code with edison

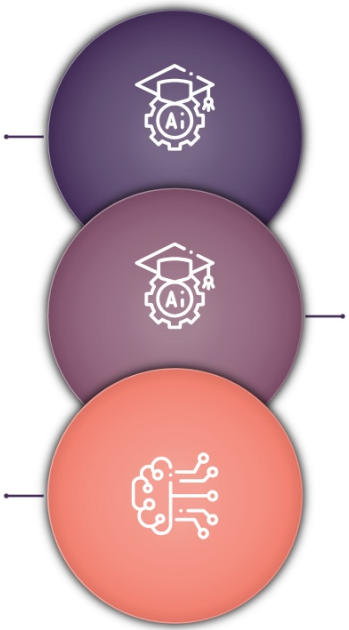
Introduction to Machine Learning Concepts

Apply Data Pre-processing

Introduction to Machine Learning Concepts

Definition of Machine Learning

Machine learning (ML) is an area of artificial intelligence where models learn patterns from data, enabling them to make predictions or decisions without explicit programming.



Applications of Machine Learning

ML is utilized in various applications, including Netflix recommendations, Google search, and autonomous cars.

Real-life Example of ML

An example of machine learning in daily life is spam filters in emails, which analyze incoming messages for patterns associated with spam and filter them accordingly.

Types of Machine Learning

01

Supervised Learning

Trained on labeled data. For example, a spam detection system learns from labeled emails (spam vs. not spam) and uses this data to classify new emails.

02

Unsupervised Learning

Works with unlabeled data. For instance, clustering algorithms in customer segmentation group customers by purchasing behavior without predefined categories.

03

Reinforcement Learning

Used in robotics and gaming, where an agent learns by receiving rewards for certain actions, like a self-driving car adjusting its route based on road conditions.

Preparing the Machine Learning Environment

Setting up Python and Libraries

Tools Required

The essential tools for setting up the machine learning environment include Python, Jupyter Notebook, and libraries such as Pandas, NumPy, and Scikit-Learn.

Exercise

Students should install the required libraries and test the environment by importing each library in a Jupyter Notebook.

Example

An example of loading a sample dataset is provided using the pandas library with the Iris dataset from Scikit-Learn, displaying its first few rows.

Data Collection & Acquisition

Understanding Data Sources and Types

Data Sources

Sources include social media, IoT sensors, and transactional databases.

Types of Data

Data can be categorized as Structured (e.g., tables in databases), Semi-structured (e.g., JSON files, XML), and Unstructured (e.g., images, audio, video).

Example of Data Collection

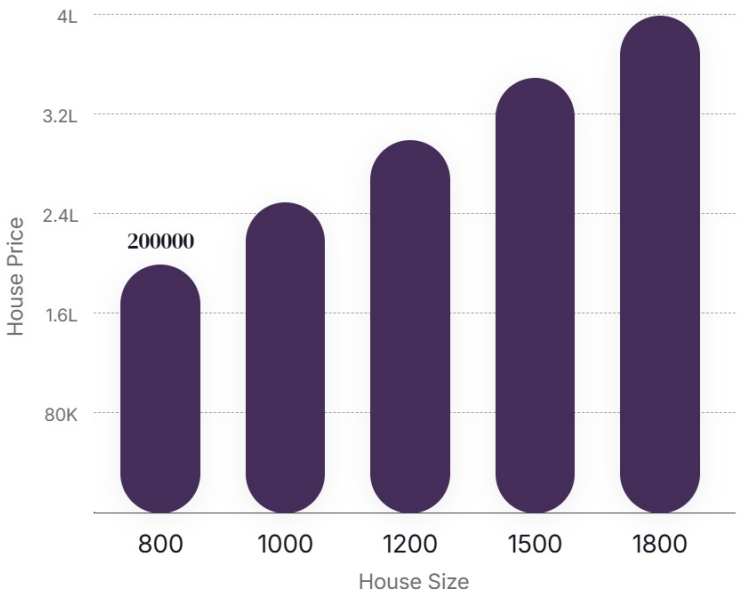
Amazon collects transactional data on customer purchases, product categories, and browsing history to make recommendations.

Classifying Data Exercise

Engage students in categorizing data they interact with daily, such as identifying photos as unstructured data and bank statements as structured data.

