UCI Extension

Data Science Certificate Program

Accelerate Your Career

extension.uci.edu/datascience

Offered in partnership with

Predictive
Analytics
WORLD

University of California, Irvine

Improve Your Career Options with a Professional Certificate

University of California, Irvine Extension's professional certificate and specialized studies programs help you increase or enhance your current skills or prepare for a new career. Courses are highly practical and instructors are qualified leaders in their field. Convenient online courses make it easy to learn on your own time, in your own way. A certificate bearing the UC seal signifies a well-known, uncompromising standard of excellence.

Data Science Certificate Program

Corporations have dramatically increased investments in their "digital enterprise" in the past few years. It has been estimated that by 2020, IT departments will be monitoring 50 times more data than they are today. This tidal wave of data is driving unprecedented demand for those with the skills required to manage and leverage these very large data sets into a competitive advantage.

This program has been designed to help meet the expanding needs for these data scientists who are skilled in the utilization of a unique blend of science, art and business. These professionals are skilled in automating methods of collecting and analyzing data and utilizing inquisitive exploring techniques to discover previously hidden insight from this data that can profoundly impact the success of any business.

Not easy to obtain through a single more traditional undergraduate or graduate degree, this program provides the skills needed to become a courageous data explorer with a "hacker mentality" utilizing concepts in on- and off-cloud, scalable, data engineering (inspecting, cleaning, transforming, and modeling data), unstructured data and NoSQL, computational statistics, pattern recognition, data mining /predictive analytics, machine learning, data visualization, and high performance software and hardware to tackle the demands of the resulting "big data" sets.

Data science has become the broader term encompassing data engineering, data analysis, business intelligence and advanced computer science. This program provides the skills required to become a data scientist and provides existing data analysts opportunities to broaden their skills while leveraging their unique domain expertise.

In a recent article, USA Today noted that the talent pool for these data scientists "… is at best 20% of the demand." This has resulted in incredibly strong demand, salaries and job security for those with these skills. This program provides individuals the skills needed to effectively collect and manage big data and perform data-driven discovery and prediction, extracting value and competitive intelligence for their organizations.

Curriculum

Who Should Enroll

This program is intended for professionals in a variety of industries and job functions who are looking to help their organization understand and leverage the massive amounts of diverse data they collect. Others who would benefit from this program include: data engineers, data analysts, computer scientists, business analysts, database administrators, researchers, and statisticians.

Certificate Eligibility and Requirements

A certificate is awarded upon completion of 15 credit units (6 required and 9 elective credit units) with a grade point average of "C" or better.

Students interested in pursuing both the Data Science and Predictive Analytics programs may overlap two courses to satisfy certificate requirements into both programs.

To become an official candidate in the program, students pursuing the certificate must submit a Declaration of Candidacy. To receive the certificate after completing all program requirements, students must submit a Request for Certificate. All requirements must be completed within 5 years after the student enrolls in his/her first course. Students not pursuing the certificate program are welcome to take as many individual courses as they wish.

Program Benefits

- Learn from industry experts how to utilize a combination of science, art & business techniques to deliver new insights and competitive intelligence
- Describe the phases of the analytics lifecycle
- Utilize a variety of statistical and computer science tools and techniques to analyze data
- Describe and use the typical tools and technologies required to model and analyze large (big) datasets
- Explain the use of typical tools to explore data (R, STATISTICA, Hadoop, etc.)
- Utilize an inquisitive "hacker" mentality to uncover new meaning from existing data
- Effectively design, model and manage databases
- Describe and utilize unstructured and structured data sets leveraging text analytics tools.
- Define requirements, develop an architecture and implement a data warehouse plan

Program Fees

The total cost of the program varies depending on the electives chosen. Actual fees may differ from the estimate below. Fees are subject to change without prior notice.

Course Fees	$3,500
Candidacy Fee	$125
Textbooks	$145
Total Estimated Cost	**$3,770**

Required Courses (6 units)

Introduction to Data Science
I&C SCI X426.60 (3 units)
Learn the data design, management and manipulation tools and processes commonly used by data scientists and understand how to apply this knowledge to yield measurable business value. Gain an overview of the basic techniques of data science, including data analysis, statistical modeling, data engineering, relational databases, SQL and NoSQL, manipulation of data at scale (big data), algorithms for data mining, data quality, remediation and consistency operations.

Data Exploration, Analytics and Visualization
I&C SCI X426.61 (3 units)
Once data has been effectively collected, structured and cleaned it's time to explore and uncover applicable meaning from the data. Learn the basics of how to use processes and tools from machine learning, data mining and Predictive analytics to begin this journey. Understand how to collect and gain value from the vast amount of often untapped un-unstructured (text) data. Key industry processes for understanding and utilizing data and developing predictive and related models will be covered. An overview of related topics including Graph Analytics (PageRank, community detection, recursive queries, iterative processing) and data Provenance, privacy, ethics, governance will also be covered Students will also learn techniques for visualizing data including multivariate, temporal, text-based, hierarchical, and network/graph-based data.

