PROGRAM OVERVIEW

The online Data Analytics Certificate teaches you how to find valuable insights from quantitative and qualitative data collected on a massive scale. You will learn machine learning techniques to solve a broad range of data science problems by interpreting the results to make predictions and establish the reliability of those predictions. All while developing a data analytics portfolio to showcase and highlight your skills.

LEARNING OUTCOMES

- Apply Python, R, R Studio, Spark, and Amazon Web Services to solve data science problems
- Translate business objectives into analytical opportunities using data mining
- Become broadly competent in the use and evaluation of the statistical machine learning techniques classification, regression and association
- Construct, justify, and apply custom data science processes
- Apply parametric and nonparametric machine learning models and avoid overfitting models

DURATION

- Full time: 20 weeks
- Part time: 40 weeks

PROGRAM FEATURES

- 100% project-based program that takes a learn-by-doing approach
- The program is a series of five online instructor-led courses
- Students are supported by an expert mentor and work with a small team of fellow students to move through the program
- Career coaching designed to help you land a job in data analytics, also includes mock interviews, resume assistance and portfolio development

WHO SHOULD ENROLL

- Those with career goals of becoming a data analyst or data scientist
- Professionals looking to gain extensive skills in data analytics

A minimum of two years’ experience and a basic familiarity with Microsoft Office or an equivalent productivity suite is required.
In this course, you'll learn how to use statistical machine learning techniques to understand the relationship between customer demographics and purchasing behavior, and then develop a model for predicting the future sales volume of products. You will learn how to apply the Python programming language to solve data analytics and machine learning problems.

**Data Analytics Tasks:**

- Analyze differences in customer behavior
- Uncover relationships in purchasing behavior

**Course Topics:**

- Identifying types of business problems for which data analysis can provide significant insights in support of business decision-making
- Translating business objectives into analytical opportunities using data mining
- Applying statistical machine learning tools and methods to different kinds of data
- Preprocessing data (e.g., feature engineering, addressing missing data, transforming numeric values to nominal values, discretizing data)
- Communicating data mining results to management and other nontechnical audiences
- Using data mining tools to investigate patterns in complex data sets
- Using decision tree classifiers to investigate classification problems
- Installing and Configuring Security Appliances
- Installing and Configuring Wireless and Physical Access Security
- Applying cross-validation methods
- Interpreting and drawing inferences from the results of data mining
- Applying Python to data science problems

**Tools used:** Python, Jupyter Notebook, Matplotlib, Pandas, Scikit-learn
In this course, you will continue to use the Python programming language to perform machine learning and data science tasks. More specifically, you will use Pandas, NumPy, Matplotlib and Sci-kit learn to develop machine learning and predictive analytics solutions.

Data Analytics Tasks:

- Develop predictive models for appropriate customer credit limits

Course Topics:

- Using the Sci-kit learn machine learning library for Python
- Identifying and solving collinearity through feature engineering and feature selection
- Constructing, justifying, and applying custom data science processes
- Drawing relationships between learner performance and measured features to help understand model performance
- Conducting feature selection to investigate the correlation between different features in a dataset
- Defining the business purpose of a data analytics project and making a principled, realistic analysis plan
- Using decision tree classifiers to investigate classification problems
- Assessing the predictive performance of classifiers by examining key error metrics
- Identifying where learning methods fail and gaining insight into why with error analysis
- Selecting and justifying appropriate types of data analysis and statistical procedures
- Using data mining tools and different classifiers (e.g., k-nearest neighbor, decision trees, support vector machines) to develop predictive models

Tools used: Python, Jupyter Notebook, Matplotlib, Pandas, Scikit-learn, SQL
In this course, you will learn to use statistical machine learning techniques to predict brand preferences based on customer characteristics, and then you will develop a model to recommend new products based on customer purchases. You will learn and use the R statistical programming language and the R Studio analytics environment.

