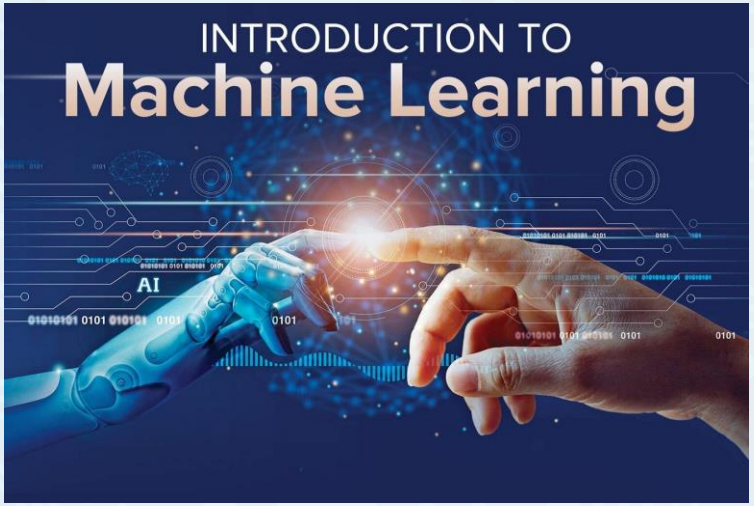
IEEE
Advancing Technology for Humanity

Introduction to machine learning (lecture 3)



By/ Aly Maher Abdel Fattah

IEEE
Galala University Student Branch

Lesson Outline

- *Case studies that use machine learning*
- *Using ML to predict a book's genre*
- *Using ML to detect spills*

Example1: House price prediction

Details:
- Room count
- Lot size

→ Model → House Value

Step 1: Define the Problem → Step 2: Build the Dataset → Step 3: Train the Model → Step 4: Evaluate the Model → Step 5: Use the Model

Details:
- Room count
- Lot size

→ Model → House Value

ML Task → Supervised Learning → Continuous Label (Regression)

Labeled Data

Step 1: Define the Problem | **Step 2: Build the Dataset** | Step 3: Train the Model | Step 4: Evaluate the Model | Step 5: Use the Model

Details:
- Room count
- Lot size

→ Model → House Value

Dataset:
- Collected
- Explored
- Cleaned
- Visualization
- Split

# of Rooms	Lot Size (ft^2)	House Value ($)
4	10,454	339,900
3	9,147	239,000
3	10,890	250,000
10	25,877	877,000

IEEE Advancing Technology for Humanity
IEEE Galala University Student Branch

Step 1: Define the Problem → Step 2: Build the Dataset → Step 3: Train the Model → Step 4: Evaluate the Model → Step 5: Use the Model

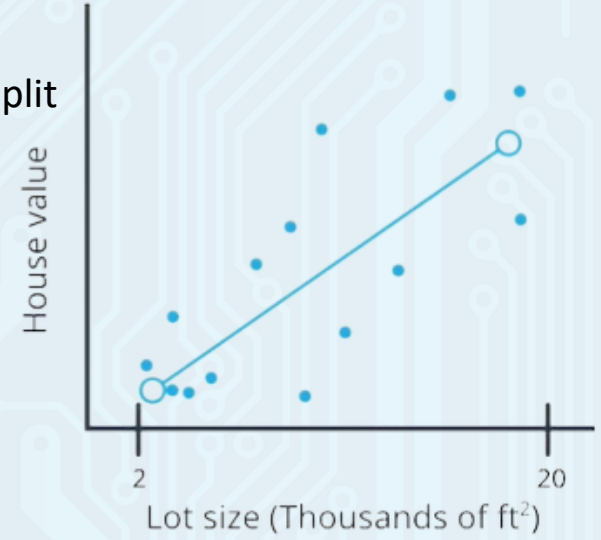
Details:
- Room count
- Lot size

→ Model →

House Value

Dataset: Collected, Explored, Cleaned, Visualization and Split

Model:
- Linear
- Python M.L. libraries



Step 1: Define the Problem → Step 2: Build the Dataset → Step 3: Train the Model → Step 4: Evaluate the Model → Step 5: Use the Model

