

O'REILLY®

Machine Learning for Data Analytics with Python

Dr. Chester Ismay





Welcome and Introduction

Dr. Chester Ismay

- PhD in Statistics
- Worked in academia, online education, corporate training, tech bootcamps, and independent consulting
- Currently, freelance data scientist and educator
- Fun Fact: Slept a night or eaten a meal in all 50 US states





Learning Objectives

By the end of this course, you will be able to:

- Leverage machine learning for business decision-making.
- Identify real-world applications of machine learning in business.
- Utilize Python's machine learning ecosystem.
- Apply data preprocessing techniques to prepare datasets for ML models.



Course schedule

- Intro: Getting Started with Machine Learning for Data-Driven Decisions
- Module 1: Data Understanding and Preprocessing for Machine Learning
- Module 2: Supervised Learning for Business Decisions
- Module 3: Unsupervised Learning and Pattern Discovery in Business
- Module 4: Implementing and Evaluating ML Models

Poll in ON24 (to be entered by O'Reilly team)

Which ML application/topic are you most eager to learn more about today ?

1. Data Exploration and Preparation
2. Predictive Modeling for Strategic Insights
3. Uncovering Patterns and Clusters in Data
4. Model Execution and Performance Analysis
5. Enhancing Model Performance with Fine-Tuning

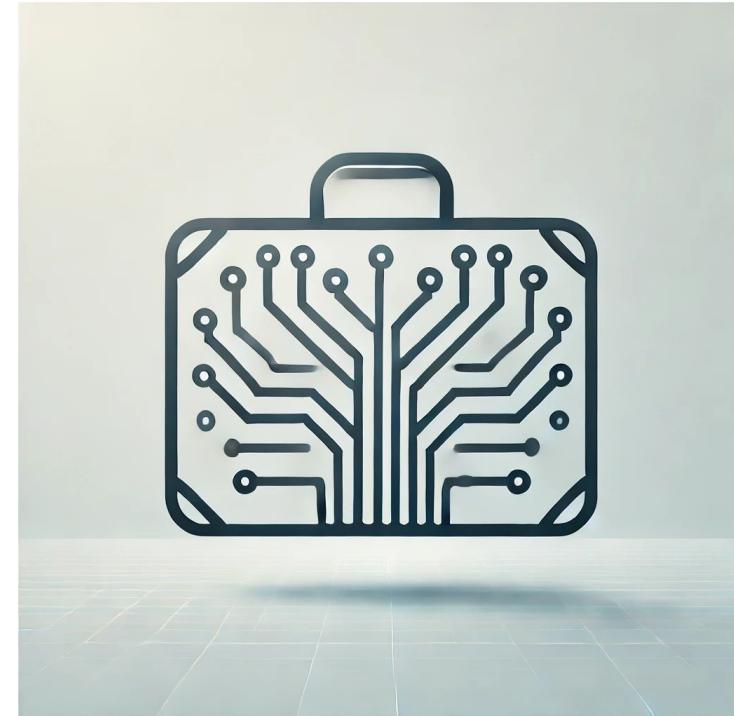
Intro

Getting Started with Machine
Learning for Data-Driven Decisions



How is Machine Learning Used in Business?

- Machine learning (ML) as a tool for predictive analytics
- Automating data-driven decision-making
- Enhancing business operations and customer insights



Practical Applications of ML in Business Analytics

- Data-driven business decision-making
- Real-world examples
 - Customer segmentation (Retail)
 - Fraud detection (Finance)
 - Demand forecasting (Supply Chain)



Python for Machine Learning in Business Analytics

- Common ML libraries in Python
 - Scikit-learn: Machine learning algorithms and model evaluation
 - Pandas: Data manipulation
 - NumPy: Numerical computing
 - Matplotlib & Seaborn: Data visualization



matplotlib



seaborn

Walkthrough and Exercise #1

Setting Up the Python Environment for ML

By completing this exercise, you will be able to

1. Install and import essential ML libraries
2. Verify successful package installation
3. Load a sample dataset for ML exploration



