NATIONAL UNIVERSITY

Syllabus

POSTGRADUATE DIPLOMA (PGD) in DATA ANALYTICS

Effective From: 2022-2023 Session

NATIONAL UNIVERSITY

POSTGRADUATE DIPLOMA in DATA ANALYTICS

Credential: Postgraduate Diploma in Data Analytics
Length of Program: One year
Credits: 40.00

Admissions Requirements
Applicants must have an undergraduate degree. Applicants with an undergraduate degree other than Science, Technology, Business, or Mathematics/Statistics will need to show proficiency in the Business/ Technology field by way of extra coursework/work experience and can be evaluated on a case-by-case basis.

Course Objectives and Outcomes
This course aims to produce technically competent, innovative graduates who will become leading practitioners in the field of data analytics. Upon completion, graduates will be able to:

- Conduct independent research and analysis in the field of data analytics.
- Demonstrate expert knowledge of the tools, techniques, and technologies of data analytics utilized in both technical and business contexts.
- Critically assess and evaluate business and technical strategies for data analytics.
- Develop and implement effective business and technical solutions using data analytics.
- Critically appreciate ethical and data governance issues relevant to data analytics.

Students undertaking this course will be exposed to a variety of programming languages/ tools that may include R, SPSS, Excel, and KOBO.

Career Prospects
This course is designed to meet the ever-growing need for deep skills in Big Data/Analytics and to fill a skills shortage in Bangladesh as well as around the globe. Graduates from this course will enter the job market in the positions of Risk Analytics Intern, QA Analyst, Pricing Analyst, Financial Reporting Consultant, Applied Data Scientist, Data Analyst, Lead Data Scientist, Managed Services Consultant, Revenue Analyst, and various other data analyst positions. Learners will also avail the huge opportunity to work as freelancer data analysts.

Graduation Requirements
Successful completion of 40 credits.
Must achieve a program CGPA of 2.00 or higher.
Evaluation of the program courses will be held with the theoretical exam at the end of the semester, and lab experiments with continuous assessment.

Curriculum Framework

Course Code	Course Title	Credits
First Semester		
816801	Fundamentals of Data Analytics	4.00
816803	Introduction to Programming	4.00
816805	Data Analytics: Tools and Techniques	4.00
816807	Database Systems	4.00
816809	Data Analysis and Software Application	4.00
Second Semester		
826811	Data Visualization	4.00
826813	Data Governance and Ethics	4.00
826815	Fundamentals of Machine Learning	4.00
826817	Professional Development & Freelancing	4.00
Project/Internship		
826818	Special Topics in Data Analytics	4.00
Total Credits		**40.00**

NATIONAL UNIVERSITY

POSTGRADUATE DIPLOMA (PGD) in DATA ANALYTICS

Detailed Syllabus

First Semester

Course Code: 816801	Credits: 4	Class Hours: 4 contact hours per week
Course Title: Fundamentals of Data Analytics		
Course Evaluation: Theoretical exam-60%, Lab with continuous assesment-40%		

Course Objectives: This course will provide essential knowledge on the basic understanding of the emerging data analytics field. The student will learn the data exploration methods, dimensional analysis, and some extent of data visualization. The theoretical knowledge and practical lessons learned from this course will equip the students to absorb their skill sin their respective industries significantly.

Course Contents:

1. **Introduction to Data Analytics:** Data overview; State of the practice in Analytics; Key roles for the new Big data ecosystem.
2. **Data Analytics Lifecycle:** Overview of data analytic lifecycle; Various phases of data analytics life cycle; Global innovation network and analysis (GINA).
3. **Data Mining Process:** Definition; Data extraction as a process; Data mining models-CRISP-DM, SEMMA; Steps in the data mining process; Data mining process in a data warehouse, Applications of data mining; Data mining challenges.
4. **Review of Basic Data Analytics Methods:** Descriptive analytics, Diagnostic analytics, Predictive analytics; Prescriptive analytics.
5. **Business Intelligence Trends and Big Data Trends:** What is big data and why it is important; 5V's of big data; Goals of big data; Big data architecture, analytics & techniques; Future trends of big data.