Data Analytics Tasks:

- Develop predictive models for customer brand preferences
- Build a recommender system using association rules

Course Topics:

- Presenting nontechnical data mining results to management
- Becoming broadly competent in the use and evaluation of the statistical machine learning techniques classification, regression and association
- Applying parametric and non-parametric machine learning models
- Avoiding overfitting models
- Optimizing classifiers by adjusting and testing classifier parameters
- Applying data analytics in e-commerce (e.g., understanding customer behavior, segmenting customers by key demographics, selecting new products strategically and predicting their profitability).
- Managing data in R with Tidyverse
- Comparing and selecting different predictive models
- Applying predictive models to test sets
- Exploring data using visualization techniques and descriptive statistics in R using ggplot2 and plot.ly

Tools used: R, R Studio and machine learning packages, SQL
In this course, you will continue to use the R statistical programming language and a variety of add-on "packages" to visualize data relationships and to implement classification and regression models for emerging engineering applications, such as understanding behavior in the “Internet of Things.”

Data Analytics Tasks:

- Model smart energy usage
- Develop an indoor locationing solution

Course Topics:

- Presenting highly technical, engineering-oriented data mining results to business management
- Applying dimensionality reduction methods to broad datasets to reduce their complexity prior to modeling
- Using the caret package for R
- Choosing and evaluating classification modeling techniques in R
- Choosing and evaluating regression modeling techniques in R
- Analyzing time series data through using data mining tools and different classifiers (e.g., k-nearest neighbor, decision trees, support vector machines) to develop predictive models
- Interpreting performance metrics
- Performing error analysis
- Applying cross-validation and train-test split methods

Tools used: R packages for data visualization, SQL, time series analysis and machine learning
BIG DATA ANALYTICS: TEXT MINING AND SENTIMENT ANALYSIS

In this course, you will learn how to mine and analyze extremely large datasets to provide insight into real-world business problems. You will conduct automated keyword extraction using natural language processing and use cloud-based computing to interpret the results and to make and communicate predictions of vital interest to business stakeholders.

Data Analytics Tasks:

- Mine textual content from a large public corpus to conduct a web sentiment analysis of users’ attitudes towards various mobile phone platforms
- Construct a sentiment analysis predictive model using natural language processing and machine learning

Course Topics:

- Acquiring, processing and analyzing extremely large data sets using cloud-based data mining methods to discover patterns or do data exploration
- Operating the Amazon AWS cloud computing platform for data analysis
- Discovering and mounting very large public datasets on the AWS platform
- Setting up and running Elastic Map Reduce (EMR) instances for data analysis
- Defining and creating a Spark cluster on Azure
- Exploring and visualizing data using Databricks and PySpark
- Performing sentiment analysis with an Azure Spark cluster using NLP
- Doing data science in the Cloud
- Assessing the predictive performance of classifiers by examining key error metrics
- Using the Matplotlib and Seaborn Python 2D plotting libraries

Tools used: Amazon Web Services, Elastic Map Reduce, Microsoft Azure, Databricks, Spark
DATA SCIENCE CAPSTONE

In the Data Science Capstone, you will propose and design your own data science solution for a problem of your choosing. The capstone is an assessment of your ability to independently utilize the tools and methods needed to be a successful data scientist. This project will begin at the end of course two.

TECHNICAL REQUIREMENTS

- PC, Mac or Linux desktop or laptop computer
- 8 GB of Ram, minimum, 16 GB of Ram recommended
- Intel or compatible processor
- 500 GB hard drive (minimum)
- PC, Mac or Linux desktop or laptop computer
- Operating System: (Windows, Mac OS or Linux; most recent OS versions recommended)
- High-speed internet connection
- Current version of Microsoft Office or Microsoft Office for Mac (Open Office or Libre Office may also be used)
- Webcam for video conferencing

JOB OUTLOOK

364,000
DATA PROFESSIONAL JOBS OPEN IN THE U.S.
FORBES

PAY SCALE

$75,205
SALARY AVERAGE FOR DATA ANALYST
INDEED.COM

$96,208
SALARY AVERAGE FOR DATA SCIENTIST
PAYSCALE.COM

JOB TITLES

DATA ANALYST
DATA SCIENTIST
BUSINESS INTELLIGENCE
PROJECT MANAGER
DATA ANALYTICS CONSULTANT

85%
COMPANIES WILL ADOPT BIG DATA TECH BY 2022
WORLD ECONOMIC FORUM

noncredit@purdue.edu
https://online.purdue.edu