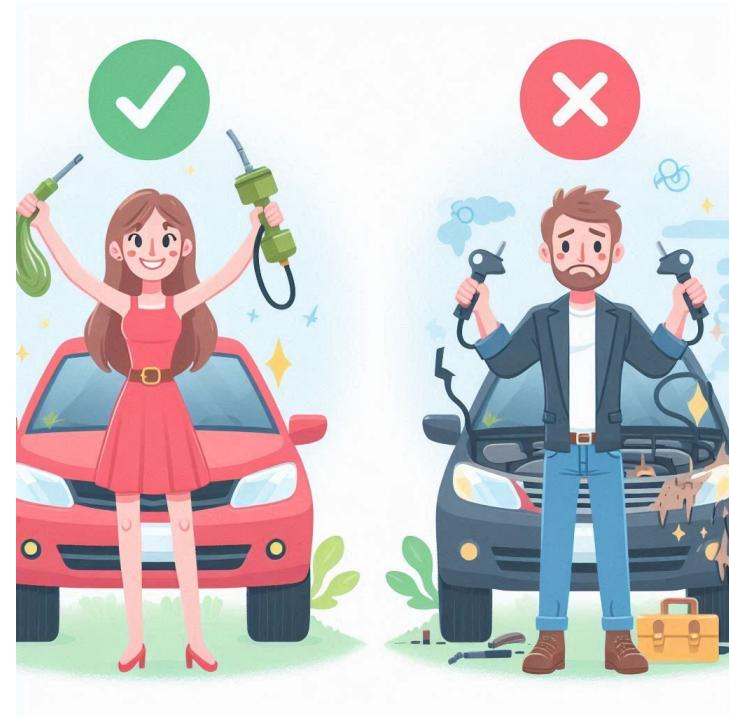
Module 1

Data Understanding and
Preprocessing for Machine
Learning



The Role of Data in Machine Learning

- Machine learning (ML) depends on high-quality data
- The “**Garbage In, Garbage Out**” principle
- Understanding **structured** vs. **unstructured** data in business



How Data is Collected and Prepared for ML

- Common data sources for business ML applications
- Why preprocessing is essential
- Key techniques for data cleaning



Exploring and Understanding Your Data

- Exploratory Data Analysis (EDA) is vital for ML
- Methods for summarizing and visualizing data
- Identifying trends and business insights



How Data Insights Shape Business Strategy

- Finding actionable insights from data
- Case study: Using EDA to predict customer churn
- Connecting ML insights to business goals



Walkthrough and Exercise #2

Exploring and Preprocessing Data with Pandas & Matplotlib

By completing this exercise, you will be able to

1. Inspect a dataset using Pandas
2. Handle missing values and clean data
3. Create visualizations to identify key business trends





Questions and Answers

Anything I can clear up regarding the *Intro* and *Module 1* content?