For more information:
Jackie Badwah
jdbadwah@uci.edu
(949) 824-3413

Elective Courses (Minimum 9 units)

Designing Relational Databases
I&C SCI X426.81 (3 units)
If you need to learn more about the design of relational database systems, this course is for you. You'll explore how the relational database approach is used in both open and closed systems and on both mainframe and client/server platforms, and learn about the design of systems of all sizes including standalone, workgroup, departmental, and enterprise-wide applications. You'll also learn how to incorporate data from legacy systems as well as how to develop entirely new systems. You'll have the opportunity to design a rational database model in class and refine it for a real-world system or case study. Topics include an overview of database systems, defining business entities, entity relationship modeling (ERM) for top-down analysis, defining relational database tables and attributes, conversion of logical design to physical design, normalization of tables, and using structured query language (SQL) to process data and generate reports.

Data Modeling
I&C SCI X425.32 (2.5 units)
In this course you will learn how to develop data models which are used to publish the information that organizations rely on for their day-to-day decision making processes. Basic data modeling techniques will be explored including capturing data requirements, analyzing data elements, identifying entities and attributes, and determining relationships between entities. Data models will be developed using the entity/relationship graphical model. Once a basic data model has been developed, students will be instructed on how to enhance the data model to manage and enhance the value of the data it contains. This will directly affect the quality of the information produced from the data, and provide an organization with the knowledge necessary to optimally perform its business functions. The emphasis of this course is on analysis and design rather than on the actual implementation of a database.

Data Management
I&C SCI X425.35 (2.5 units)
Data may be an organization's most underrated asset. Organizations often invest a very small percentage of their budget on maintaining the health of their data, which may over time lead to decreasing competitiveness, loss of market share, and even violations of regulatory requirements. This course presents data management from an organizational perspective with the goal of using data management methodologies as a foundation for enterprise data architecture. The course concentrates on data definition and specification, data quality, data organization, data integration, and data equilibrium.

Introduction to Predictive Analytics
I&C SCI X425.61 (2 units)
Gain a fundamental understanding of the art and science of predictive analytics as it relates to improving business performance. This hands-on course will cover the key concepts necessary to extract stored data elements, understand what they mean from a business point of view, transform their formats, and derive new relationships among them to produce a dataset suitable for analytical modeling. At the end of the course, participants will be able to use these skills to produce a fully processed data set compatible for building powerful predictive models that can be deployed to increase profitability.

Business Intelligence/Data Warehouse
I&C SCI X425.22 (3 units)
Learn how to make better business decisions, use fewer resources, and improve your company's bottom line by developing and using a data warehouse. This course provides an overview of business intelligence and data warehousing and gives you a look at all the major facets of developing and using a data warehouse to make effective business decisions. You'll work on a single project that allows you to develop a project plan and business case for a data warehouse, develop a dimensional model, develop a data staging process, and develop a data access process. You'll also learn about careers working with business intelligence and data warehousing as well as the educational requirements for this field.

R Programming
I&C SCI X425.20 (2.5 units)
R is a scripting language for statistical data manipulation and analysis. R is an open source package available under GNU license at no cost. R competes with SPSS, another very well-known statistical package used heavily in many industries. Statistics is used in every part of business data processing and prediction. Data captured by web analytics services need statistics. Statistics is also the foundation of predictive analytics. R business applications include correlation, regression, hypothesis testing, and all inference testing. This course will focus on R programming which is used for solving business problems related to basic math and statistics. First, all relevant math concepts will be reviewed. This will include functions, regression, descriptive and inferential statistics, and matrix operations. All these basic math problems will be solved using R. The programmatic interface and graphic capabilities of R will also be explored. In the end, several real-world business problems will be solved using R.

Hadoop: In Theory and Practice
I&C SCI X425.18 (3 units)
Today, organizations in every industry are being showered with imposing quantities of new information. Along with traditional sources, many more data channels and categories now exist. Collectively, these vastly larger information volumes and new assets are known as big data. Enterprises are using technologies such as MapReduce and Hadoop to extract value from big data. This course provides an in-depth overview of Hadoop and MapReduce, the cornerstones of big data processing. To crystalize the concepts behind Hadoop and MapReduce, you will work through a series of short/ focused exercises; you will configure and install a Hadoop cluster, write basic MapReduce programs, gain familiarity with advanced MapReduce programming practices, and utilize interfaces such as Pig and Hive to interact with Hadoop. You will also learn about real-world situations were MapReduce techniques can be used.