Data Visualization & Interpretation

Visualization Tools



Source: Companies Market Cap

Identifying relationships

Correlation between house size and price

- ✓ Students should identify relationships, such as the correlation between house size and price.

Subscribe

Identifying relationships

Correlation between house size and price

- ✓ Students should identify relationships, such as the correlation between house size and price.

Subscribe

Data Cleaning Techniques

Understanding the Importance and Methods of Data Cleaning

Purpose of Data Cleaning

Remove errors and inconsistencies to make data ready for analysis. For example, missing values in customer data can distort predictions in an ML model.

Example: Handling Missing Data

A dataset contains missing values in age columns. Students can learn to fill these values using mean imputation.

Normalization and Transformation

Normalizing scales values within a specific range is essential in models like neural networks. Transforming involves converting data types for compatibility, such as converting a categorical variable into dummy variables for analysis.

Machine Learning Algorithms

Understanding Supervised and Unsupervised Learning

01

Linear Regression

Used for predicting continuous values, like forecasting house prices based on features such as area, location, and number of rooms.

02

Classification with Logistic Regression

Predicts categorical outcomes (e.g., customer churn: yes or no). Students can use customer data to predict whether they'll stay or leave based on features like customer age and usage frequency.

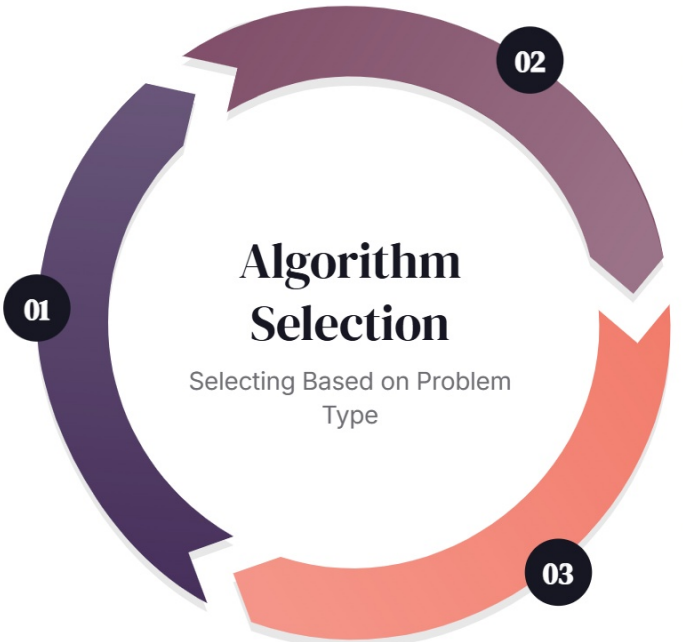
03

K-Means Clustering

Groups similar data points. An example includes clustering customer data into different segments based on purchasing behavior.

Algorithm Selection

Selecting Based on Problem Type



Classification problem

For a classification problem like spam detection, models such as Logistic Regression or Decision Trees are suitable.

Recommendation system

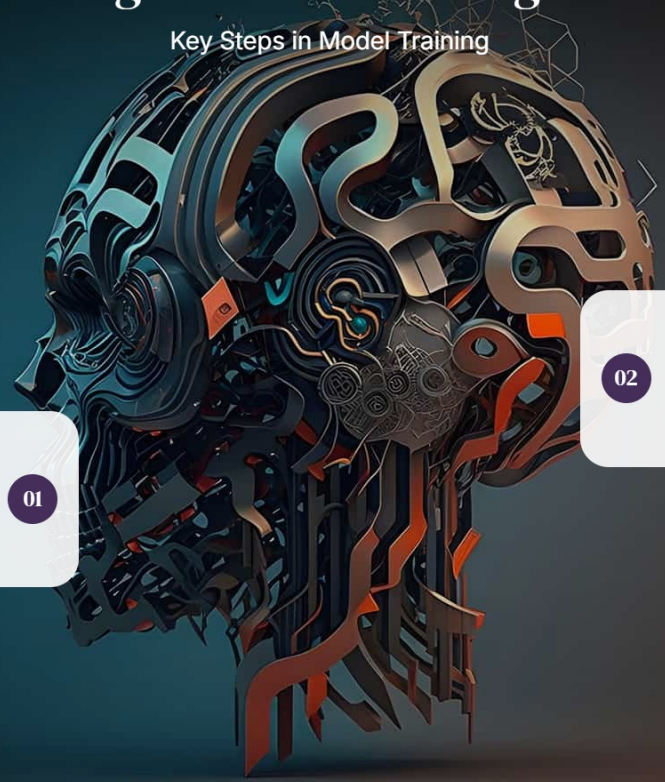
For recommendation systems, collaborative filtering algorithms are more suitable.