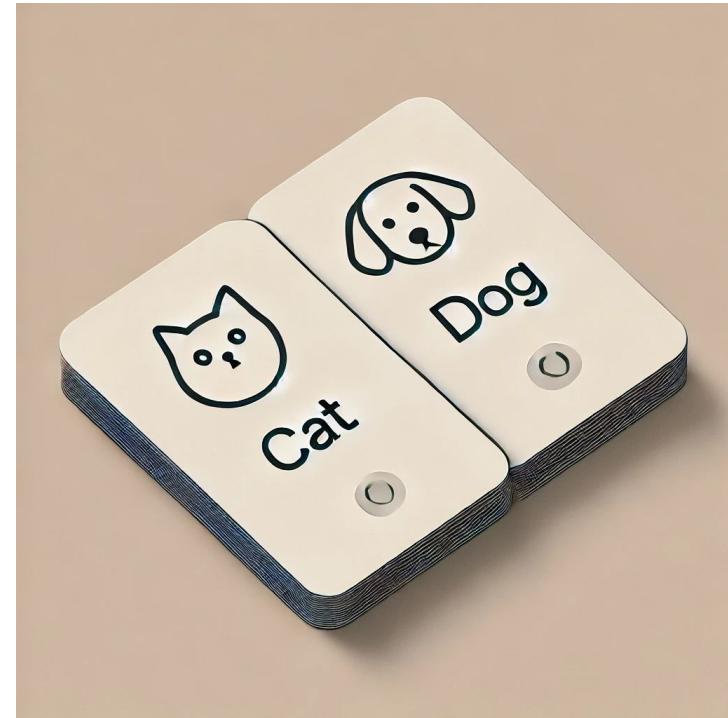
Module 2

Supervised Learning for Business
Decisions



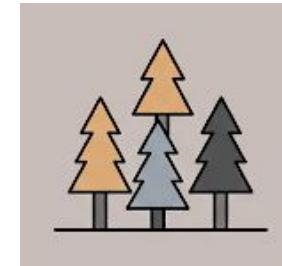
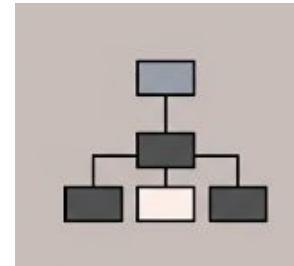
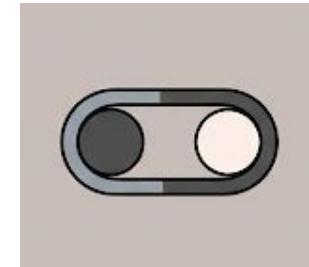
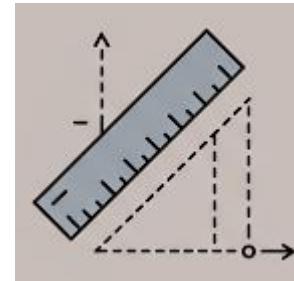
Machines Learning from Examples

- Definition of supervised learning
- Role of labeled data in supervised learning
- Types of supervised learning: classification and regression



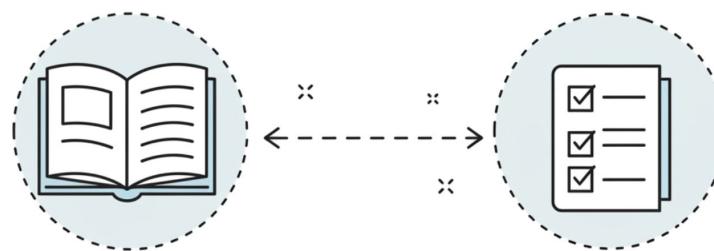
Key Supervised Learning Algorithms

- Linear regression
- Logistic regression
- Decision trees
- Random forests



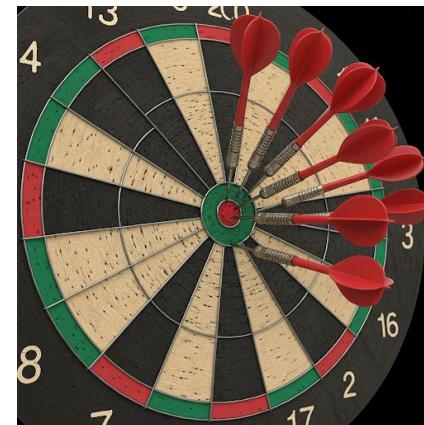
Ensuring Model Generalization

- The difference between training and validation data
- Why we don't evaluate models on the same data they were trained on
- How validation data helps assess model performance



Evaluating Regression Models

- How to assess model performance in regression
- Understanding **R-squared** and **Mean Absolute Error (MAE)**
- Choosing the right metric for business applications



Walkthrough and Exercise #3

Build a Regression Model for Pricing Optimization

By completing this exercise, you will be able to

1. Use scikit-learn to build a simple linear regression model
2. Train the model using training data and evaluate it on validation data
3. Optimize pricing based on historical data



Primer on Classification Techniques

- Definition of classification in supervised learning
- Applications of classification in business



