

CERTIFIED ARTIFICIAL INTELLIGENCE (AI) PRACTITIONER CAIP

Exam: AIP-110

Duration: 5 days; Instructor-led | Virtual Instructor-led

WHAT WILL YOU LEARN

Artificial intelligence (AI) and machine learning (ML) have become an essential part of the toolset for many organizations. When used effectively, these tools provide actionable insights that drive critical decisions and enable organizations to create exciting, new, and innovative products and services. This course shows you how to apply various approaches and algorithms to solve business problems through AI and ML, follow a methodical workflow to develop sound solutions, use open source, off-the-shelf tools to develop, test, and deploy those solutions, and ensure that they protect the privacy of users.

OBJECTIVES

In this course, you will implement AI techniques in order to solve business problems. You will:

- Specify a general approach to solve a given business problem that uses applied AI and ML.
- Collect and refine a dataset to prepare it for training and testing.
- Train and tune a machine learning model.
- Finalize a machine learning model and present the results to the appropriate audience.
- Build linear regression models.
- Build classification models.
- Build clustering models.
- Build decision trees and random forests.
- Build support-vector machines (SVMs).
- Build artificial neural networks (ANNs).
- Promote data privacy and ethical practices within AI and ML projects.

PREREQUISITES

To ensure your success in this course, you should have at least a high-level understanding of fundamental AI concepts, including, but not limited to: machine learning, supervised learning, unsupervised learning, artificial neural networks, computer vision, and natural language processing. You can obtain this level of knowledge by taking the CertNexus AIBIZ™ (Exam AIZ-110) course.

You should also have experience working with databases and a high-level programming language such as Python, Java, or C/C++. You can obtain this level of skills and knowledge by taking the following Logical Operations or comparable course:

- Database Design: A Modern Approach
- Python® Programming: Introduction

- Python® Programming: Advanced

AUDIENCE

The skills covered in this course converge on three areas—software development, applied math and statistics, and business analysis. Target students for this course may be strong in one or two or these of these areas and looking to round out their skills in the other areas so they can apply artificial intelligence (AI) systems, particularly machine learning models, to business problems.

So the target student may be a programmer looking to develop additional skills to apply machine learning algorithms to business problems, or a data analyst who already has strong skills in applying math and statistics to business problems, but is looking to develop technology skills related to machine learning.

A typical student in this course should have several years of experience with computing technology, including some aptitude in computer programming.

This course is also designed to assist students in preparing for the CertNexus® Certified Artificial Intelligence (AI) Practitioner (Exam AIP-110) certification.

COURSE CONTENTS

Module 1: Solving Business Problems Using AI and ML

Topic A: Identify AI and ML Solutions for Business Problems

- The Data Hierarchy—Making Data Useful
- Big Data
- Guidelines for Working with Big Data
- Data Mining
- Examples of Applied AI and ML in Business
- Guidelines to Select Appropriate Business Applications for AI and ML
- Identifying Appropriate Business Applications for AI and ML (ACTIVITY)

Topic B: Follow a Machine Learning Workflow

- Machine Learning Model
- Machine Learning Workflow
- Data Science Skillset
- Traditional IT Skillsets
- Concept Drift
- Transfer Learning
- Guidelines for Following the Machine Learning Workflow

- Planning the Machine Learning Workflow (*ACTIVITY*)

Topic C: Formulate a Machine Learning Problem

- Problem Formulation
- Framing a Machine Learning Problem
- Differences Between Traditional Programming and Machine Learning
- Differences Between Supervised and Unsupervised Learning
- Randomness in Machine Learning
- Uncertainty
- Random Number Generation
- Machine Learning Outcomes
- Guidelines for Formulating a Machine Learning Outcome
- Selecting a Machine Learning Outcome (*ACTIVITY*)

Topic D: Select Appropriate Tools

- Open Source AI Tools
- Proprietary AI Tools
- New Tools and Technologies
- Hardware Requirements
- GPUs vs. CPUs
- GPU Platforms
- Cloud Platforms
- Guidelines for Configuring a Machine Learning Toolset
- How to Install Anaconda
- Selecting a Machine Learning Toolset (*ACTIVITY*)

Module 2: Collecting and Refining the Dataset

Topic A: Collect the Dataset

- Machine Learning Datasets
- Structure of Data
- Terms Describing Portions of Data
- Data Quality Issues
- Data Sources
- Open Datasets
- Guidelines for Selecting a Machine Learning Dataset
- Examining the Structure of a Machine Learning Dataset (*ACTIVITY*)
- Extract, Transform, and Load (ETL)
- Machine Learning Pipeline
- ML Software Environments
- Guidelines for Loading a Dataset
- Loading the Dataset (*ACTIVITY*)

