National University of Science and Technology
POLITEHNICA Bucharest
FACULTY OFAUTOMATION AND COMPUTERS
Splaiul Independenței 313, Sector 6, 060042- București
Web www.acs.pub.ro
Tel +40 021 410 104, Fax +40 021 318 1014, Email decanat@acs.pub.ro

Master program

Advanced Analytics for Business

Disciplines in the curriculum plan

Fundamentals of Data Science in Business and Engineering

This course is a prerequisite for most statistical analysis courses for advanced data analysis. The course will provide students with an integrated view of statistical methods and their usefulness in various industries. By taking this course, students will learn how to become users of descriptive statistics and statistical inference, linear regression, logistic regression, and how to generate descriptive statistics, explore data using graphs. Upon completing this course, students will gain the following knowledge: basic concepts in statistics (descriptive statistics and statistical inference, statistical population, samples; independent/dependent variables; types of variables: quantitative/categorical, scale of measurement, etc.); key notions in descriptive statistics (data description, measures of central tendency and variability); statistical representation of data, probability distributions; statistical inference: confidence intervals and statistical mathematics elements used in hypothesis testing; ANOVA test and simple/multiple linear regression; how to apply multiple comparison techniques, practically perform linear regressions and evaluate hypotheses, use regression model selection techniques to help choose predictor variables in multiple regression; work in real situations with diagnostic statistics to evaluate statistical hypotheses and identify potential outliers in multiple regression.

Programming Essentials for Data Processing

This course offers students the opportunity to learn how to write software programs to access, explore, and prepare for advanced data analysis. The course is structured into two parts: essential programming elements in data processing; data manipulation and an introduction to programming languages that support the organization of structured and unstructured information. The second part of the course includes knowledge about: algorithms and understanding the fundamentals of data processing; evaluating how advanced analysis processes can be used to address business problems; accessing and creating data structures; data management, report generation, and error handling; learning SAS programming for data science vs. programming in R/Python.

Visual Analytics Techniques

This course presents the main concepts related to data preparation and visual exploration to provide an understanding of the nature of the data and prepare it for further advanced analysis. The data visualization techniques presented in the course are used to extract relevant insights from the data (clusters, exceptions, outliers, correlations between data) without complex processing, especially for large volumes of data

characterized by volume (large amount of data), variety (structured, semi-structured, unstructured data), and velocity (the speed at which data is collected and changes). Students will learn: what data preparation, analysis, and report generation using advanced analysis software means; what it means to restructure and structure data for advanced analysis (geographical analysis, forecasting, network analysis, path analysis, text analysis); how to apply complex data structuring procedures for advanced analysis; what "data cleansing" means; to understand the advanced features of data preparation, analysis, and report generation; to create customized reports for use in decision management; to create advanced data reports and filters, as well as advanced interactive reports with parameters.

Python programming for business intelligence

This is a *project-based course* that focuses on creating practical applications, aiming to develop students' programming skills in the Python programming language and dual programming in Base SAS/Python. The course focuses on: installing and configuring software in a programming environment; describing generic programming language concepts; Python programming and integration in Base SAS; using Python functions in Base SAS (native SAS programming); an overview of specific applications in the field of machine learning and examples of machine learning applications (classification, regression, supervised/unsupervised learning); learning the Python language through the use of algorithms specific to the machine learning field; an introduction to the basic components of building and applying a range of machine learning methods based on models and algorithms (regression, classification trees, Naive Bayes, random forest, etc.), with an emphasis on practical applications (datasets, tests, error rates, etc.); intelligent methods for performance evaluation; downloading results and using native Python syntax to compare models.

R programming for business intelligence

This is a project-based course that focuses on creating practical applications, aiming to develop students' programming skills in the R statistical programming language. The course focuses on: an introduction to the R programming language; using the R API to take control of Cloud Analytic Services (CAS) actions from Jupyter Notebook; loading data into the distributed in-memory environment (CASLib), data analysis, and creating predictive models in CAS using R functionality through the SWAT package (SAS Wrapper for Analytics Transfer); learning the module for downloading client results and using native R syntax to compare models; an overview of specific applications in the field of machine learning, with a focus on examples of machine learning applications (classification, regression, supervised/unsupervised learning).

Introduction to Exploratory Data Analysis

This course presents the main steps of data preprocessing, immediately after data collection from various sources, through simple visualization and graphical representation, without assumptions, to help evaluate the data quality a priori. Students will learn how to: use EDA (Exploratory Data Analysis) techniques to understand, summarize, and investigate the data; examine how the collected data is distributed, identify outliers and anomalies in data presentation, in order to correctly inform the formation and further exploration of hypotheses; better understand how to analyze the structure of a database for data analysis; discover patterns in the data; synthesize and visualize important characteristics of a dataset; understand the structure and adjacent variables of the data, and how to extract and create relevant variables for building predictive models and advanced analysis; develop intuition about the dataset, consider how the dataset was generated, and decide how it can be investigated through more formal statistical methods; perform visual data analysis for a better understanding of quality issues (data quality); process and apply data "cleansing" procedures.

Predictive modelling

This course presents the main concepts related to the field of predictive analysis, a statistical method that uses specific machine learning and data mining algorithms to predict future outcomes based on historical and current data. Students will learn: the essential steps in the predictive modeling process, data collection, formulation of the statistical model, model evaluation, result analysis, and refining the predictive model as new data is collected; understanding the specifics of predictive analysis in business and its main challenges; how to apply predictive models in various case studies; what analyzing historical and current data and generating predictive models entails; the main models used in predictive modeling; what predictive analysis means; evaluating predictive models; assessing the factors that can affect the accuracy of the results (predictor efficiency).