Evaluating Classification Models

- Understanding model evaluation metrics on validation data
- The role of the confusion matrix
- Accuracy vs. precision vs. recall in business decisions

		Predicted Class	
		Yes	No
Actual Class	Yes	TP	FN
	No	FP	TN

Walkthrough and Exercise #4

Implement a Classification Model for Customer Churn Prediction

By completing this exercise, you will be able to

1. Build a classification model to predict whether a customer will churn
2. Split data into training and testing sets to assess model generalization
3. Evaluate model performance using key classification metrics



Business Applications of Supervised Learning

- Forecasting business trends
- Customer segmentation for targeted marketing
- Fraud detection and risk management





Questions and Answers

Anything I can clear up regarding the *Module 2* content?



Review of Modules 1 & 2



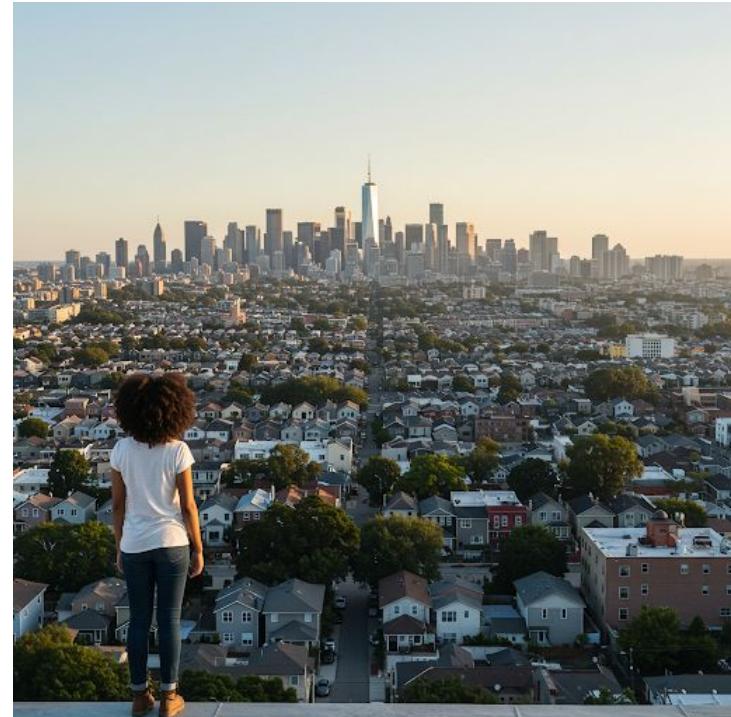
Module 3

Unsupervised Learning and Pattern
Discovery in Business



Unsupervised Learning and Its Business Applications

- Definition of unsupervised learning
- Key differences between supervised and unsupervised learning
- Common applications in business



Clustering for Customer Insights

- Definition of clustering
- Business use cases (customer segmentation, product recommendations)
- Overview of K-Means clustering



Walkthrough and Exercise #5

Exploring K-Means Clustering for Customer Segmentation

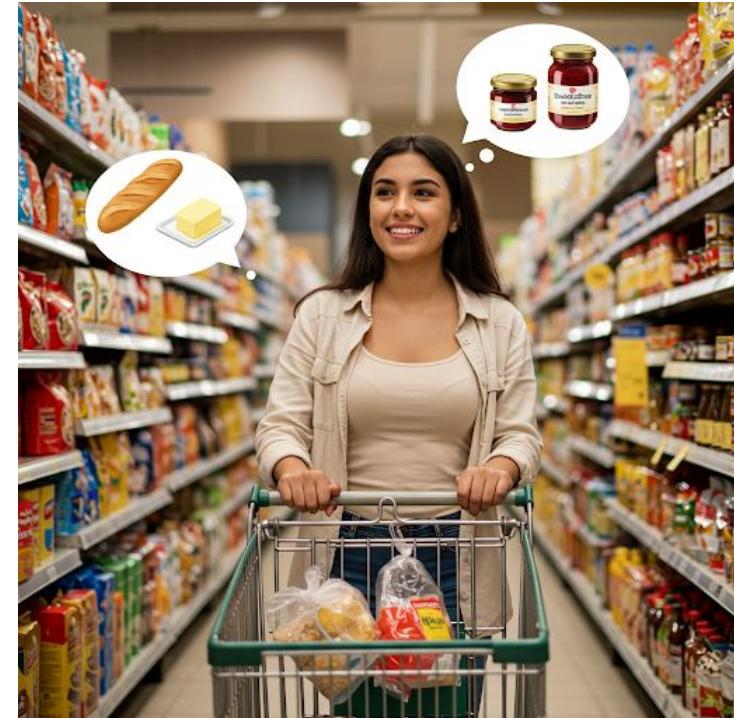
By completing this exercise, you will be able to

