	Consider applicability of specific algorithms
3	Train a model using the selected algorithm
4	Select specific model after experimentation, avoiding Overengineering
5	Evaluate model performance
6	Look for potential sources of bias in the algorithm
7	Evaluate model sensitivity
8	Confirm adherence to regulatory requirements
9	Obtain stakeholder approval

Types of algorithms in AI

AI algorithms play a crucial role in various AI tasks, ranging from machine learning and deep learning to natural language processing and computer vision. Here are some popular algorithms and their applications:

- **Linear Regression**
- **Decision Trees**
- **Random Forest**
- **Support Vector Machines (SVM)**
- **Neural Networks**
- **Convolutional Neural Networks (CNN)**
- **Recurrent Neural Networks (RNN)**
- **Long Short-Term Memory (LSTM)**
- **Gaussian Mixture Models (GMM)**
- **K-Means Clustering**
- **Reinforcement Learning Algorithms**


How to select machine learning algorithms
What do you want to do with your data?
Algorithm Cheat Sheet
Additional requirements
Accuracy
Training time
Linearity
Number of parameters
Number of features

Applicability of algorithms in AI

When selecting algorithms for various tasks, it's important to consider factors such as the problem domain, data characteristics, computational efficiency, and the specific goals you aim to achieve.

Here are some common use cases along with algorithm considerations:

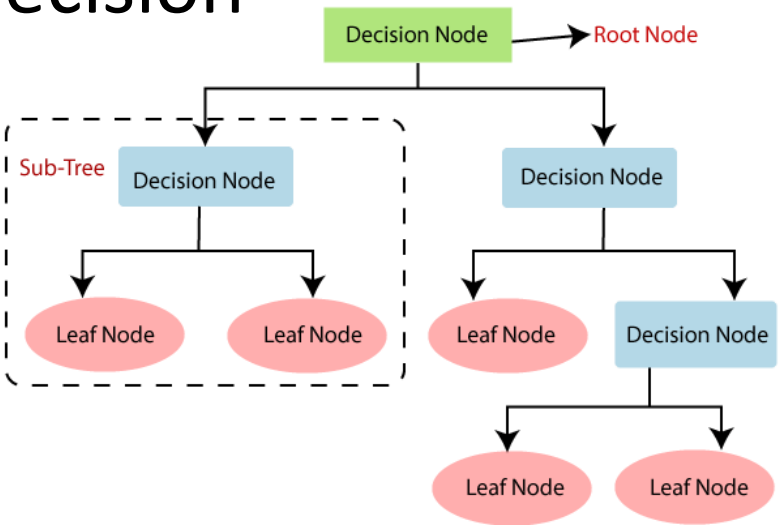
Classification:

Problem: Assigning data points to predefined classes or categories.

Algorithms:
Decision Trees
Random Forest
Support Vector Machines (SVM)
Neural Networks

Classification: Decision Trees **sample program**

- Calculating percentage of 3 scores in three subjects
- Checking the condition and deciding on the category (FCD/FC/SC/TC/Fail)
- Display the graph of the decision



Applicability of algorithms in AI

Clustering:

Problem: Grouping similar data points into clusters based on similarity.

Algorithms:
K-Means
Hierarchical Clustering
DBSCAN

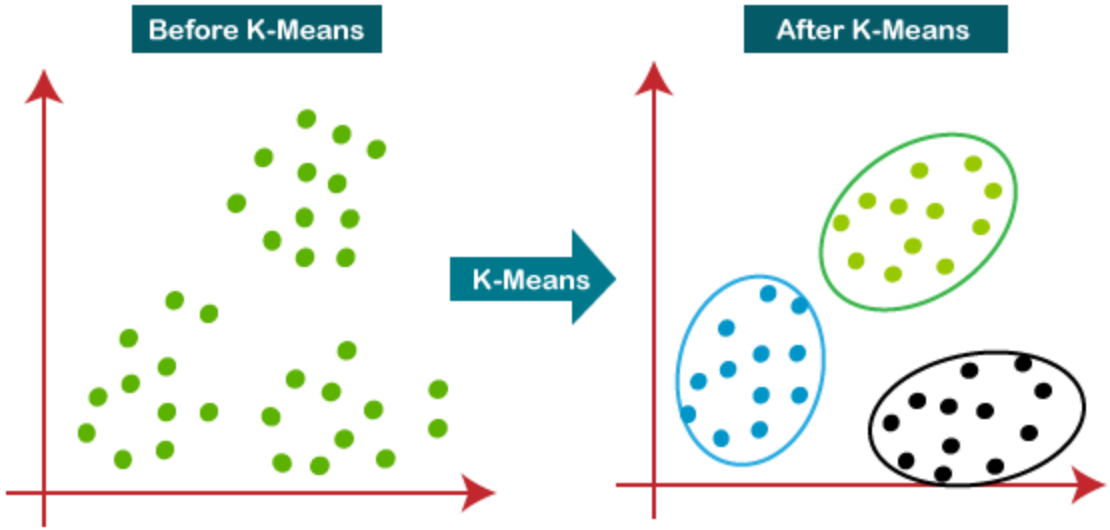
What is K mean?