Topic B: Analyze the Dataset to Gain Insights

- Dataset Structure
- Guidelines for Exploring the Structure of a Dataset
- Exploring the General Structure of the Dataset (*ACTIVITY*)
- Normal Distribution
- Non-Normal Distributions
- Descriptive Statistical Analysis
- Central Tendency
- When to Use Different Measures of Central Tendency
- Variability
- Range Measures
- Variance and Standard Deviation
- Calculation of Variance

- Variance in a Sample Set
- Calculation of Standard Deviation
- Skewness
- Calculation of Skewness Measures
- Kurtosis
- Calculation of Kurtosis
- Statistical Moments
- Correlation Coefficient
- Calculation of Pearson's Correlation Coefficient
- Guidelines for Analyzing a Dataset
- Analyzing a Dataset Using Statistical Measures (*ACTIVITY*)

Topic C: Use Visualizations to Analyze Data

- Visualizations
- Histogram
- Box Plot
- Scatterplot
- Geographical Maps
- Heat Maps
- Guidelines for Using Visualizations to Analyze Data
- Analyzing a Dataset Using Visualizations (*ACTIVITY*)

Topic D: Prepare Data

- Data Preparation
- Data Types
- Operations You Can Perform on Different Types of Data
- Continuous vs. Discrete Variables
- Data Encoding
- Dimensionality Reduction
- Impute Missing Values
- Duplicates
- Normalization and Standardization
- Summarization
- Holdout Method
- Guidelines for Preparing Training and Testing Data
- Splitting the Training and Testing Datasets and Labels (*ACTIVITY*)

Module 3: Setting Up and Training a Model

Topic A: Set Up a Machine Learning Model

- Design of Experiments
- Hypothesis
- Hypothesis Testing
- Hypothesis Testing Methods
- p-value
- Confidence Interval
- Machine Learning Algorithms
- Algorithm Selection
- Guidelines for Setting Up a Machine Learning Model
- Setting Up a Machine Learning Model (*ACTIVITY*)

Topic B: Train the Model

- Iterative Tuning
- Bias
- Compromises
- Model Generalization
- Cross-Validation

- k-Fold Cross-Validation
- Leave-p-Out Cross-Validation
- Dealing with Outliers (*ACTIVITY*)
- Feature Transformation
- Transformation Functions
- Scaling and Normalizing Features
- The Bias–Variance Tradeoff
- Parameters
- Regularization
- Models in Combination
- Processing Efficiency
- Guidelines for Training and Tuning the Model
- Refitting and Testing the Model (*ACTIVITY*)

Module 4: Finalizing a Model

Topic A: Translate Results into Business Actions

- Know Your Audience
- Visualization for Presentation
- Guidelines for Presenting Your Findings
- Translating Results into Business Actions (*ACTIVITY*)

Topic B: Incorporate a Model into a Long-Term Business Solution

- Put a Model into Production
- Production Algorithms
- Pipeline Automation
- Testing and Maintenance
- Consumer-Oriented Applications
- Guidelines for Incorporating Machine Learning into a Long-Term Solution
- Incorporating a Model into a Long-Term Solution (*ACTIVITY*)

Module 5: Building Linear Regression Models

Topic A: Build a Regression Model Using Linear Algebra

- Linear Regression
- Linear Equation
- Linear Equation Data Example
- Straight Line Fit to Example Data
- Linear Equation Shortcomings
- Linear Regression in Machine Learning
- Linear Regression in Machine Learning Example
- Matrices in Linear Regression
- Normal Equation
- Linear Model with Higher Order Fits
- Linear Model with Multiple Parameters
- Cost Function
- Mean Squared Error (MSE)
- Mean Absolute Error (MAE)
- Coefficient of Determination
- Normal Equation Shortcomings
- Guidelines for Building a Regression Model Using Linear Algebra
- Building a Regression Model Using Linear Algebra (*ACTIVITY*)

Topic B: Build a Regularized Regression Model Using Linear Algebra

- Regularization Techniques
- Ridge Regression

- Lasso Regression
- Elastic Net Regression
- Guidelines for Building a Regularized Linear Regression Model
- Building a Regularized Linear Regression Model (*ACTIVITY*)

Topic C: Build an Iterative Linear Regression Model

- Iterative Models
- Gradient Descent
- Global Minimum vs. Local Minima
- Learning Rate
- Gradient Descent Techniques
- Guidelines for Building an Iterative Linear Regression Model
- Building an Iterative Linear Regression Model (*ACTIVITY*)