1. Apply K-Means clustering to segment customers.
2. Visualize customer groups to derive business insights.



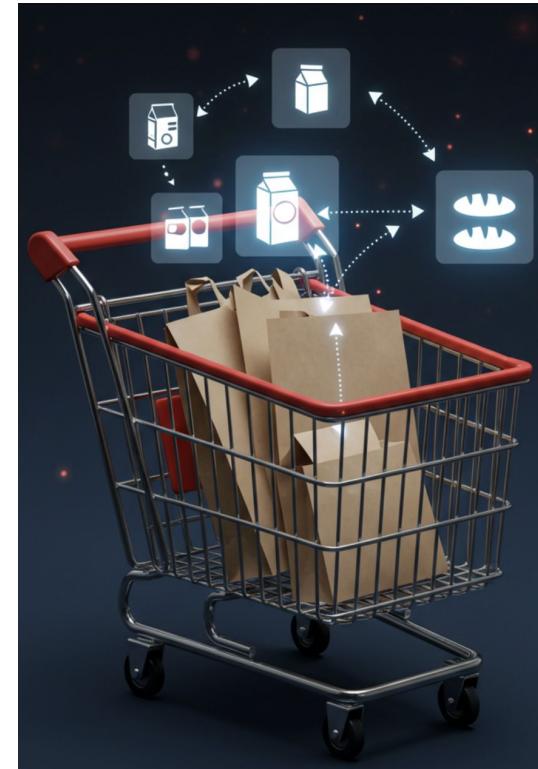
Association Rule Learning for Market Basket Analysis

- Definition of association rule learning
- Business applications (recommendation systems, cross-selling strategies)



The Apriori Algorithm for Finding Patterns

- Steps in discovering association rules
 - Find frequent itemsets
 - Generate association rules
 - Measure rule strength



Walkthrough and Exercise #6

Market Basket Analysis with Apriori Algorithm

By completing this exercise, you will be able to

1. Apply association rule learning to uncover purchasing patterns.
2. Extract actionable insights for cross-selling strategies.



Real-World Applications of Unsupervised Learning

- Case study: Amazon's staff-less stores and personalization
- How clustering helps businesses tailor marketing strategies
- Best practices for applying unsupervised learning





Questions and Answers

Anything I can clear up regarding the *Module 3* content?

Module 4

Implementing and Evaluating ML
Models



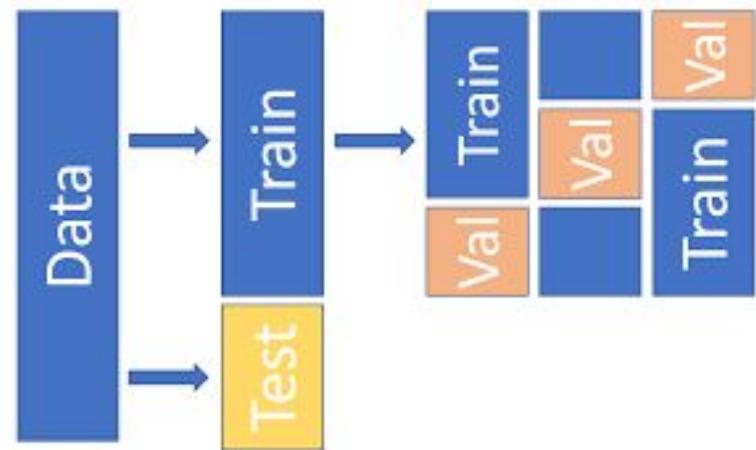
Picking the Right ML Approach

- Factors influencing model selection
- Trade-offs: accuracy vs. interpretability vs. complexity
- Supervised vs. unsupervised models for business