Limited computational resources

When computational resources are limited, prefer simpler algorithms like Logistic Regression or Naive Bayes over deep learning methods, which require high computational power.

Training Machine Learning Models

Key Steps in Model Training



Dataset Splitting

01

Divide data into training and testing sets, typically with 80% of the data used for training and 20% for testing.

Model Fitting

02

Train the model using the training dataset (X_train) and evaluate its performance on the testing dataset (X_test).

Evaluating Machine Learning Models

Understanding Key Evaluation Metrics

01

Accuracy

Percentage of correctly predicted instances.

02

Precision and Recall

Useful for imbalanced datasets, such as fraud detection.

03

Error Metrics

Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are important for regression problems.

04

Example of Model Evaluation

Use Scikit-Learn to compute accuracy, precision, and recall.

05

Scikit-Learn Code Example

Python code to compute accuracy using Scikit-Learn: `from sklearn.metrics import accuracy_score; y_pred = model.predict(X_test); accuracy = accuracy_score(y_test, y_pred); print("Accuracy:", accuracy)`.

Hyperparameter Tuning

Example of Hyperparameter Tuning with Grid Search

Grid Search Method

Grid Search is a systematic approach to tuning hyperparameters by evaluating a specified set of parameters through cross-validation.

Importance of Hyperparameter Tuning

Tuning parameters like learning rate in neural networks or max_depth in Decision Trees can significantly improve model performance.

Example Code Snippet

The following Python code demonstrates how to implement Grid Search using scikit-learn's GridSearchCV function.

Model Deployment as REST API

Deploy a trained model

Utilize Flask to deploy a trained model, enabling predictions via HTTP requests.

Flask application setup

Initialize a Flask app to handle incoming requests for model predictions.

Define prediction route

Create a '/predict' endpoint that accepts POST requests for prediction.

Handle JSON input

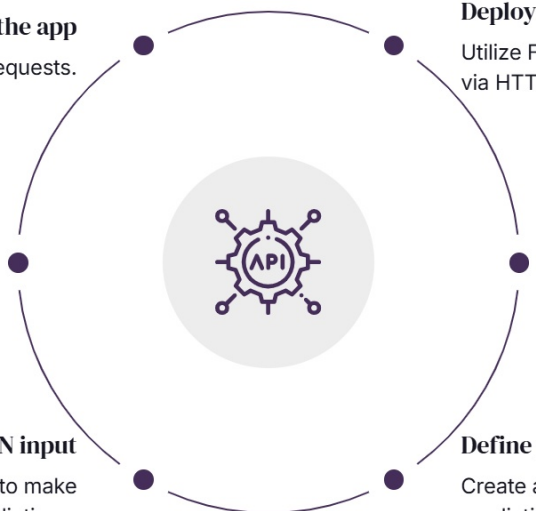
Extract features from the incoming JSON request to make predictions.

Return predictions

Format the model predictions into a JSON response for the client.

Run the app

Start the Flask application to listen for incoming requests.



Model Integration in a Web App

Use this API in a web application to provide real-time predictions, such as a sentiment analysis tool where users input text and receive a sentiment score.

Monitoring and Maintenance

Example of API Monitoring Use logging tools to monitor API performance, track response times, and record errors.

Delivering Predictions and Error Handling

Error Handling Example

Error Handling in Predictions

Implement error handling to manage incorrect data formats effectively.

Example of Bad Request Handling

The provided Python code demonstrates how to handle bad requests with a 400 status code.

Returning Error Responses

Use appropriate response codes to inform users about the nature of the error.

Classroom Activities and Assessments

Project Implementation

Students can implement a model from start to finish, choosing their dataset, applying data cleaning, training a model, and deploying it as an API.

Quizzes and Exams

Test knowledge on concepts, algorithms, and code implementation.

Exploratory Data Analysis Assignment

Perform exploratory data analysis (EDA) on a real-world dataset and document findings.

Contact Us for More Info

We are here to assist you with any questions or information you need.