Module 6: Building Classification Models

Topic A: Train Binary Classification Models

- Linear Regression Shortcomings
- Logistic Regression
- Decision Boundary
- Cost Function for Logistic Regression
- A Simpler Alternative for Classification
- k-Nearest Neighbor (k-NN)
- k Determination
- Logistic Regression vs. k-NN
- Guidelines for Training Binary Classification Models
- Training Binary Classification Model (*ACTIVITY*)

Topic B: Train Multi-Class Classification Models

- Multi-Label Classification
- Multi-Class Classification
- Multinomial Logistic Regression
- Guidelines for Training Multi-Class Classification Models
- Training a Multi-Class Classification Model (*ACTIVITY*)

Topic C: Evaluate Classification Models

- Model Performance
- Confusion Matrix
- Classifier Performance Measurement
- Accuracy
- Precision
- Recall
- Precision–Recall Tradeoff
- F1 Score
- Receiver Operating Characteristic (ROC) Curve
- Thresholds
- Area Under Curve (AUC)
- Precision–Recall Curve (PRC)
- Guidelines for Evaluating Classification Models
- Evaluating a Classification Model (*ACTIVITY*)

Topic D: Tune Classification Models

- Hyperparameter Optimization
- Grid Search
- Randomized Search
- Bayesian Optimization
- Genetic Algorithms
- Guidelines for Tuning Classification Models

- Tuning a Classification Model (*ACTIVITY*)

Module 7: Building Clustering Models

Topic A: Build k-Means Clustering Models

- k-Means Clustering
- Global vs. Local Optimization
- k Determination
- Elbow Point
- Cluster Sum of Squares
- Silhouette Analysis
- Additional Cluster Analysis Methods
- Guidelines for Building a k-Means Clustering Model
- Building a k-Means Clustering Model (*ACTIVITY*)

Topic B: Build Hierarchical Clustering Models

- k-Means Clustering Shortcomings
- Hierarchical Clustering
- Hierarchical Clustering Applied to a Spiral Dataset
- When to Stop Hierarchical Clustering
- Dendrogram
- Guidelines for Building a Hierarchical Clustering Model
- Building a Hierarchical Clustering Model (*ACTIVITY*)

Module 8: Building Advanced Models

Topic A: Build Decision Tree Models

- Decision Tree
- Classification and Regression Tree (CART)
- Gini Index Example
- CART Hyperparameters
- Pruning
- C4.5
- Continuous Variable Discretization
- Bin Determination
- One-Hot Encoding
- Decision Tree Algorithm Comparison
- Decision Trees Compared to Other Algorithms
- Guidelines for Building a Decision Tree Model
- Building a Decision Tree Model (*ACTIVITY*)

Topic B: Build Random Forest Models

- Ensemble Learning
- Random Forest
- Out-of-Bag Error
- Random Forest Hyperparameters
- Feature Selection Benefits
- Guidelines for Building a Random Forest Model
- Building a Random Forest Model (*ACTIVITY*)

Module 9: Building Support-Vector Machines

Topic A: Build SVM Models for Classification

- Support-Vector Machines (SVMs)
- SVMs for Linear Classification
- Hard-Margin Classification
- Soft-Margin Classification
- SVMs for Non-Linear Classification
- Kernel Trick

- Kernel Trick Example
- Kernel Methods
- Guidelines for Building an SVM Model
- Building an SVM Model (*ACTIVITY*)

Topic B: Build SVM Models for Regression

- SVMs for Regression
- Guidelines for Building SVM Models for Regression
- Building an SVM Model for Regression (*ACTIVITY*)

Module 10: Building Artificial Neural Networks

Topic A: Build Multi-Layer Perceptrons (MLP)

- Artificial Neural Network (ANN)
- Perceptron
- Multi-Label Classification Perceptron
- Perceptron Training
- Perceptron Shortcomings
- Multi-Layer Perceptron (MLP)
- ANN Layers
- Backpropagation
- Activation Functions
- Guidelines for Building MLPs
- Building an MLP (*ACTIVITY*)

Topic B: Build Convolutional Neural Networks (CNN)

- Traditional ANN Shortcomings
- Convolutional Neural Network (CNN)
- CNN Filters
- CNN Filter Example
- Padding
- Stride
- Pooling Layer
- CNN Architecture
- Generative Adversarial Network (GAN)
- GAN Architecture
- Guidelines for Building CNNs
- Building a CNN (*ACTIVITY*)

