هيئة تقويم التعليم والتدريب
Education & Training Evaluation Commission

2023
TP-153

Course Specification
(Bachelor)

Course Title:	Data Science Fundamentals
Course Code:	CSC 1407
Program:	Bachelor in Computer Science
Department:	Computer Science
College:	Faculty of Computers and Information Technology
Institution:	University of Tabuk
Version:	1.0
Last Revision Date:	27 July 2022

Table of Contents

A. General information about the course:

1. Course Identification

1. Credit hours: (3 hours)					
2. Course type					
A.	☐University	☐College	☒ Department	☐Track	☐Others
B.	☐Required		☒ Elective		
3. Level/year at which this course is offered: (Level 7/8 – Year 4)					
4. Course general Description:					
Data Science is a rapidly emerging discipline at the intersection of statistics, machine learning, data visualization, and mathematical modeling. This course is designed to provide a hands-on introduction to Data Science by challenging student groups to build predictive models for upcoming events, and validating their models against the actual outcomes.					
5. Pre-requirements for this course (if any)**:**					
CIT 1305, STAT1101					
6. Co-requisites for this course (if any)**:**					
NA					
7. Course Main Objective(s):					
• **Provide Insights About the Roles of a Data Scientist** • **Enable You to Analyze Big Data** • **Learn Techniques and Tools for Transformation of Data** • **Make You Understand Data Mining** • **Familiarize You with Different Formats** • **Figure Out how to use Machine Learning Algorithm on your data** • **Learn Data Visualization and Optimization**					

2. Teaching mode (mark all that apply)

No	Mode of Instruction	Contact Hours	Percentage
1	Traditional classroom	45	100%
2	E-learning		
3	Hybrid		

No	Mode of Instruction	Contact Hours	Percentage
	• Traditional classroom • E-learning		
4	Distance learning		

3. Contact Hours (based on the academic semester)

No	Activity	Contact Hours
1.	**Lectures**	45
2.	**Laboratory/Studio**	
3.	**Field**	
4.	**Tutorial**	
5.	**Others (specify)**	
Total		45

B. Course Learning Outcomes (CLOs), Teaching Strategies and Assessment Methods

Code	Course Learning Outcomes	Code of PLOs aligned with program	Teaching Strategies	Assessment Methods
1.0	**Knowledge and understanding**			
1.1	Choose an appropriate strategy to analyze the problems.	K2	Lectures, Case studies, Group discussion	Exams Project Assignments
1.2	Understand the role of probability theory in the solution of data science	K1		
1.3	Understand the appropriate logic and computer syntax for solution development.	K3		
1.4	Utilize appropriate models for data understanding.	K4		
2.0	**Skills**			
2.1	Apply data science in the solution of computing problems.	S1	Lectures	Exams

Code	Course Learning Outcomes	Code of PLOs aligned with program	Teaching Strategies	Assessment Methods
2.2	Utilize data science tools to solve and visualize problems.	S4	Research Activities	Project Assignments
2.3	Use appropriate models for solution development.	S2		
2.4	Design a successful intelligent solution to address required problem.	S3		
3.0	**Values, autonomy, and responsibility**			
3.1	Utilize a modern software development environment and programming tools.	V1	Lecture	Project

C. Course Content

No	List of Topics	Contact Hours
1.	**Introduction to data science and mathematical preliminaries:** Probability vs. Statistics, Compound Events and Independence, Conditional Probability, Bayes Theorem, Distributions of Random Variables, Probability/Cumulative Distributions, Descriptive Statistics, Centrality Measures, Aggregation as Data Reduction, Variance Metric, Parameterizing Distributions, Interpreting Variance	3
2.	**Data correlation, munging, and cleaning:** Correlation Analysis, The Pearson Correlation Coefficient, Interpreting Correlations, Variance Reduction and R^2, Interpreting Correlation: Significance, Spearman Rank Correlation, Correlation vs. Causation, Autocorrelation and Periodicity, Logarithms and Ratios, Logarithms and Power Laws, Normalizing Skewed Distributions	3
3.	**Statistical distribution and significance (part 1):** The Central Dogma of Statistics, Statistical Data Distributions, Significance of Classical Distributions, Binomial Distributions, The Normal Distribution, Lifespan Distributions, The Poisson Distribution, Power Law Distributions	3
4.	**Statistical distribution and significance (part 2):** Significance and Classification, Differences in Distributions, The T-Test, The Kolmogorov-Smirnov Test, Normality Testing, The Bonferroni Correction, The Significance of Significance, Measures of Effect Size, Bootstrapping P-values, Permutation Tests, Sampling from Distributions, Monte Carlo Simulation	3
5.	**Practice of data visualization:** Box and Whisker Plots, Stacked Area vs. Line Plots, Data Maps and Cartograms, Non-Geographic Data Maps, Tools for Data Visualization, Repetitions for Multivariate Data, Understanding Color Scales,	3