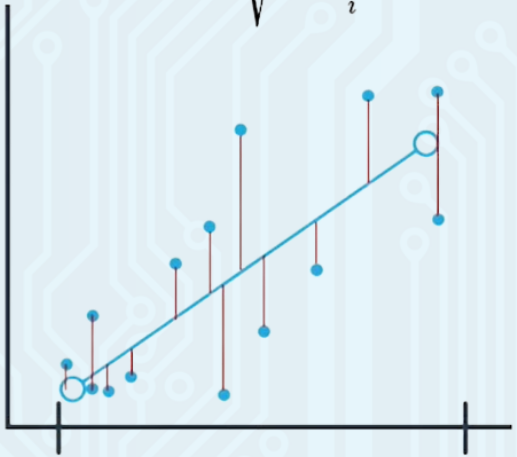
Details:
- Room count
- Lot size

→ Model → House Value

Dataset: Collected, Explored, Cleaned, Visualization and Split

Model: Linear

$$RMS = \sqrt{\frac{1}{n}\sum_{i} x_i^2}$$



Evaluation:
- Root Mean Square (RMS)
- Manual Threshold Check

IEEE Advancing Technology for Humanity
IEEE Galala University Student Branch

Step 1: Define the Problem → Step 2: Build the Dataset → Step 3: Train the Model → Step 4: Evaluate the Model → **Step 5: Use the Model**

Details:
- Room count
- Lot size

→ Model → House Value

Dataset: Collected, Explored, Cleaned, Visualization and Split

Model: Linear

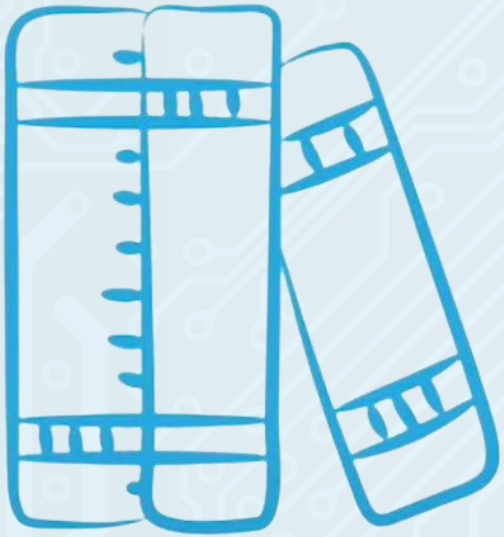
Evaluation: RMS & Manual Threshold Check

Inference: Try your model with real data

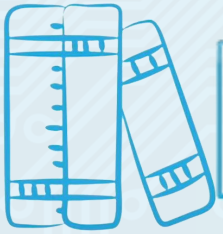
Room count = 4
Lot size = 20,000

→ Model → $435K

Example2: Microgenre Exploration



- Identify book trends
- Identify micro-genres

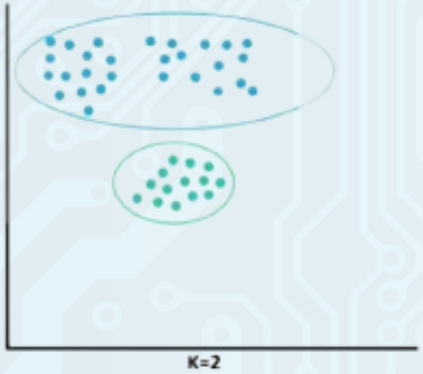


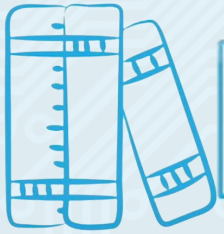
Step 1: Define the Problem → Step 2: Build the Dataset → Step 3: Train the Model → Step 4: Evaluate the Model → Step 5: Use the Model

Book Description → Model → Cluster Label

ML Task → Unsupervised Learning → Clustering

Unlabeled Data





Step 1: Define the Problem → **Step 2: Build the Dataset** → Step 3: Train the Model → Step 4: Evaluate the Model → Step 5: Use the Model

Dataset (800 Romance books):

- Collected
- Explored
- Cleaned
- Vectorized *(transformed into numbers)*

"Little did he know, she was secretly a vampire."