Practical:

Review of basic data analysis methods using MS Excel and R.

Textbook:

- EMC Education Services (2015). **Data Science & Big Data Analytics**. Wiley.
- Michael Minelli, Michele Chambers, Ambiga Dhiraj (2013).**Big Data**, **Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses**, Wiley.

Course Code: 816803	Credits: 4	Class Hours: 04 Contact hours per week
Course Title: Introduction to Programming		
Course Evaluation: Theoretical exam-40%, Lab with continuous assesment-60%		

Course Objectives: This course introduces students to computing science and programming. It includes fundamental concepts and terminologies of computing science, program design, and fundamental building blocks for programming in a high-level language. At the end of this course, the successful students will be able to: explain the process of program design and development; design a GUI by using objects (forms and controls) and managing properties of controls; design a program to solve a well-defined problem; and implement a program from a design using structured programming.

Course Contents:

1. **Programming in general:** Program development cycle; Programming tools – flowcharts, pseudo code.
2. **Programming in C#, NET:** Forms, controls, properties, events; Numbers, variables, constants; Strings; Input and output using text boxes; Built-in functions – numeric functions, strings functions.
3. **Programming in Python:** Basics, Data types and Control structures, Modularization and Classes.
4. **Methods:** Value returning and void; Scope of variables; Value and reference parameters.
5. **Decisions:** Relational and logical operators; Switch statements
6. **Structure:** Repetitions, Arrays, Classes, and Fundamental algorithms in C# and Python.
7. **Files:** Reading and writing text files in C# and Python.

Practical:

Writing short programs in C# and Python for various statistical computing and simulations.

Textbook:

- Gaddis, Tony (2017). **Starting Out with Visual C#**. Fourth Edition, Pearson.
- Mark Lutz (2013).**Learning Python**, O'Reilly Media.

Course Code: 816805	Credits: 4	Class Hours: 04 Contact hours per week
Course Title: Data Analytics: Tools and Techniques		
Course Evaluation: Theoretical exam-40% , Lab with continuous assesment-60%		

Course Objectives: This course will provide knowledge regarding the advanced tools and techniques that are used for various data types. The concentration will be given to the application of suggestive tools in sophisticated software packages rather than enormous theories crunching. Students will learn the possible solutions for the situations of the respective problem for which the data was generated.

Course Contents:

1. **Overview of Data Analytic Tools.**

2. **Descriptive Measures of Data (Basic statistical measures):** Types and measurement scales of data; Measures of central tendency; Measures of dispersion.
3. **Inferential Decision from Data (basic estimation and test of hypothesis):** Basics of inferential statistics; Power of test and p-value; Confidence intervals and tests of means and proportions.
4. **Correlation and Regression:** Simple and multiple correlations; Simple and Multivariate linear regression; Logistic Regression; Additional Regression Models.
5. **Classification:** Overview; Decision Trees; Naïve Bayes; Diagnostics of Classifiers.
6. **Time Series Analysis:** Decomposition of time series data; Plotting time series data; Box-Jenkins Methodology; ARMA models; ARIMA models.
7. **Text analysis:** Text Analysis Steps; Collecting Raw Text; Representing Text; Term Frequency-Inverse Document Frequency (TFIDF); Categorizing Documents by Topics; Determining Sentiments.

Practical:

Solution for various real-life situations depending on the observed data.

Textbook:

- EMC Education Services (2015). **Data Science & Big Data Analytics**– Wiley.

Course Code: 816807	**Credits: 4**	**Class Hours: 04 Contact hours per week**
Course Title: Database Systems		
Course Evaluation: Theoretical exam-40% , Lab with continuous assesment-60%		

Course Objectives: This course will provide the student with knowledge of database concepts. Emphasis will be placed on database design, development, and querying using DBS installed on both LAN and cloud environments. Concepts covered include ER modeling, normalization, and database design theory. Hands-on activities will allow students to create, maintain and query various databases. An introduction and overview of stored procedures, triggers, and data warehousing will also be covered.