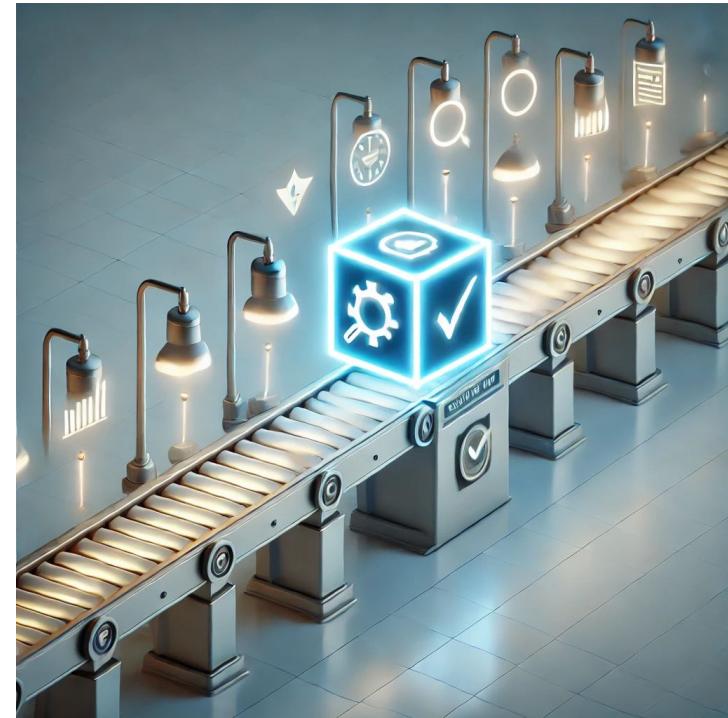
Cross-Validation Fundamentals

- What is cross-validation?
- Why is it better than a single train-test split?
- How it helps prevent overfitting



Applying Performance Metrics with Cross-Validation

- How accuracy, precision, recall, and RMSE are used with cross-validation
- Avoiding misleading model performance metrics



Walkthrough and Exercise #7

Exploring Cross-Validation for Model Evaluation

By completing this exercise, you will be able to

1. Apply cross-validation to assess model performance.
2. Compare multiple evaluation metrics.



Hyperparameter Tuning for Model Optimization

- What are hyperparameters?
- Grid search vs. Randomized search
- Business impact of model tuning



Walkthrough and Exercise #8

Hyperparameter Tuning with GridSearchCV

By completing this exercise, you will be able to

1. Optimize a model's hyperparameters.
2. Use GridSearchCV to find the best parameters.



Deploying ML Models in Business Settings

- Steps for ML model deployment
- Common deployment platforms (Flask, FastAPI, cloud services)
- Challenges in deploying ML models





Questions and Answers

Anything I can clear up regarding the *Module 4* content?



Review of Modules 3 & 4





Learning Objectives

By the end of this course, you will be able to:

- Leverage machine learning for business decision-making.
- Identify real-world applications of machine learning in business.
- Utilize Python's machine learning ecosystem.
- Apply data preprocessing techniques to prepare datasets for ML models





Takeaways

Additional resources:

- [pandas](#)
- [numpy](#)
- [matplotlib / seaborn](#)
- [scikit-learn](#)
- [mlxtend](#)
- More details on deploying models ([Practical MLOps O'Reilly book](#) and [Designing Machine Learning Systems O'Reilly book](#))

LinkedIn: <https://www.linkedin.com/in/chesterismay/>

Personal website: <https://chester.rbind.io/>

Images made by DALL-E/Microsoft Copilot/Google Gemini or found via Creative Commons search

O'REILLY®