K-means is a popular unsupervised machine learning algorithm used for clustering data points into groups based on similarity.

The goal of the K-means algorithm is to partition a given dataset into K clusters, where each data point belongs to the cluster with the nearest mean (centroid).

how the K-means algorithm works:

- Initialization
- Assignment
- Update Centroids
- Iteration



Applicability of algorithms in AI

Regression:

Problem: Predicting a continuous numerical output based on input features.

Algorithms:
Linear Regression
Ridge and Lasso Regression
Random Forest Regressor
Gradient Boosting Regressor

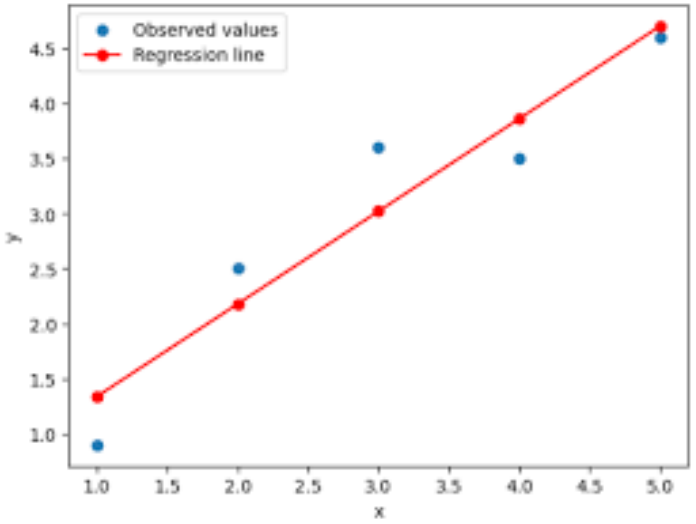
What is linear regression?

Linear regression is a fundamental statistical method and a machine learning algorithm used for modeling the relationship between a dependent variable (also called the target) and one or more independent variables (also called predictors or features).

The primary goal of linear regression is to find the best-fitting linear relationship that describes the data's patterns.

linear regression attempts to draw a straight line through the data points that best captures the overall trend or pattern of the data.

This line is often referred to as the "regression line" or "best-fit line."



Applicability of algorithms in AI

Natural Language Processing (NLP):

Problem: Processing and understanding human language text.

Algorithms:
Bag of Words (BoW)
Word Embeddings (Word2Vec, GloVe)
Recurrent Neural Networks (RNN)
Transformer Models (e.g., BERT, GPT)

What is Bag of Words (BoW)

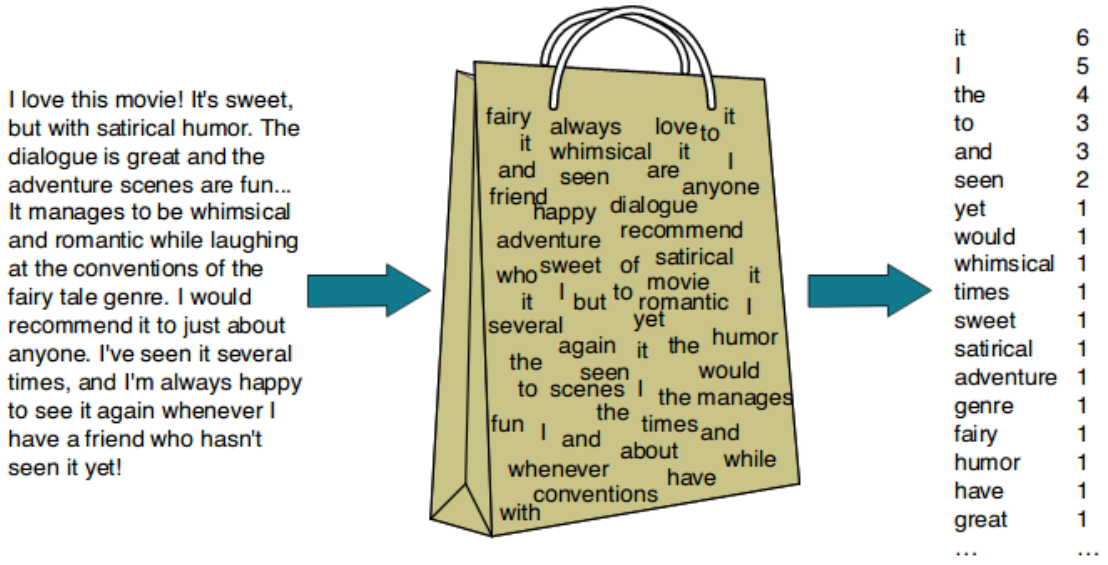
The "Bag of Words" (BoW) is a text representation technique used in natural language processing and information retrieval.