Big Data Analytics
I&C SCI X425.70 (2 units)
The amount and complexity of both internal and external data that can be used to improve the operational efficiency of today's businesses continues to expand dramatically. Having the skills and understanding the tools required to effectively leverage these resources is critical order for businesses to remain competitive. This course covers methods and algorithms for mining massive data sets and discusses how to utilize this data to make better business decisions and improve business efficiencies. The course will begin by ensuring you have an understanding of how to frame business goals as a data/analytics challenge and to develop a structured solution. Next, learn how to select and apply an appropriate analytics solution and use open source tools such as R and WEKA, and parallelization tools including Hadoop MapReduce. Big data types, use cases and design elements will be also be covered along with a discussion of the related organizational structure and cultural issues.

Defining Business Goals for Predictive
I&C SCI X425.62 (2 units)
Predictive analytics can be used to seek out increasingly small margins and understand a company's customers, products, channels, partners, and more. However, predictive analytics is only part of the solution. These analytic insights must be put to work and used properly to aid in making better business decisions. This course covers the framework to identify which decisions can be improved. Learn how to build business support for predictive analytic efforts, scope them appropriately, and develop a plan for successful implementation. The curriculum also includes an introduction to decision management, a proven approach to linking predictive analytics to a business and its goals.

Effective Data Preparation
I&C SCI X425.63 (2 units)
Broadly speaking, data preparation for data mining consists of three (3) elements: 1. Data Mining Process delineation (understanding the overall process) 2. Data Understanding (data cannot be properly prepared without first understanding it) 3. Data Pre-processing (transforming data into a form compatible with data mining). This intensive hands-on course gives students the skills necessary to extract stored data elements, understand what they mean in the company, transform their formats, and derive new relationships among them to produce a dataset suitable for analytical modeling. Students will learn how to produce a fully processed data set compatible for building powerful predictive models that can be deployed to increase business profitability.

Java Programming I
I&C SCI X460.10 (3 units)
Java is an excellent choice for those new to programming, wishing to enhance their current skillset or change their career. The aim of this course is to provide students with the knowledge and competencies to be able to write and design sophisticated professional programs using Java through extensive hands-on exercises. Included are core language concepts including fundamental data types, flow control, and standard function libraries. The course emphasizes object-oriented programming and modular design to support distributed development environments. Included are the design of classes and objects inheritance and polymorphism, and file I/O. The course also includes coverage of the Java Collections API, debugging, exceptions, and the Eclipse IDE. Optional modules throughout the course enforce basic concepts for those who are new to programming. After completion of this course, students are prepared for Java Programming II (I&CSCI X460.11).

Object-Oriented Analysis and Design
I&C SCI X471.91 (3 units)
Learn how to develop better, more complex software by applying good analysis and design methodologies. You'll learn how to approach software development systematically and how standardized notation such as the Unified Modeling Language (UML) allows you to map out the functionality of an application before writing any code. You'll also learn good practices for defining classes and methods so that your applications run efficiently and are easier to maintain. Anyone using object-oriented programming language including C+ +, Java, VB.NET, or C# can benefit from this course.

Introduction to Python for Data Analysis
I&C SCI X426.62 (1.5 units)
An introductory course aimed at students who want to learn python for the purposes of importing and analyzing data. The goal of the course is to introduce enough python so that, after the course, the student can immediately use what they have learned to ingest data, produce plots and analysis, and fit models. Students should note that not everything with-in the python language will be covered (such as user interfaces, web services, and object oriented programming). The main python libraries introduced will be numpy, matplotlib, pandas, and scikit-learn. Along with the basics of the python language, students will learn how to import data and manipulate it efficiently using numpy, how to produce plots and data visualizations with matplotlib, how to run statistical analysis using pandas, and how to build predictive models scikit-learn. There will be a final project that ties everything we have learned together. Students will also learn how to use eclipse, a very handy development environment!

Advisory Committee

Dean Abbott, President, Abbott Analytics

John Elder, Ph.D., Chief Scientist, Elder Research

Bernie Jeltema, Principal Consultant, Strategic Frameworks, Inc.

Derrick Lam, Director, Advancement Information Management Department, Office of Information Technology, University of California, Irvine

Amit Manghani, Director and Solution Manager, Business Analytics & Technology, SAP AG Business Objects Division

Gary Miner, Ph.D., Senior Consultant/Statistician, Statsoft

Bob Nisbet, Ph.D., Consulting Data Scientist

James Taylor, CEO, Decision Management Solutions

Eric Siegel, Ph.D., Chair, Predictive Analytics World

Jim Sterne, President, Target Marketing, Chairman, Digital Analytics Association, Founder, eMetrics Marketing Optimization Summit

Academic Management

Dave Dimas, Ph.D., Director, Engineering, Sciences and Information Technologies

Data Science
Certificate Program

UCI Extension

Jackie Badwah ■ (949) 824-3413 ■ jdbadwah@uci.edu

09.11.15

extension.uci.edu/datascience