↓

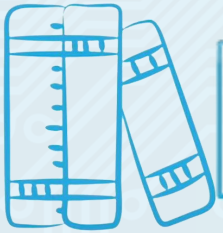
['little', 'does', 'he', 'know', 'she', 'is', 'secretly', 'vampire']

↓

Bag of Words

↓

[0, 0, 1, 0, 1, ...]

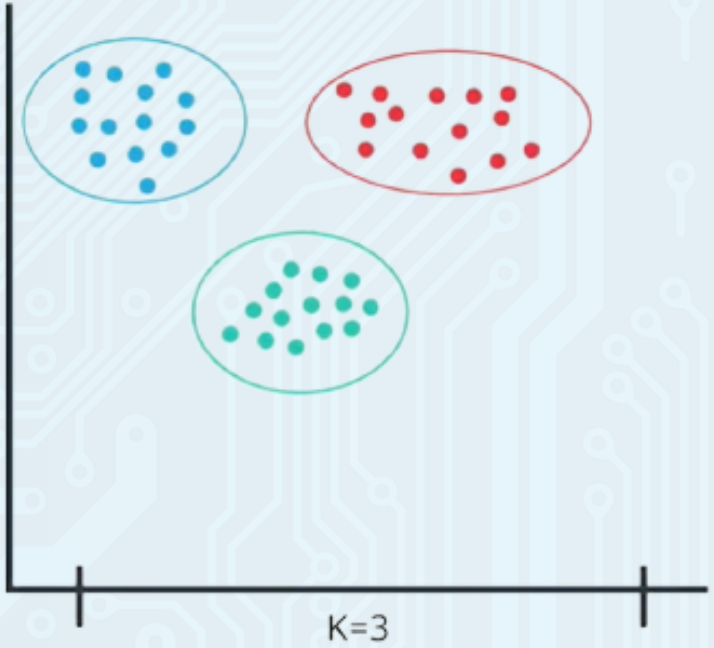


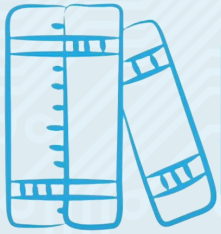
Step 1: Define the Problem → **Step 2: Build the Dataset** → **Step 3: Train the Model** → Step 4: Evaluate the Model → Step 5: Use the Model

Book Description → Model → Cluster Label

Dataset: Collected, Explored, Cleaned, Vectorized

Model: k-means



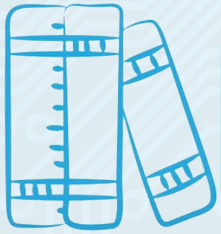


Step 1: Define the Problem → Step 2: Build the Dataset → Step 3: Train the Model → Step 4: Evaluate the Model → Step 5: Use the Model

Evaluation metrics:

- Silhouette coefficient
- Rand index
- Mutual information
- V-measure
- Completeness
- Fowlkes-Mallows
- Contingency Matrix
- Homogeneity
- Pair confusion matrix
- Calinski-Harabasz index
- Davies-Bouldin index

IEEE Advancing Technology for Humanity
IEEE Galala University Student Branch



Step 1: Define the Problem → Step 2: Build the Dataset → Step 3: Train the Model → Step 4: Evaluate the Model → Step 5: Use the Model