It's a simple and commonly used method for converting text data into numerical vectors that machine learning algorithms can work with.

BoW treats a document as an unordered collection of words and focuses solely on the occurrence of words in the text, disregarding grammar and word order.

how the Bag of Words technique works:

- Tokenization
- Vocabulary Creation
- Vectorization
- Sparse Representation
- Normalization



Applicability of algorithms in AI

Anomaly Detection:

Problem: Identifying rare or unusual data points.

Algorithms:
Isolation Forest
One-Class SVM
Autoencoders

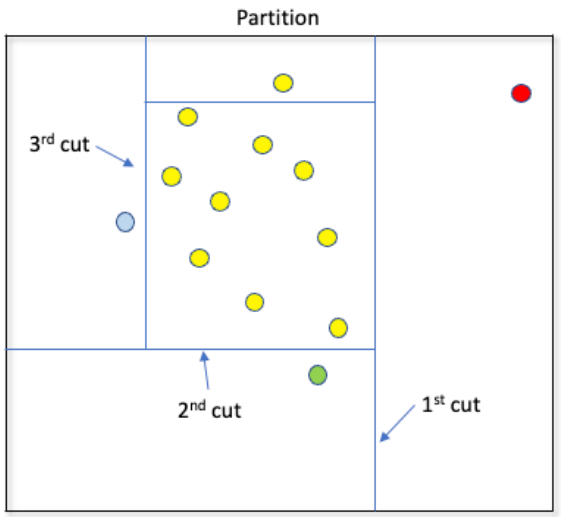
What is Isolation Forest?

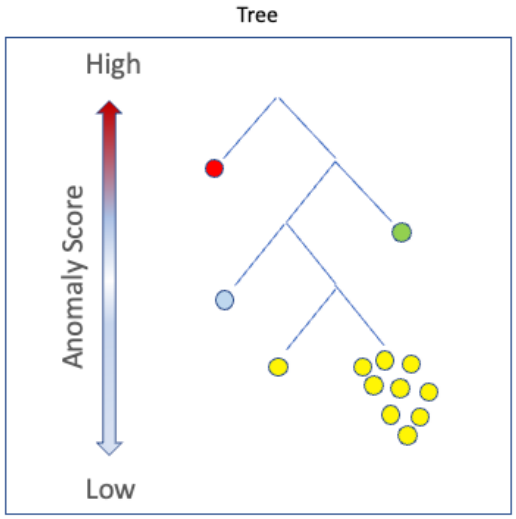
Isolation Forest is an unsupervised machine learning algorithm used for outlier detection and anomaly detection.

It is designed to identify observations in a dataset that are significantly different from the majority of the data points.

Key characteristics of the Isolation Forest algorithm include

- Tree-Based Approach
- Anomaly Scoring
- Path Length
- Scalability
- Parameter Tuning





Applicability of algorithms in AI

Recommendation Systems:

Problem: Suggesting items or content based on user preferences.

Algorithms:
Collaborative Filtering
Content-Based Filtering
Matrix Factorization

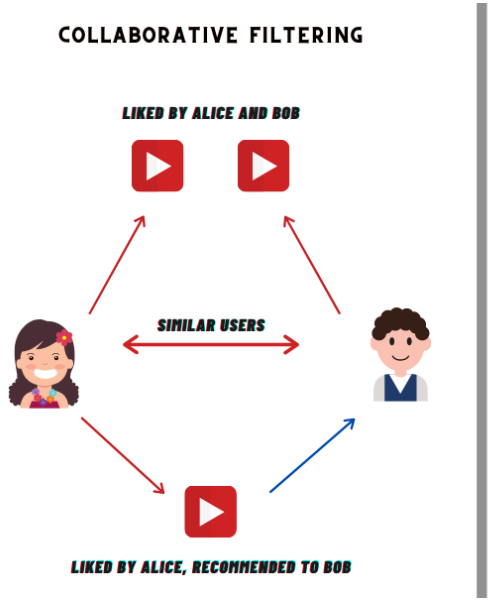
what is Collaborative Filtering?

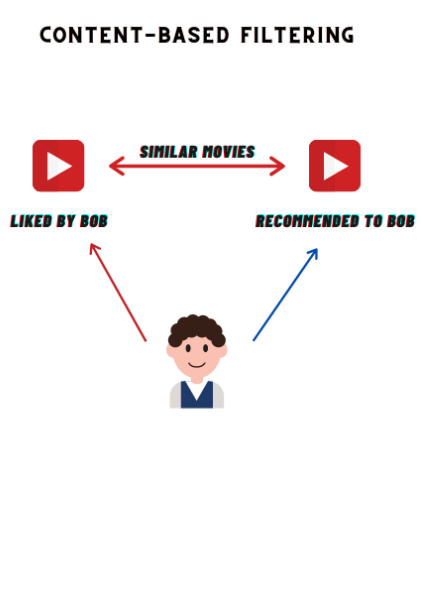
Collaborative Filtering is a popular recommendation technique used in the field of recommender systems.

