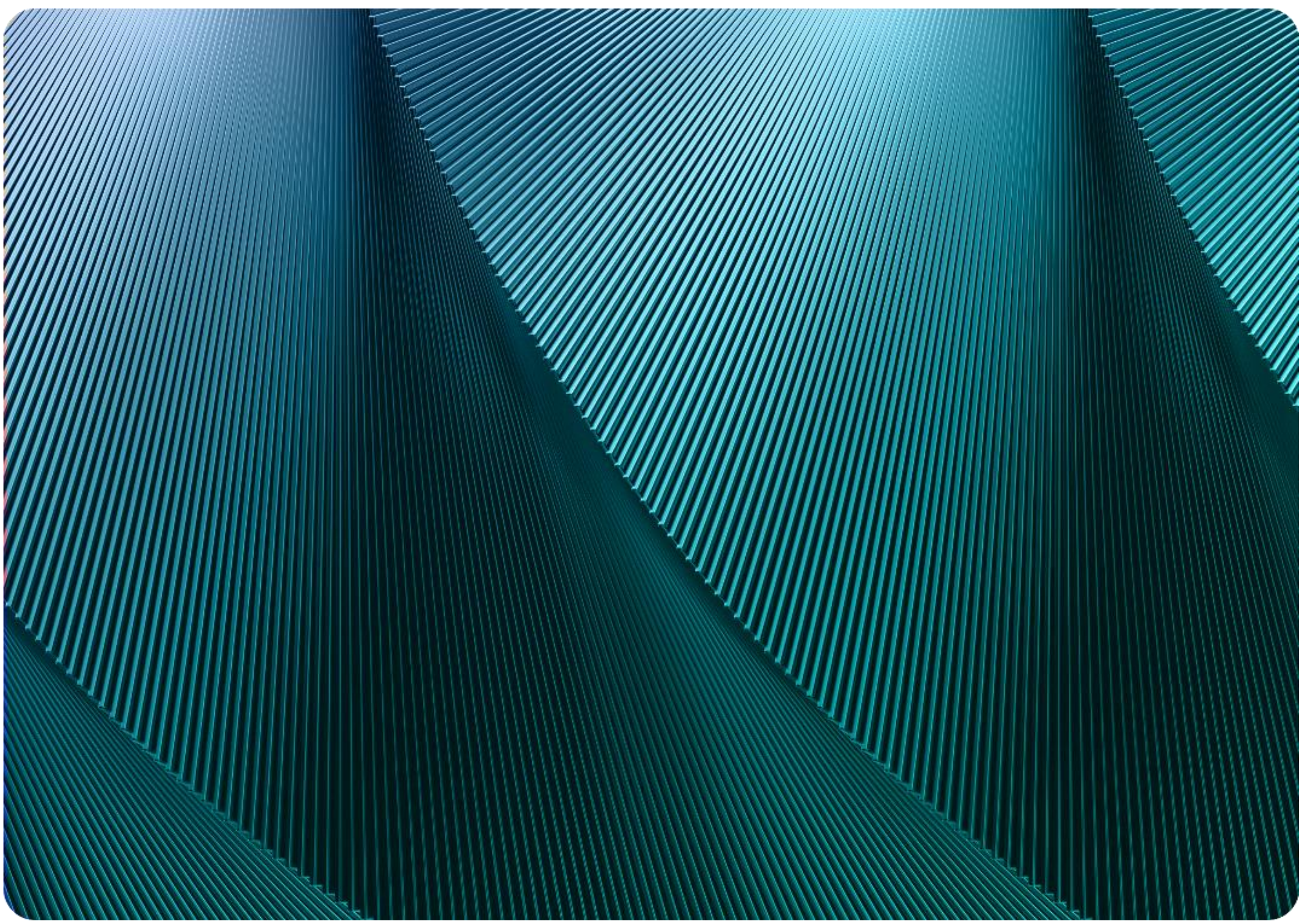
Resource provided by

Course Outline Details

Introduction to Machine Learning

This introductory course provides hands-on experience with machine learning concepts, techniques, and real-world applications.