Module 11: Promoting Data Privacy and Ethical Practices

Topic A: Protect Data Privacy

- Protected Data
- Obligation to Protect PII
- Relevant Data Privacy Laws
- Privacy by Design
- Data Privacy Principles at Odds with Machine Learning
- Guidelines for Complying with Data Privacy Laws and Standards
- Complying with Applicable Laws and Standards
- Open Source Data Sharing and Privacy
- Data Anonymization
- Guidelines for Data Anonymization
- The Big Data Challenge
- Guidelines for Protecting Data Privacy
- Protecting Data Privacy (*ACTIVITY*)

Topic B: Promote Ethical Practices

- Preconceived Notions
- The Black Box Challenge
- Prejudice Bias
- Proxies for Larger Social Discriminations
- Ethics in NLP
- Guidelines for Promoting Ethical Practices
- Promoting Ethical Practices (*ACTIVITY*)

Topic C: Establish Data Privacy and Ethics Policies

- Privacy and Data Governance for AI and ML
- Intellectual Property
- Humanitarian Principles
- Guidelines for Establishing Policies Covering Data Privacy and Ethics
- Establishing Policies Covering Data Privacy and Ethics (*ACTIVITY*)

**APPENDIX A: MAPPING COURSE CONTENT TO
CERTNEXUS® CERTIFIED ARTIFICIAL INTELLIGENCE (AI)
PRACTITIONER (EXAM AIP-110)**

Course-specific Technical Requirements

Hardware

For this course, you will need one computer for each student and one for the instructor. Each computer will need the following minimum hardware configurations:

- 2 gigahertz (GHz) 64-bit (x64) processor that supports the VT-x or AMD-V virtualization instruction set and Second Level Address Translation (SLAT).
- 8 gigabytes (GB) of Random Access Memory (RAM).
- 32 GB available storage space.
- Monitor capable of a screen resolution of at least 1,024 × 768 pixels, at least a 256-color display, and a video adapter with at least 4 MB of memory.
- Bootable DVD-ROM or USB drive.
- Keyboard and mouse or a compatible pointing device.
- Fast Ethernet (100 Mb/s) adapter or faster and cabling to connect to the classroom network.
- IP addresses that do not conflict with other portions of your network.
- Internet access (contact your local network administrator).
- (Instructor computer only) A display system to project the instructor's computer screen.

Software

- Microsoft Windows 10 64-bit.
- Oracle® VM VirtualBox version 6.0.10 (VirtualBox-6.0.10-132072-Win.exe).
- VirtualBox is distributed with the course data files under version 2 of the GNU General Public License (GPL).
- If necessary, software for viewing the course slides. (Instructor machine only.)

NOTE:

- While it is possible to run VirtualBox on other operating systems, this course was written and tested using Windows

10. If your classroom computers will use a different operating system, it is highly recommended that you install and test VirtualBox and the course VM on the computers to make sure you can key through the course successfully before delivering a class.

- The Linux operating system is already installed on the VM that will be loaded in VirtualBox. Specifically, this VM runs the Debian 10 ("Buster") distribution.
- The system on the VM is configured to log the user in automatically. If you or your students are prompted at any time to log in, the account is named student and the password is Pa22w0rd.

Datasets

This course uses several third-party datasets to demonstrate machine learning concepts. Some of these datasets come packaged with the scikit-learn and Keras libraries:

- Boston house prices dataset
- Iris plants dataset
- Fashion-MNIST database of fashion articles
- IMDB movie reviews sentiment classification

In addition, several datasets were obtained from other sources. These are listed along with the relevant license information or citation:

- House Sales in King County, USA
 - Public domain. Retrieved from <https://www.kaggle.com/harlfoxem/housesalesprediction>.
- Combined Cycle Power Plant Data Set
 - Tüfekci, P. (2014, September). Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. International Journal of Electrical Power &

Energy Systems, 60, 126-140. doi:
10.1016/j.ijepes.2014.02.027.

- Kaya, H., Tüfekci, P., & Gürgen, S. F. (2012, March). Local and Global Learning Methods for Predicting Power of a Combined Gas & Steam Turbine. Proceedings of the International Conference on Emerging Trends in Computer and Electronics Engineering (ICETCEE), 13-18.
- Titanic: Machine Learning from Disaster
 - Public domain. Retrieved from <https://www.kaggle.com/c/titanic>.
- Wine Data Set
 - Dua, D., & Graff, C. (2019, November). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.
- Occupancy Detection Data Set
 - Candanedo, L. M., & Feldheim, V. (2016, January). Accurate occupancy detection of an office room from light, temperature, humidity, and CO2 measurements using statistical learning models. Energy and Buildings, 112(15), 28-39. doi: 10.1016/j.enbuild.2015.11.071.