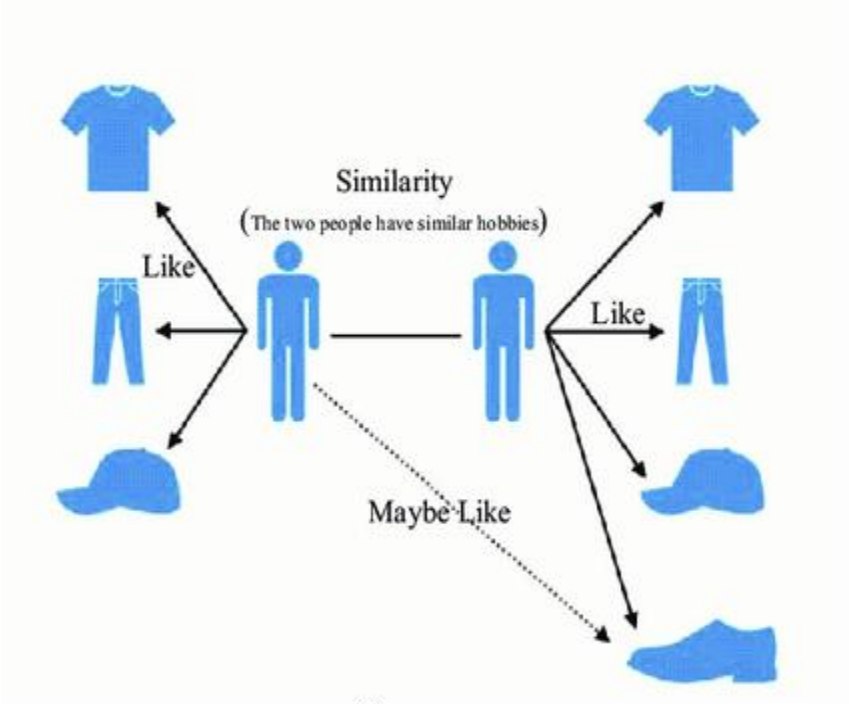
Its main objective is to provide personalized recommendations to users by leveraging the preferences and behaviors of similar users or items.

There are two main types of Collaborative Filtering:

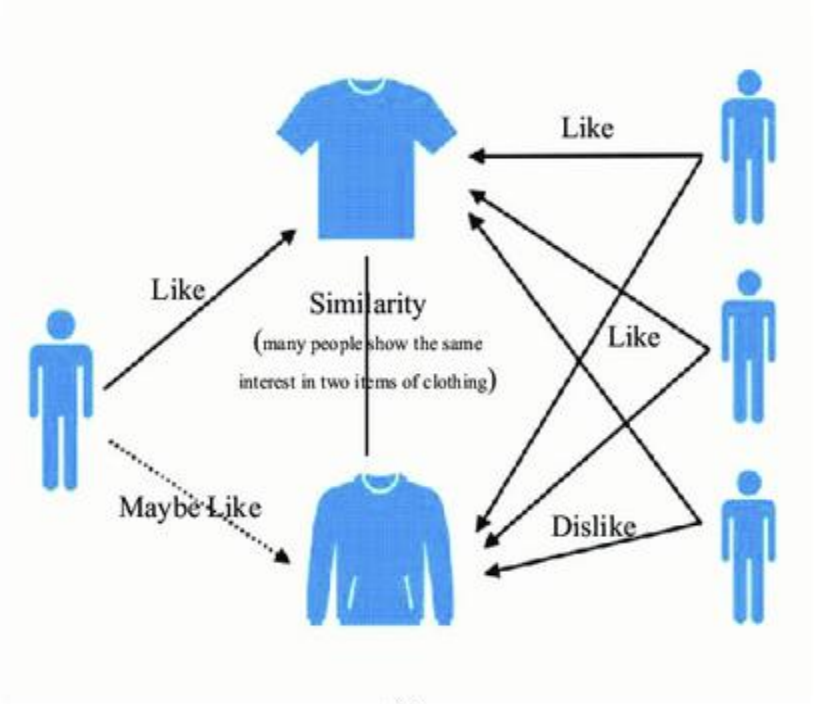
- User-Based Collaborative Filter
- Item-Based Collaborative Filteri