Course Outline

Week starting on / Module	Module Topic
Week 1	Mod 01: Welcome, Introduction, and What is Machine Learning?
Week 2	Mod 02: Tools Used in Machine Learning
Week 3	Mod 03: Types of Machine Learning and Successful ML Solutions
Week 4	Mod 04: ML Lifecycle, Overfitting, and Underfitting
Week 5	Mod 05: AutoML and AutoGluon
Week 6	Mod 06: Data in Machine Learning & Exploratory Data Analysis (EDA)
Week 7	Mod 07: Basic Feature Engineering
Week 8	Mod 08: Tree-Bades Models, Optimization, and Regularization
Week 9	Mod 09: Hyperparameter Tuning
Week 10	Mod 10: Ensembling (Bagging, Boosting, Stacking)
Week 11	Mod 11: Introduction to Fairness and Bias Mitigation in ML
Week 12	Mod 12: Fair Model Formulation and Metrics
Week 13	Mod 13: Fairness Criteria
Week 14	Mod 14: Bias Mitigation During Pre-processing, Training, and Post-Processing
Week 15	Mod 15: Course Wrap-Up
Week 16	Final Project/Final Exam

Course Outline Details

Module 01: Welcome, Introduction, and What is Machine Learning?

- **Introduction to the Course**: Overview of the course structure, objectives, and expectations
- **What is Machine Learning**: Define machine learning and its importance. Discuss the difference between machine learning, AI, and traditional programming
- **Historical Context**: Brief history and evolution of machine learning
- **Key Concepts**: Basic terminology like algorithms, models, training data, features, labels, and learning objectives (supervised, unsupervised, reinforcement)

Module 02: Tools Used in Machine Learning

- **Introduction to Tools**: Overview of the tools such as Jupyter (including Google Colab and Amazon SageMaker Studio Lab), GitHub, and Visual Studio Code
- **Machine Learning Libraries and Frameworks**: Introduction to popular ML libraries such as Scikit-Learn, Pandas, NumPy, and Matplotlib
- **Introduction to Hugging Face**: Basics of Hugging Face and its importance in the ML ecosystem
- **Hands-on Lab**:
    - **Jupyter Notebook Basics**: Creating, formatting, and organizing notebooks
    - **GitHub Basics**: Version control, repositories, and collaboration
    - **Supplemental Formatting**: Using Markdown and LaTeX in Jupyter notebooks

Module 03: Types of Machine Learning and Successful ML Solutions

- **Types of Machine Learning**: Overview of supervised, unsupervised, and reinforcement learning
- **Examples of ML Applications**: Case studies of successful ML applications (e.g., recommendation systems, image classification, fraud detection)
- **Understanding Data Types**: Discuss different data types (numerical, categorical, text, and images) and their significance in ML

Module 04: ML Lifecycle, Overfitting, and Underfitting

- **Machine Learning Lifecycle**: Introduction to the end-to-end ML workflow—from data collection to model deployment
- **Overfitting vs. Underfitting**: Definition, causes, and impacts on model performance
- **Strategies to Combat Overfitting/Underfitting**: Cross-validation, regularization, and selecting the right model complexity

Module 05: AutoML and AutoGluon

- **Introduction to AutoML**: What is AutoML, its benefits, and when to use it
- **AutoGluon**: Overview and importance of AutoGluon in the ML workflow
- **Hands-on Lab**:
  - **Using AutoGluon**: Creating a model and generating predictions
  - **Batch Predictions**: How to generate and evaluate batch predictions using AutoGluon

Module 06: Data in Machine Learning & Exploratory Data Analysis (EDA)