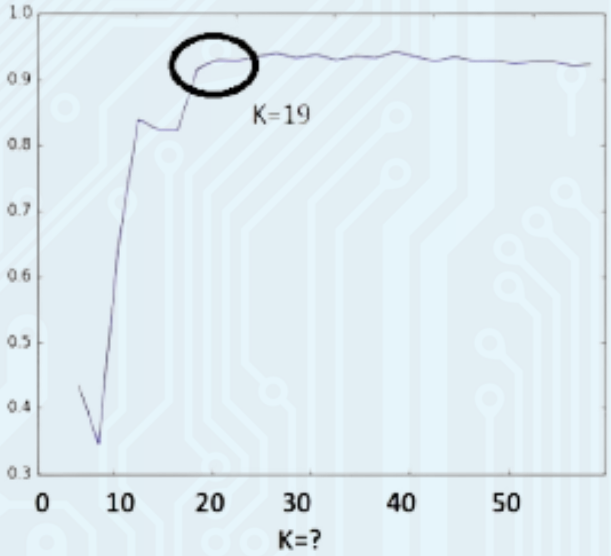
Book Description → Model → Cluster Label

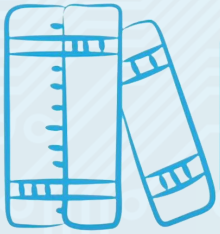
Dataset: Collected, Explored, Cleaned, Vectorized

Model: k-means

Evaluation:

- Silhouette Coefficient
- Manual Inspection





Step 1: Define the Problem → Step 2: Build the Dataset → Step 3: Train the Model → Step 4: Evaluate the Model → Step 5: Use the Model

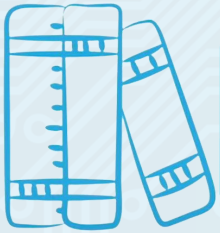
Dataset: Collected, Explored, Cleaned, Vectorized

Model: k-means

Evaluation:

- Silhouette Coefficient
- Manual Inspection

Cluster Label	Book Description
7	"Susan's crush just moved away.."
7	"Can Gurpuran and Jorge keep their relationship together three hundred miles apart?"
7	"When Ali's fiance got offered a new job in New York..."



Step 1: Define the Problem → Step 2: Build the Dataset → Step 3: Train the Model → Step 4: Evaluate the Model → **Step 5: Use the Model**

Dataset: Collected, Explored, Cleaned, Vectorized

Model: k-means

Evaluation:

- Silhouette Coefficient
- Manual Inspection

IEEE Advancing Technology for Humanity
IEEE Galala University Student Branch

Example3: Spill Detection From Video

- Fast response time needed
- Automated detection

Step 1: Define the Problem | Step 2: Build the Dataset | Step 3: Train the Model | Step 4: Evaluate the Model | Step 5: Use the Model

Image → Model → Contains Spill?

ML Task → Supervised Learning → Categorical Label (Image Classification)

Labeled Data

"contains spill"

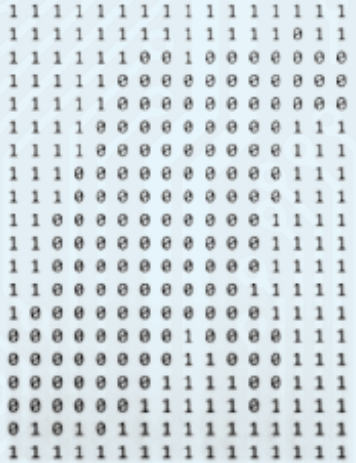
"does not contain spill"

Step 1: Define the Problem | **Step 2: Build the Dataset** | Step 3: Train the Model | Step 4: Evaluate the Model | Step 5: Use the Model

Dataset:

- Collected
- Explored and cleaned
- Split
- Vectorized

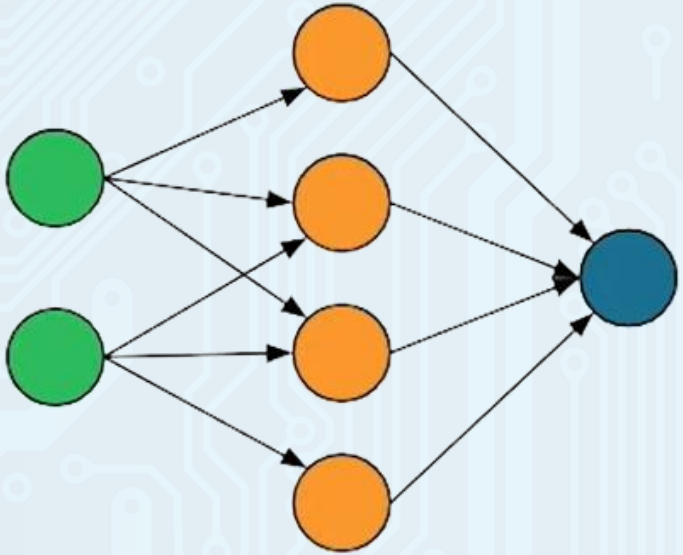
Spilled liquid



IEEE Advancing Technology for Humanity
IEEE Galala University Student Branch

Step 1: Define the Problem | Step 2: Build the Dataset | Step 3: Train the Model | Step 4: Evaluate the Model | Step 5: Use the Model

Input layer Hidden layer Output layer



Dataset: Collected, Explored, Cleaned, Split, and Vectorized

Model: Convolutional Neural Network (CNN)

IEEE Advancing Technology for Humanity
IEEE Galala University Student Branch

Step 1: Define the Problem | Step 2: Build the Dataset | Step 3: Train the Model | Step 4: Evaluate the Model | Step 5: Use the Model

IEEE Advancing Technology for Humanity

Evaluation Metrics:

ROC curve

Accuracy

Confusion matrix

Precision

Recall

False positive rate

False negative rate

Negative predictive value

Specificity

Log Loss

F1 Score

IEEE Galala University Student Branch

Dataset: Collected, Explored, Cleaned, Split, and Vectorized
Model: CNN
Evaluation:

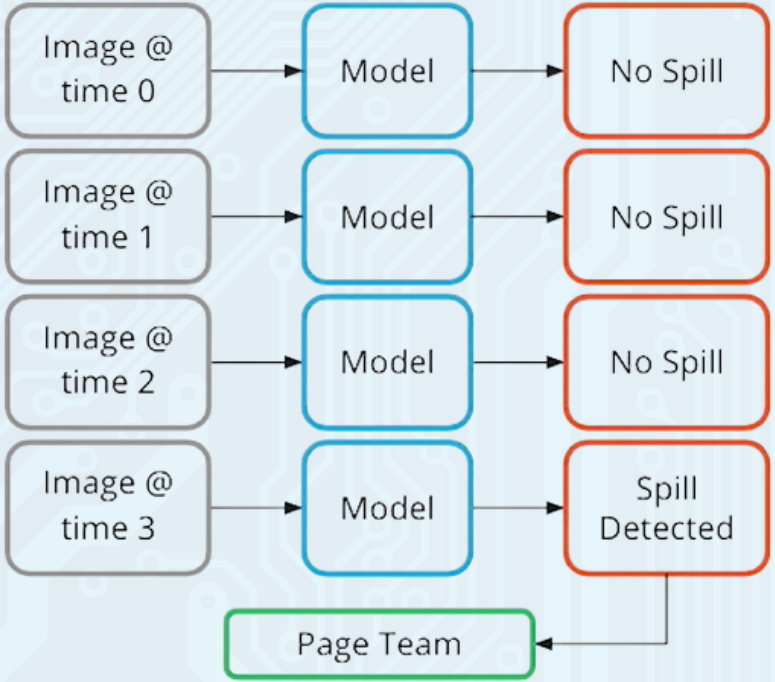
- Value of “accuracy”
- Precision
- Recall
- Manual testing

Step 1: Define the Problem → Step 2: Build the Dataset → Step 3: Train the Model → Step 4: Evaluate the Model → **Step 5: Use the Model**

Dataset: Collected, Explored, Cleaned, Split, and Vectorized

Model: CNN

Evaluation: Precision and Recall



IEEE
Advancing Technology for Humanity

Thank you

Any Question?

IEEE
Galala University Student Branch