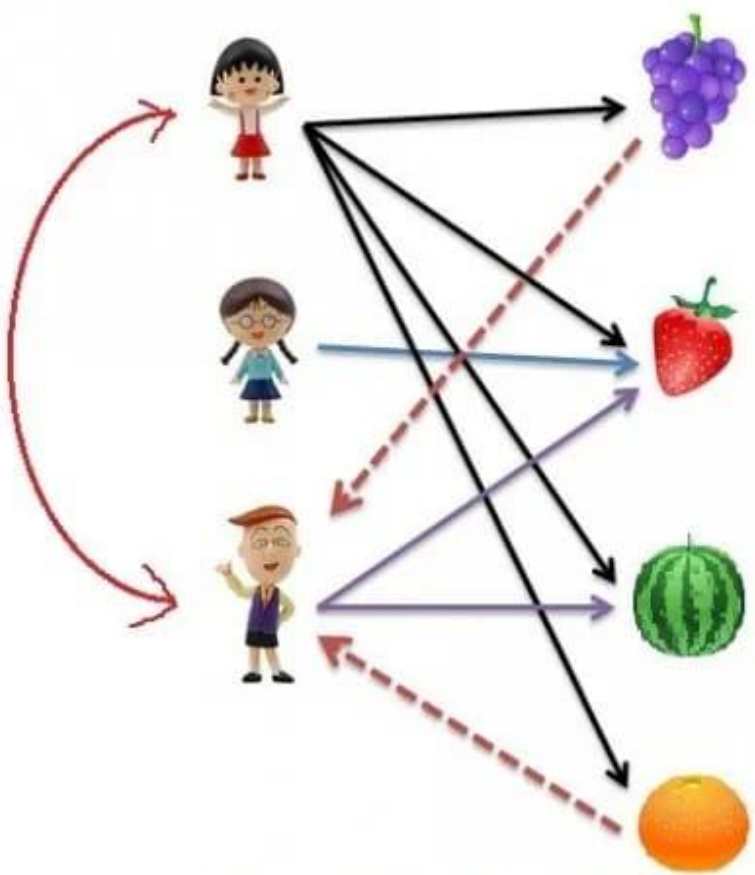



Similarity
(The two people have similar hobbies)
Like
Like
Maybe Like

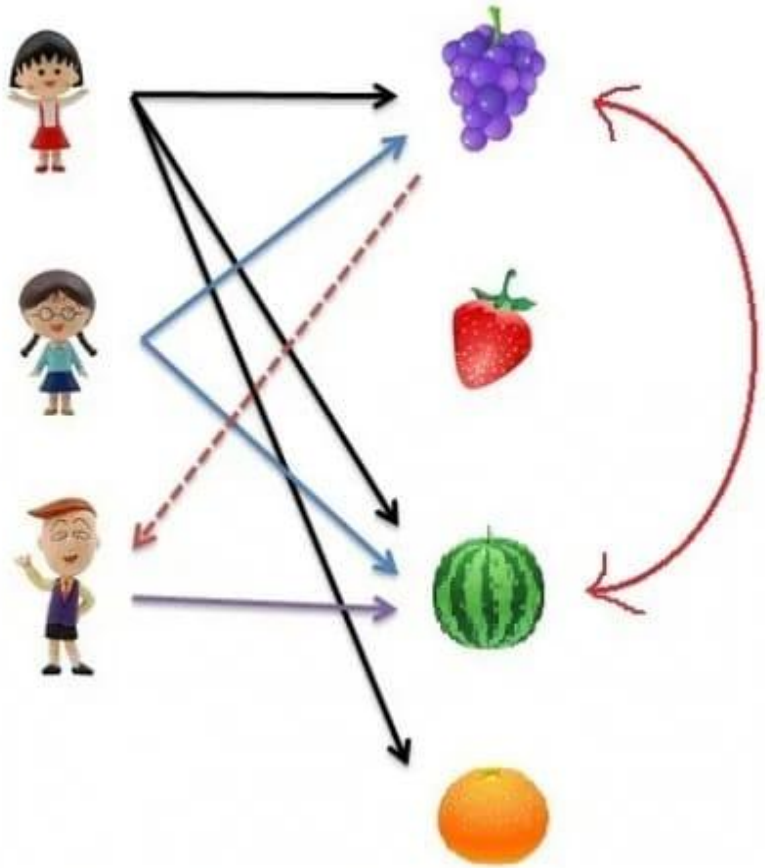
(a)


Like
Like
Similarity
(many people show the same
interest in two items of clothing)
Like
Maybe Like
Dislike

(b)



User-based filtering



Item-based filtering

Applicability of algorithms in AI

Image Processing:

Problem: Analyzing and extracting information from images.

Algorithms:
Image Segmentation
Convolutional Neural Networks (CNN)
Feature Extraction

What is Image Segmentation?

Image segmentation is a computer vision task that involves dividing an image into different segments or regions based on certain criteria.

The goal of image segmentation is to extract meaningful and semantically coherent regions from an image, enabling a more detailed understanding and analysis of its contents.

several types of image segmentation techniques:

- Thresholding
- Region-based Segmentation
- Edge Detection
- Watershed Segmentation
- Semantic Segmentation
- Instance Segmentation

G - Original Image

H - Semantic segmentation

I - Instance segmentation

J - Panoptic segmentation

Training a model using the selected algorithm

Training a model using a selected algorithm refers to the process of teaching the algorithm to learn patterns and relationships from a dataset.

The goal of training is to enable the algorithm to make accurate predictions or decisions on new, unseen data based on the patterns it has learned during the training process

- Feed the training data (features and labels) into the algorithm.
- The algorithm uses the training data to adjust its internal parameters through an iterative process.
- The process aims to minimize the difference between the model's predictions and the actual target values.

