Syllabus – Data Analytics and Big Data Certificate Program

Applications and Benefits:

Nearly all businesses collect data about their operations and examine this data for insight into how to improve their operations. As the amount of data that businesses collect becomes increasingly large, insights from the data can no longer be effectively derived manually. There is a growing trend among companies, organizations, and individuals to exploit data analytics’ potential to help them discover and act on the most important patterns contained within the data they collect. Data analytics has a myriad of business applications and is used increasingly to drive decisions about all aspects of business including spotting sales trends, developing smarter marketing campaigns, accurately predicting customer loyalty, and predicting and protecting against fraud. In fact, data analytics can be applied anywhere in a business or organization where a company is interested in identifying and exploiting predictable outcomes. The skills you will learn are applicable to a wide variety of data analytics and big data projects, and will enable you to start working with business problems that require “supervised learning.”

How you will work:

The Data Analytics and Big Data Certificate Program is delivered 100% online, but you will not work alone. You will work (synchronously) with a small team of fellow students supported by an expert mentor on a series of nine projects distributed across five courses and a capstone project. You and your teammates will collaborate on the conceptual and design aspects of each project, but then you must do the hands-on implementation individually. The program employs a mastery-based approach, which means that you will continue to work on a project until your mentor is convinced that you have mastered the underlying knowledge and skills.

Course 1: Understanding Customers

Overview:

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.
Details:

You will be working as a member of the fictional Blackwell Data Analytics Team. Blackwell Electronics has been a successful consumer electronics retailer in the southeastern United States for over 40 years. Your job is to use data mining and machine-learning techniques to investigate the patterns in customer sales data and provide insight into customer buying trends and preferences. You will then use machine learning methods to predict which potential new products will be the most profitable for the company to add to its sales mix. The inferences you draw from the patterns in the data will help businesses make data-driven decisions about sales and marketing activities. Finally, you will present to management, explaining your insights and suggestions for data mining process improvements.

Course 2: Data Science with Python

Overview:

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.

Details:

You will be working as a Data Scientist for Credit One, a third-party credit issuing authority that provides customer credit approval services to businesses that sell to customers. Credit One has tasked you with examining current customer demographics to better understand what traits might relate to whether or not a customer might or might not default on their current credit offerings. Your job as a Data Scientist will be to identify which customer features are statistically significant to customer default rates and build predictive models that Customer One can use to better classify these ‘at-risk’ customers better than previously implemented models they have used. You will use machine learning classification and regression methods to identify customer demographic characteristics and account attributes that possibly contribute to credit defaults. Finally, you will present to management, explaining your insights and suggestions for data mining process improvements.
Course 3: Predicting Profitability and Customer Preferences

Overview:

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 for work in course 3.

Details:

You will work with Thomas Bayes, the Chief Technology Officer at SuperMart. They have recently begun to leverage the data collected from online and in-store transactions to gain insight into their customers’ purchasing behavior. Your job is to extend their application of data analytics and data science methods to develop models to predict which brands of computer products SuperMart customers prefer based on customer demographics collected from a marketing survey. You will also develop an e-commerce product recommendation system based on the historical sales of products that will help SuperMart identify cross-selling opportunities that can be used to increase the profitability of their online store. Finally, you will present to management, explaining your insights and suggestions for data mining process improvements.

Course 4: Deep Analytics and Visualization

Overview:

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.”

Details:

In this course, you’ll be working for an "Internet of Things" analytics firm to use Data Analytics to solve two difficult problems in the physical world:

- Smart energy usage: Modeling patterns of energy usage by time of day and day of the year in a typical residence whose electrical system is monitored by multiple sub-meters.
- Indoor locationing: Determining a person’s physical position in a multi-building indoor space using Wi-Fi fingerprinting (the pattern of radio signals from multiple WiFi hotspots that are received by a smart phone at a given location).
You will create visualizations, and then you will generate descriptive statistics and predictive models using both statistical classifiers and linear regression techniques. Finally, you’ll present the results to the company’s management, explaining the strengths and weaknesses of the approaches you implemented and making suggestions for further improvement.

Course 5: Big Data Analytics

Overview:

In this course, you will learn how to mine and analyze extremely large data sets to provide insight to real-world business problems. You will conduct automated keyword extraction using Natural Language Processing and cloud-based computing to interpret the results to make and communicate predictions of vital interest to business stakeholders.

Details:

You will be working as a data analyst for Alert Analytics, a data analytics consulting firm. You will be asked to take over for a recently-transferred analyst who has been working on a big data project for Helio, a customer service consulting firm. Helio is working with a large online retailer to create an automated process of gauging customer sentiment about many of the products sold by the retailer. This will provide a better understanding of which products their customers prefer and some of the possible reasons why their customers may not prefer others. After examining some of the customer reviews, which have been collected from their own website, the retailer has provided Helio with a specific list of products to be examined.

To perform this analysis, Helio has engaged Alert Analytics to conduct an automated keyword extraction using Apache Spark and Amazon Web Service’s elastic map-reduce to gain insight into users’ attitudes towards the products. In the course of creating this process, you’ll also learn to identify and remove commonly used words found in the customer reviews and assign values to the words that are more relevant to customer sentiment. Your job is to conduct this analysis and assemble a report with the specific percentages of positive and negative customer sentiment surrounding each previously identified product.

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.
Applications and Benefits:

This is your opportunity to propose your own data science project that will be the highlight of your work in the program and the signature piece of work in your professional portfolio. You are free to utilize any technologies, methods, and skills you have learned in the program and can either gather your own data or use freely available data from the web.