Neural Networks Essentials

The course combines theory and practice to deepen the basic concepts of neural network models and the essential practices for using them in real-world applications. The course covers the following aspects: the main paradigms for constructing ANN models (supervised/unsupervised/reinforcement learning); selecting essential model parameters in various case studies; the main neural network models used in practical applications (including multilayer perceptrons and RBF networks); how to choose a neural network architecture for a specific application and the appropriate training method; how to avoid the overfitting phenomenon; an introduction to deep learning; interpreting predictions made with ANN models; creating and evaluating ANN models in the workplace, and implementing ANN models.

Advanced visual text analytics

This course presents specific techniques for discovering information and hidden patterns in unstructured datasets, using the combined power of natural language processing, machine learning, and linguistic rules. The five characteristic components of visual text analysis are introduced: parsing, concept derivation, topic derivation, text classification, and sentiment analysis. It explains how to analyze documents (unstructured data) to determine the dominant themes in document collections. The course also covers the construction of complex linguistic queries for extracting and structuring information. By taking this course, students will learn to: explore text document collections to discover important topics; automatically identify key textual topics from large document collections; create robust content classification models based on the specific needs of an organization; create Boolean rules for classifying documents according to categorical target variables; extract sentiment scores at the document level; practically apply text processing knowledge in business applications, e.g., chatbots; use these skills on a large scale as part of marketing analysis and advanced market research.

Advanced visual data mining techniques

The course takes an interactive approach to presenting the visualization, evaluation, and implementation of data-driven models. Students will practically explore various machine learning techniques and case studies. The course provides a complex visual perspective on the main principles of data mining. The key components of the course focus on: presenting, training, and evaluating the performance of various architectures such as Bayesian networks, forest models, predictive models obtained through the gradient boosting method, artificial neural networks, SVM (Support Vector Machine) networks, and general supervised learning algorithms (factorization machines); comparing the various models obtained; and transferring analytical models into data pipelines.

Research Design

This course presents the main methods of qualitative, quantitative, and mixed analysis, which involve the primary collection and processing of data. These methods will be used by students to formulate hypotheses (questions) in research projects, including the analysis of real data collected through questionnaires and hypothesis testing in decision-making processes. Students will learn to: understand what quantitative data analysis and qualitative data analysis mean in business and research projects; understand the main data collection methods (interviews, questionnaires, etc.), the advantages and disadvantages of using them in various approaches (i.e., action research, case studies, surveys, etc.); create complex questionnaires (surveys) in a business environment and analyze data collected from real case studies; understand what statistical analysis for qualitative data entails and how to choose statistical tests; define a research model (composed of variables and hypotheses); apply the Likert scale for measuring variables associated with the research model.

Software Engineering Essentials

This course has the general objectives of acquiring the fundamental knowledge, principles, and methods of "Software Engineering," understanding the complexity and challenges of industrial-scale software development, as well as familiarizing students with teamwork in software projects. Understanding the specifics of software engineering, compared to other engineering and technical disciplines, is essential today, as even beyond the ICT industry, in many industrial sectors, almost all projects have a software component, whose importance is even greater in data-driven projects focused on data collection, analysis, and usage. The following aspects will be studied: common technical activities in software development — to understand the specifics and challenges of each; related management and quality assurance activities in software development; various common software development life cycle (SDLC) models, to understand the specifics, advantages, limitations, and challenges of each, and to develop the ability to choose the optimal model per project or integrate optimally into teams using certain models; pragmatic principles of software modeling and the UML language; and several other related topics through presentations by students.

Forecasting and Decision Support Systems

This course presents the main concepts related to forecasting models that use complex algorithms and how these forecasts can be used to support decision-making in a business environment. The course describes the main steps for forecasting and their use in decision processes, including preparing and loading data into memory, modeling and visualizing time series, and defining attribute-type variables for creating complex visualizations. Additionally, the course covers key notions regarding the use of data pipelines to generate forecasts and select the best pipeline for a project. This course will give students the opportunity to: use data analysis for practical purposes; understand how to use complex models to make forecasts; learn how decision support systems (DSS), i.e., interactive software-based decision-making systems, can assist future project managers in making decisions by accessing large volumes of information generated by various related information systems involved in organizational business processes; and incorporate large-scale forecasting practices into their forecasting projects.

Optimization Concepts for Data Science

This course focuses on the practical applicability of various optimization concepts (linear, nonlinear, and mixed linear), presenting and illustrating various case studies. The course covers: a brief introduction to optimization techniques; the mathematics of optimization; a short introduction to optimization methods: linear/nonlinear/integer and mixed integer linear programming; identifying and formulating the most useful approach for solving various optimization problems; improving the efficiency of formulations;

demonstrations with examples of portfolio formulation and optimization; specific procedures and open-source tools for formulating and solving optimization problems; and optimization models commonly used in industry.

Data Ethics and smart service design

This course addresses issues related to how modern methods of collecting and analyzing increasingly large amounts of data can raise ethical concerns in society. The course describes the design problem using data in various formats, which requires understanding how to ethically approach a data-intensive project. Data ethics is relevant at all stages of such a project's development: data stewardship – collection, management, sharing; extracting information from data – in the form of services, products, analyses, insights, visualizations, and data dashboards; decision-making – informing decision-makers by aggregating information from multiple sources and analysis experience. Students will learn how to: define the difference between morality, ethics, and law; explain why data ethics applies in a much broader context than just data usage; understand what responsible data science means – ensuring fairness, accuracy, confidentiality, and transparency at the service or product design level; identify and learn to manage various ethical aspects at different stages of a project; understand various frameworks for ethical data use; and apply the knowledge gained in designing smart services that heavily rely on data.