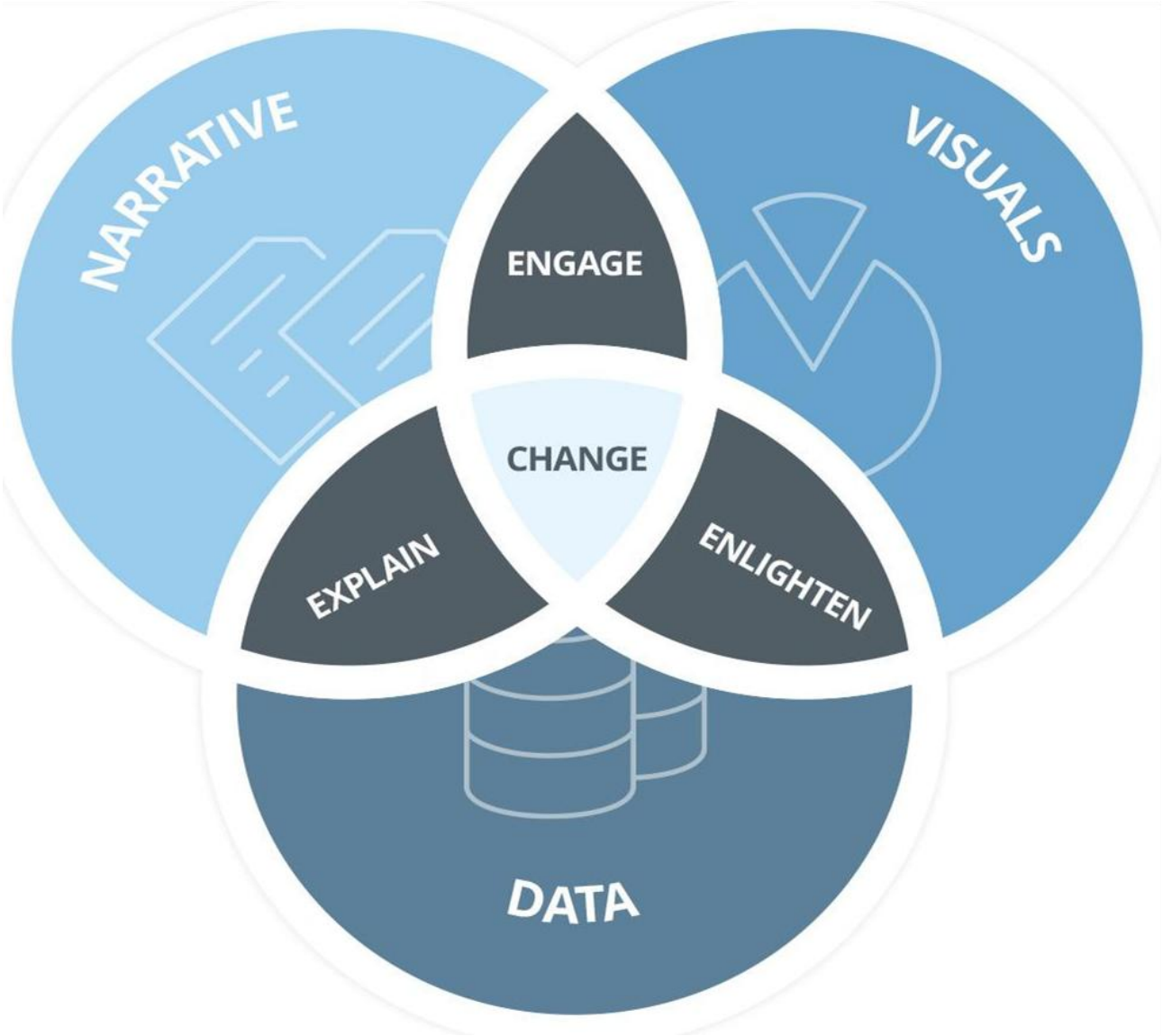
Tell data stories

Data storytelling is a powerful way to communicate insights, trends, and findings derived from data analysis.

It involves transforming raw data into a narrative that engages and informs your audience.