Course Contents:

1. **Overview:** Purpose of Database Systems; Database Languages; Database Design; Database Engine; Database and Application Architecture; History of Database Systems.
2. **Relational Models:** Structure of Relational Databases; Database Schema; Relational Query Languages; The Relational Algebra; Introduction to SQL; Intermediate SQL; Advanced SQL.
3. **Database Design:** Database Design Using E-R Model; Relational Database Design.
4. **Application Design and Development:** Complex Data Types; Application Development.
5. **Big Data Analytics:** Big Data; Data Analytics.

6. **Storage Management and Indexing:** Physical Storage Systems; Data Storage Structures; Indexing.
7. **Query Processing and Optimization:** Query Processing; Query Optimization.
8. **Transaction Management:** Transactions; Concurrency Control; Recovery Systems.
9. **Parallel and Distributed Databases:** Database-System Architecture; Parallel and Distributed Storage; Parallel and Distributed Query Processing; Parallel and Distributed Transaction Processing.
10. **Advanced Topics:** Advanced Indexing Techniques; Block chain Databases.

Practical:

Processing different types queries for data preparation using SQL**.**

Textbook:

- Abraham Silberschatz, Henry F. Korth, S. Sudarshan (2019). **Database System Concepts**. Seventh Edition, McGraw-Hill, New Delhi, India.
- Kroenke. **Database Concepts**. Latest Edition. Prentice Hall.
- Renee M. P. Teate (2021). SQL for Data Scientists. Wiley

Course Code: 816809	**Credits: 4**	**Class Hours: 04 Contact hours per week**
Course Title: Data Analysis & Software Application		
Course Evaluation: Theoretical exam-40%, Lab with continuous assesment-60%		

Course Objectives: This course will provide the student with knowledge of various data collection and data analysis software packages and their uses from database concepts. Emphasis will be placed on hands-on activities that will allow students to create, maintain and explore different queries and questions from various datasets.

Course Contents:

1. **Data collection and organization in KOBO Toolbox:** Introduction; Getting started with KoBo Toolbox; Creating a project; Adding questions; Implementing skip checks; Implementing validation logic; Deploying the form & collecting data; Configuration, data entry & uploading; Viewing and downloading data.
2. **Analysis of quantitative data using SPSS:** Introduction to statistical analysis; Data manipulation and data transformation; Computing descriptive statistics; Different types of graphs: Bar chart, Pie chart, Histogram, Scatter plot, etc.; Test of hypothesis: Z-test, t-test, Chi-square test, F-test, Nonparametric test; Correlation and Regression: simple and multiple correlations, simple and multiple regression, logistic regression; Analysis of Variance (ANOVA): CRD and RBD; Factor Analysis and Discriminant Analysis
3. **Qualitative data analysis with NVivo:** Introduction to NVivo; Designing an NVivo project; Coding foundation; Advanced coding; Cases, classifications, and comparisons; Surveys and mixed methods; Querying data; Working with multimedia files; Social media data; Teamwork in NVivo.

4. **Data Analysis with R:** Introduction to R; Data manipulation using *Tidyverse*; Control flow and functions; Data visualization with *GGplot*; Exploratory data analysis; Correlation and regression; Basic machine learning techniques.

Practical:

In-practice lessons with ***KOBO, SPSS, NVivo,*** and***R***.

Textbook:

- Garrett Grolemund, and Hadley Wickham (2017). **R for Data Science: Import, Tidy, Transform, Visualize, and Model Data**, 1st Edition, O' Reilley Media.
- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani (2013). **An Introduction to Statistical Learning with Application in R**, Springer.
- Kristi Jackson and Pat Bazeley (2019). **Qualitative Data Analysis with NVivo**. Third Edition, SAGE Publications Ltd.
- Stephen A. Sweet and Karen A. Grace-Martin (2011). **Data Analysis with SPSS: A First Course in Applied Statistics**- 4th Edition, Pearson.
- References for **SPSS, NVivo & KOBO manuals**.