	Tufte's Visualization Aesthetic, Great Data Visualizations, Terrible Professional Visualizations	
6.	**Principle of data visualization:** Exploratory Data Analysis, Ascombe's Quartet, Maximize Data-Ink Ratio, Graphical Integrity: Scale Distortion, Aspect Ratios and Lie Factors, Reduce Chartjunk, MatPlotLib, Tabular Data, Dimensions for Improvement, Multivariate Data visualization types	3
7.	**Building models**: Data Science Analysis Pipeline, Philosophies of Modeling, Occam's Razor, Bias-Variance Tradeoffs, Principles of Nate Silver, Properties of Probabilities, Scores to Probabilities, Modeling Methodologies, Baseline Models, Bayesian Reasoning, Taxonomy of Models, Discrete vs. Continuous Models, General vs. Ad Hoc Models, Steps to Build Effective Models	3
8.	**validating models:** Evaluating Classifiers, Threshold Classifiers, Accuracy, Precision, Recall, F-Score, Receiver-Operator (ROC) Curves, Evaluating Multiclass Systems, Confusion Matrix, Summary Statistics: Numerical Error, Evaluation Data, Evaluation Environment Architecture, Error Histograms: Dating Documents, Cross-Validation, Probability Similarity Measures, Evaluation Statistics	3
9.	**Linear Algebra Review:** Linear Algebra Formulae, Points vs. Vectors, Visualizing Matrix Operations, Linear Combination, Properties of Matrix Multiplication, Multiplying Feature Matrices, Matrix Inversion and Linear Systems, Factoring Matrices, LU Decomposition, Eigenvalues and Eigenvectors, Reconstructing a Covariance Matrix, Error Declines with Dimensionality	3
10.	**Linear Regression:** Singular Value Decomposition, Reconstruction from SVD, Linear Regression and Duality, Error in Linear Regression, Linear Regression in One Variable, Connections with Correlation, Outliers and Linear Regression, Fitting Non-Linear Functions, Feature Scaling: Z-scores, Dominance of Power Law Features	3
11.	**Logistic regression and classification:** Classification Problems, Regression for Classification, Decision Boundaries, The Logit Function, Scoring for Logistic Regression, Logarithms of Probabilities, Cost/Loss Function, Logistic Regression via Gradient Descent, Balanced Training Classes, Multi-class Classification, Hierarchical Classification	3
12.	**Clustering:** Supervised / Unsupervised Learning, K-Means Clustering, Expectation Maximization (EM), Agglomerative Clustering, Similarity Graphs, Spectral Clustering, Singular Value Decomposition	3
13.	**Introduction and topics in machine learning:** Subjective Rankings, XOR and Linear Classifiers, Decision Tree Classifiers, Information-Theoretic Entropy, Voting Classifiers, Boosting	3
14.	**Introduction to Big Data System and Data Analytics:** What is Big Data, How to Use Big Data, How to Manage Big Data, Data Analytics, Future of Big Data	3
15.	**Hadoop Intro Ecosystem Architecture:** what is Hadoop is and how it addresses big data challenges, the guiding principles behind Hadoop, the major components of the Hadoop ecosystem	3
Total		**45**

D. Students Assessment Activities

No	Assessment Activities *	Assessment timing (in week no)	Percentage of Total Assessment Score
1.	Mid Term Exam 1	6	15%
2.	Mid Term Exam 2	11	15%
3.	Project (3 phases)	7,11, and 15	20%
4.	Assignments	8, and 12	10%
5.	Final Exam	17	40%

*Assessment Activities (i.e., Written test, oral test, oral presentation, group project, essay, etc.).

E. Learning Resources and Facilities

1. References and Learning Resources

Essential References	1. Skiena, Steven S. The data science design manual. Springer, 2017. ISBN-10 : 9783319554433, ISBN-13 : 978-3319554433 2. Kotu, Vijay, and Bala Deshpande. Data science: concepts and practice. Morgan Kaufmann, 2018. ISBN-10 : 012814761X, ISBN-13 : 978-0128147610 3. Tom White, Hadoop The Definitive Guide (O'REILLY) 3rd Edition,2012, ISBN-10 : 1449311520, ISBN-13 : 978-1449311520.
Supportive References	**NA**
Electronic Materials	**NA**
Other Learning Materials	**NA**

2. Required Facilities and equipment

Items	Resources
facilities (Classrooms, laboratories, exhibition rooms, simulation rooms, etc.)	Classrooms
Technology equipment (projector, smart board, software)	Projector
Other equipment (depending on the nature of the specialty)	

F. Assessment of Course Quality

Assessment Areas/Issues	Assessor	Assessment Methods
Effectiveness of Teaching	Faculty, Program Leaders, and Advisory Board	Both Direct and Indirect

Assessment Areas/Issues	Assessor	Assessment Methods
	Students	Indirect
Effectiveness of Students Assessment	Faculty, Program Leaders, Advisory Board, and Independent Opinion	Both Direct and Indirect
Quality of Learning Resources	Faculty, Students, and Advisory Board	Indirect
The Extent to which CLOs have been Achieved	Faculty, Program Leaders, Advisory Board, and Independent Opinion	Direct (as in section B) and Indirect/Surveys
	Students	Indirect
Other	-	-

Assessors (Students, Faculty, Program Leaders, Peer Reviewer, Others (specify)

Assessment Methods (Direct, Indirect)

G. Specification Approval

COUNCIL /COMMITTEE	
REFERENCE NO.	
DATE	