- **How Data is Used in ML**: Understanding the role of data in training models, data quality, and preprocessing steps
- **Exploratory Data Analysis (EDA)**: Introduction to EDA, its importance, and techniques
- **Hands-on Lab**:
    - **EDA for Numerical Data**: Visualization, summary statistics, and identifying patterns
    - **EDA for Categorical Data**: Analyzing frequency distributions, relationships, and using bar plots

Module 07: Basic Feature Engineering

- **Introduction to Feature Engineering**: What is feature engineering, and why is it crucial?
- **Common Techniques**: Techniques like scaling, encoding, binning, and interaction terms
- **Hands-on Lab**: Practical exercises in applying feature engineering techniques to a dataset. In this lab, students learn common techniques that are used to transform numerical features, encode categorical features, and vectorize processed text features

Module 08: Tree-Based Models, Optimization, and Regularization

- **Tree-Based Models**: Overview of decision trees, random forests, and gradient boosting
- **Optimization Techniques**: Understanding loss functions, gradient descent, and optimization strategies
- **Regularization**: Techniques to prevent overfitting in models, such as L1, L2 regularization
- **Hands-on Lab**: Logistic regression using Scikit-Learn. (to decision trees, the Iterative Dichotomiser (ID3) algorithm, and information gain or impurity

Module 09: Hyperparameter Tuning

- **Importance of Hyperparameter Tuning**: What are hyperparameters, and why are they critical?
- **Techniques for Tuning**: Grid search, random search, and Bayesian optimization
- **Hands-on Lab**: Practical hyperparameter tuning on a simple model. Grid search and randomized search. Students will use a decision tree model to learn the basics of hyperparameter tuning that can be applied to any model

Module 10: Ensembling (Bagging, Boosting, Stacking)

- **Introduction to Ensemble Methods**: Explanation of ensemble techniques and why they improve model performance
- **Bagging, Boosting, and Stacking**: Detailed exploration of each method
- **Hands-on Lab**: Implementing ensembling techniques, with a focus on boosting (e.g., using XGBoost)

Module 11: Introduction to Fairness and Bias Mitigation in ML

- **Understanding Bias in ML**: What is bias in ML, and why is fairness critical?
- **Mitigation Strategies**: Introduction to basic techniques for identifying and reducing bias
- **Hands-on Lab**: Introduction to ResML (Responsible Machine Learning) frameworks

Module 12: Fair Model Formulation and Metrics

- **Fairness Metrics**: Overview of fairness metrics (e.g., demographic parity, equal opportunity)
- **Formulating Fair Models**: Strategies for ensuring fairness in model design
- **Hands-on Lab**: Exploring bias in data and applying fairness metrics. practice how to measure bias in data and apply various measures of bias quantification (including accuracy difference and difference in proportion of labels)

Module 13: Fairness Criteria

- **Deep Dive into Fairness Criteria**: Detailed exploration of criteria used to evaluate fairness
- **Hands-on Lab**: Practical application of fairness criteria on a given dataset. Implementing a DI Remover

Module 14: Bias Mitigation During Pre-processing, Training, and Post-processing

- **Bias Mitigation Techniques**: Overview of pre-processing (e.g., reweighting), during-training (e.g., adversarial debiasing), and post-processing methods
- **Hands-on Lab**: Implementation of bias mitigation techniques across the ML pipeline. model training, and postprocessing by using CI (normalized), DPL, fairness penalty terms, equalized odds, and ROC/calibration curves

Module 15: Course Wrap-Up

- **Summary and Reflection**: Recap of key concepts and skills learned throughout the course
- **Open Discussion**: Address any remaining questions and explore advanced topics or future learning paths

Module 16: Final Project/Final Exam

- Submission and presentation of final project
- Final Exam option

Teaching Methods and Strategies

- Total Course Duration: 96 contact hours; 16 weeks
- Weekly Contact Time: 6 hours
- Weekly Structure:
    - Lecture: 2-3 hours
    - Lab: 3-4 hours