Second Semester

Course Code: 826811	Credits: 4	Class Hours: 04 Contact hours per week
Course Title: Data Visualization		
Course Evaluation: Theoretical exam-40%, Lab with continuous assesment-60%		

Course Objectives: This course introduces data literacy required as a key twenty-first-century skill. Students will learn the nature of data across different domains and the concepts and skills of data visualization by understanding, questioning, and problematizing how data are generated, analyzed, and used. Students will be able to apply its concepts and skills to visualize their own data, interpret the findings, and examine the impacts of the data-driven decision.

Course Contents:

1. **Introduction:** Introduction to data visualization; Brief history of data visualization.
2. **Principles:** Data graphics; Static graphics; Data visualization through their graphic presentation; Graph-theoretic graphs, High-dimensional data visualization; Multivariate data glyphs: principles and practice; Linked data views; Visualizing trees and forests.
3. **Methodologies:** Linked micro-map plots; Grand tours, projection pursuit guided tours, and manual control; Multidimensional scaling; Multivariate visualization by density estimation; Structured sets of graphs; Regression by parts: fitting visually interpretable models with GUIDE; Smoothing techniques for visualization; Visualizing cluster analysis and finite mixture models; Visualizing contingency table; Mosaic plots and their variants; Matrix visualization; Web-based statistical graphics using XML technologies.
4. **Application:** Reconstruction, visualization, and analysis of medical images; Exploratory graphics of a financial dataset, Graphical data representation in bankruptcy analysis, Visualization tools for insurance risk process.

Practical:

Lab experiments on data visualization in ***Excel, R,*** and ***Python***.

Textbook:

- Chun-houh Chen, Wolfgang Härdle, Antony Unwin (2008). **Handbook of Data Visualization**, First Edition, Springer-Verlag Berlin Heidelberg.
- Sosulski, K. (2018). **Data Visualization Made Simple: Insights into Becoming Visual**. New York: Routledge.
- Hadley Wickham (2016). **ggplot2, Elegant Graphics for Data Analysis**, Springer.
- Winston Chang (2012). **R Graphics Cookbook**, O'Reilly Media.
- Jake VanderPlas (2016). **Python Data Science Handbook: Essential Tools for Working with Data**, O'Reilly Media.

Course Code: 826813	Credits: 4	Class Hours: 04 Contact hours per week
Course Title: Data Governance and Ethics		
Course Evaluation: Theoretical exam-60% , Lab with continuous assesment-40%		

Course Objectives: This course will provide an understanding of data governance policy and ethical considerations while working as a data analyst. The students will be able to critically interpret the governance and regulatory frameworks associated with the capture, processing, and stewardship of data. They will also gain skills to critically interpret the roles and responsibilities of data security, privacy, and data protection. Learners will also build their capacity to analyze and evaluate the intersection of data and ethics in socio-technical environments.

Course Contents:

1. **Data Governance:** Data quality and provenance; Data management-Roles and responsibilities. Management of data policies, processes, and procedures; Data integrity & security; Risk management; Models and tools for data governance.
2. **Privacy and Data Protection:** The right to privacy - constitutional and statutory protections, privacy, and the European Convention on Human Rights and EU Charter of Fundamental Rights; Common law protection; Data Protection Regulation Scope, processing of personal data, legitimate bases; Principles of data protection; Sensitive data; Issues of consent; Rights, supervision and enforcement; Data Protection in practice includes international transfers, surveillance, cloud computing, and auditing; The current reform of the area.
3. **Ethical Issues Pertaining to Data:** Ethics and Computing - examining moral problems when using the Internet - spam, censorship and free speech, the anonymity offered by the Internet; Ethical issues that arising from the increasing use and pervasiveness of information technology and socio-technical systems; Health technology; Pervasive monitoring and tracking; Image, video and sound capture; Identity; Perpetuity of data storage; Trans nationality; Copyright; IOT.