1. Choose Your Data	**6. Visualize the Data**	**11. Engage the Audience**
2. Identify Your Audience	**7. Craft the Narrative**	**12. Create a Flow**
3. Define the Objective	**8. Support with Insights**	**13. Highlight Takeaways**
4. Exploratory Data Analysis	**9. Use Real-Life Examples**	**14. Practice and Refine**
5. Create a Structure	**10. Maintain Clarity**	**15. Presentation**

Tell data stories



Evaluate model performance in AI

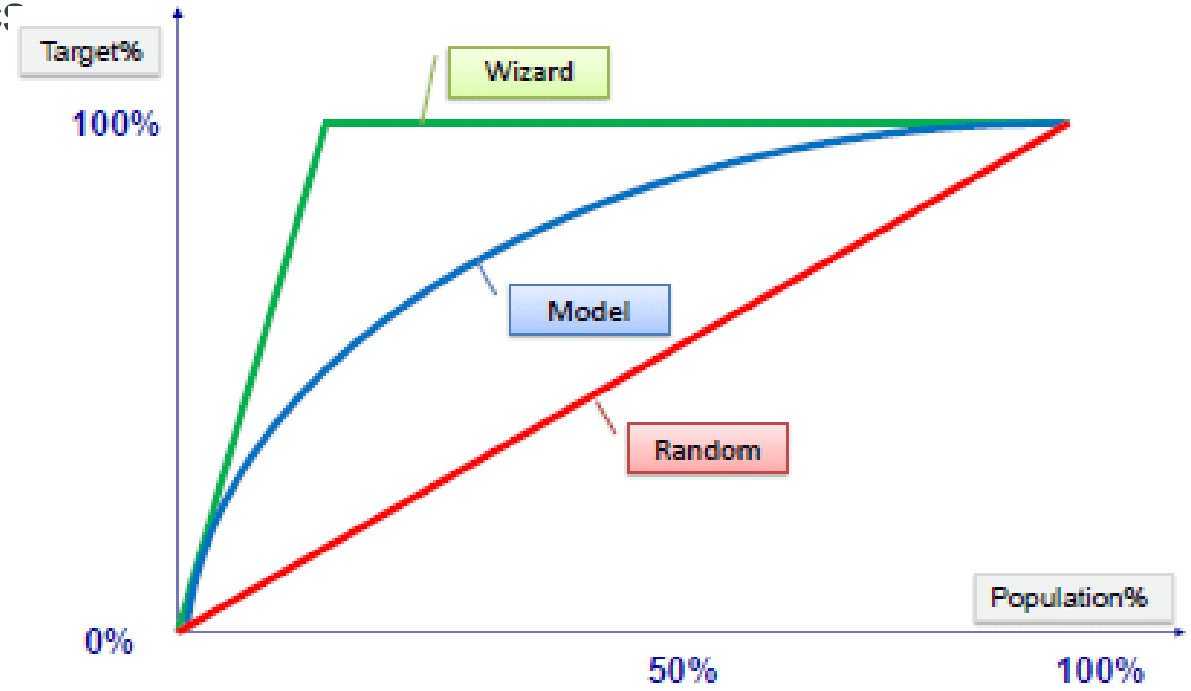
Evaluating model performance in artificial intelligence (AI) involves assessing how well a trained model performs on new, unseen data.

Proper evaluation is essential to determine the model's ability to generalize and make accurate predictions or decisions.

1. Classification Metrics
   - Accuracy
   - Precision
   - Recall (Sensitivity or True Positive Rate)
   - F1-Score
   - ROC Curve and AUC
2. Regression Metrics
   - Mean Squared Error (MSE)
   - Root Mean Squared Error (RMSE)
   - Mean Absolute Error (MAE)
   - R-squared (Coefficient of Determination)

Evaluate model performance in AI

3. Cross-Validation
4. Hyperparameter Tuning
5. Confusion Bias-Variance Trade-Off
6. Overfitting and Underfitting Detection
7. Domain-Specific Metrics
8. Visualizations
9. Comparing Models



Evaluate model sensitivity in AI

Evaluating model sensitivity involves assessing how variations in input data or parameters affect the model's predictions, outputs, or performance.

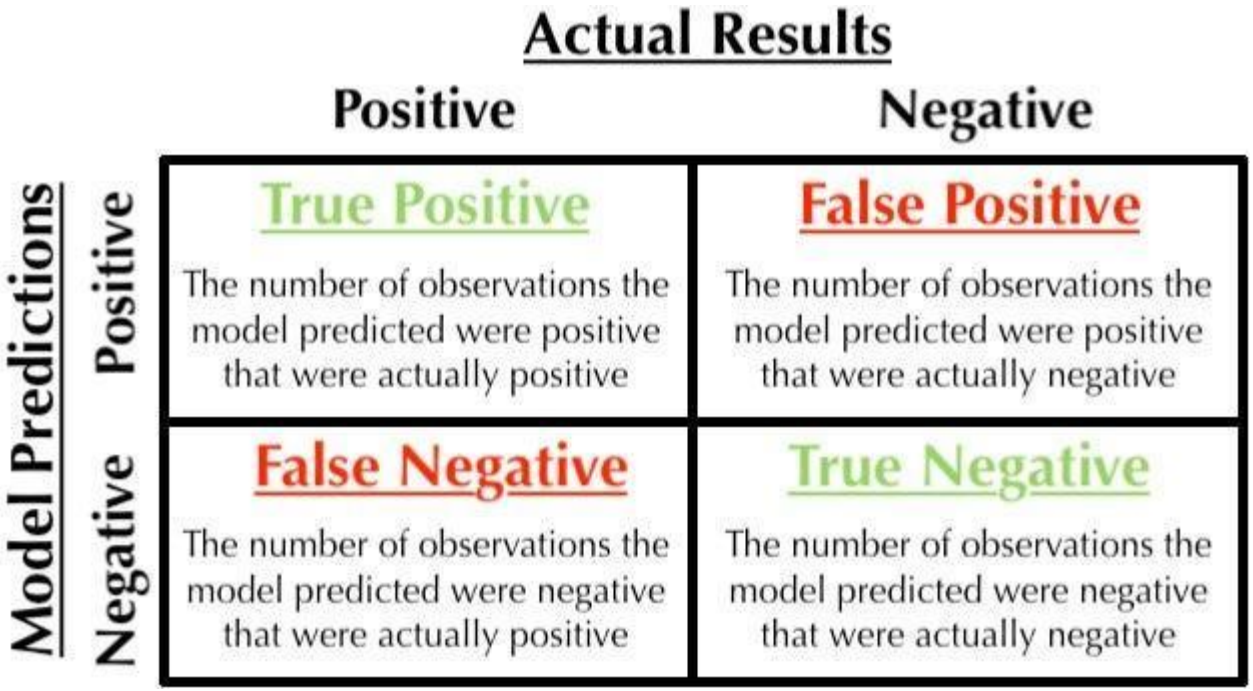
It helps understand the robustness and reliability of AI models.