Textbook:

- Katherine O'Keefe, Daragh O Brien (2018). **Ethical Data and Information Management**, Kogan.
- Herman T. Tavani (2013). **Ethics and Technology**, Wiley.

Course Code: 826815	Credits: 4	Class Hours: 04 Contact hours per week
Course Title: Fundamentals of Machine Learning		
Course Evaluation: Theoretical exam-40%, Lab with continuous assesment-60%		

Course Objectives: In this course, students will learn to apply machine learning concepts to analyze data and make predictions. Students will learn how to collect and wrangle data, explore data using statistics and visualizations, transform data for further modeling, model data using machine learning algorithms to predict data patterns, and evaluate these model-based predictions.

Course Contents:

1. **Supervised learning Vs. Unsupervised learning.**

2. **Maximum likelihood.**
3. **Naive Bayes.**
4. **Logistic regression.**
5. **Gradient descent.**
6. **Regularization.**
7. **Information theory.**
8. **Decision trees.**
9. **Representation learning / PCA.**
10. **Clustering.**
11. **Latent variable models.**
12. **Neural networks**

Textbook:

- Shai Shalev-Shwartz and Shai Ben-David (2014)**. Understanding Machine Learning: From Theory to Algorithms**. Cambridge University Press. *Available free online.*
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman (2009). **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**, 2nd Edition. Springer. *Available free online*
- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani (2013). **An Introduction to Statistical Learning with Application in R**, Springer.
- Jake VanderPlas (2017). Python Data Science Handbook: Essential Tools for Working with Data, O'Reilly Media.

Course Code: 826817	**Credits: 4**	**Class Hours: 04 Contact hours per week**
Course Title: Professional Development and Freelancing		
Course Evaluation: Theoretical exam-40% , Lab with continuous assesment-60%		

Course Objectives: This course will connect students' education in the program and their work experiences to their post-graduation goals, whether that is obtaining a new job, earning a promotion, and/or freelancing. The goal is for students to focus on professional development topics so they feel confident in marketing their skills after graduation and have concrete materials that will make them stand out to employers and/or clients. By the end of the class, students should be aware of job search trends, the targeted documents necessary to be employable after graduation, and the importance of networking. Students will also learn job-search skills specific to the online communications field such as writing resumes and cover letters, presenting analytic works effectively, networking, and facing data-analytics job interviews. They will also be able to execute tactics that establish themselves as industry experts and explore additional sources of income.

Course Contents:

1. Career Development Theories; Conducting Career Research.
2. Networking; Networking Tools; E-Portfolios.
3. Conducting a Targeted Job Search; Resumes and Cover Letters.
4. Interviewing; Effective job presentation.
5. Decision Making; Best Practices in the Workplace.
6. Setting Business Goals; Working with Clients.

7. Client Documents and Forms – Proposals, Bids, Contracts, Policies, and Procedures; Creating Forms (ex. Needs Assessments).
8. Project Organization; Requests for Proposals.

Textbook:

- John J. Sonmez (2015). **Soft Skills.** Manning Publications, Shelter Island, NY 11964.
- Kevin Donlin (2000). **Résumé and Cover Letter Secrets Revealed**! Available online for free.
- Adam Sinicki (2019). **Thriving in the Gig Economy: Freelancing Online for Tech Professionals and Entrepreneurs.** Apress Media LLC.

Course Code: 826818	**Credits: 4**	**Class Hours:**
Course Title: Special Topics in Data Analytics		

Students will get an opportunity to conduct innovative research for Data Analytics in an industry of their choice. This course will provide students with the required breadth to jumpstart their careers in the Data Analytics field. During the first quarter of the Second semester, students will be asked to select a specialized topic according to his/her interest and submit a thesis/experimental project report during the semester final examination. The work will be supervised by an internal supervisor and the submitted thesis/report will be evaluated by an expert in a similar research field.