1. Identify Sensitivity Factors	**6. Visualize Sensitivity**	**11. Decision Support**
2. Define Ranges and Intervals	**7. Analyze Findings**	**12. Interpretability**
3. Select Performance Metrics	**8. Validation and Robustness**	**13. Communication**
4. Vary Factors and Observe Outcomes	**9. Risk Assessment**	**14. Iterative Process**
5. Record Results	**10. Optimization and Fine-Tuning**	

Evaluate model sensitivity in AI

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$



Confirm adherence to regulatory requirements in AI

Ensuring adherence to regulatory requirements in AI is essential to maintain ethical standards, fairness, transparency, and legal compliance.

1. Understand Relevant Regulations	**6. Ethical Considerations**	**11. Cross-Functional Collaboration**
2. Data Privacy and Security	**7. Data Governance**	**12. Third-Party Services and Contracts**
3. Algorithmic Transparency and Explainability	**8. Model Accountability**	**13. Testing and Validation**
4. Fairness and Non-Discrimination	**9. User Consent and Transparency**	**14. Document Policies and Procedures:**
5. Intellectual Property Rights	**10. Auditability and Compliance Monitoring**	**15. Continuous Education and Updates**

Obtain stakeholder approval in AI

Obtaining stakeholder approval in AI projects is crucial to ensure alignment, transparency, and successful implementation.

Stakeholders can include executives, management, legal teams, data subjects, customers, and any other parties impacted by or involved in the AI project.

1. Identify stakeholders	**6. Demonstrate value and benefits**	**11. Establish a formal approval process**
2. Communicate project benefits and goals	**7. Discuss potential risks and mitigation**	**12. Document agreements and conditions**
3. Address concerns transparently	**8. Engage in open dialogue**	**13. Maintain ongoing communication**
4. Discuss ethical considerations	**9. Customize presentations for each group**	**14. Adapt based on stakeholder feedback**
5. Present data privacy measures	**10. Collaborate in decision-making**	


QUIZ

Question 1:

What is the primary goal of the k-means clustering algorithm?

a) Maximize inter-cluster variance
b) Minimize intra-cluster variance
c) Maximize total variance
d) Minimize total variance

Answer:

Question 2:

In linear regression, what is the purpose of the cost function?

a) To measure model accuracy
b) To compute gradient descent
c) To determine feature importance
d) To quantify the error between predictions and actual values

Answer:

Question 3:

Which technique is used to handle missing data in a dataset?

a) Data augmentation
b) Principal Component Analysis (PCA)
c) Imputation
d) One-Hot Encoding

Answer: 

Question 4:

What does the F1-score measure?

a) Accuracy of a classification model
b) Precision of a classification model
c) Recall of a classification model
d) Balance between precision and recall in a classification model

Answer:

Question 5:

What is the term for the unfair advantage or disadvantage that certain groups may experience due to biased AI systems?

a) Data bias
b) Algorithmic bias
c) Feature bias
d) Model bias

Answer:

Question 6:

Which of the following is NOT a consideration when ensuring fairness in AI?

a) Transparency
b) Explainability
c) Data privacy
d) Model complexity

Answer:

Question 7:

Why is obtaining stakeholder approval important in AI projects?

a) To increase project complexity
b) To satisfy regulatory requirements only
c) To align the project with organizational goals and values
d) To expedite project completion

Answer:

Question 8:

Who are stakeholders in AI projects?

a) Only the development team
b) Only end-users
c) Individuals and groups impacted by or involved in the project
d) Only executive management

Answer: c)

Any Questions?

Next session: Data Collection, Processing and Engineering