East West University
Department of Computer Science and Engineering
Course Outline of CSE303
Spring 2023 Semester

Course: CSE303 Statistics for Data Science

Credits and Teaching Scheme

	Theory	Laboratory	Total
Credits	3	1	4
Contact Hours	3 Hours/Week for 13 Weeks + Final Exam in the 14th Week	2 Hours/Week for 13 Weeks	5 Hours/Week for 13 Weeks + Final Exam in the 14th Week

Prerequisite

STA102 – Statistics and Probability

Instructor Information

Instructor: **Dr Md Rifat Ahmmad Rashid**
Assistant Professor, Department of Computer Science and Engineering
Office: 635
E-mail: rifat.rashid@ewubd.edu

Course Objective

The objective of the course is to introduce the statistical methods, techniques and tools that are essential for Data science domain. The course focuses on examining descriptive and inferential statistics and analyzing the output of these methods. The course also focuses on relevant linear algebra methods which are necessary for understanding Data Science. The course also emphasizes techniques for result estimation and anomaly detection. Statistical machine learning methods that "learn" from data will be also introduced such as Linear Regression, Logistic Regression, Support Vector Machine and so on.

Course Outcomes (COs) with Mappings

After completion of this course students will be able to:

CO1	**Apply** various techniques of exploratory data analysis and inferential statistics for understanding the dataset and **analyze** the outcomes to explore interesting characteristics of the dataset.
CO2	**Apply** Linear Algebra, Regression and other statistical methods to determine outliers for smoothing and cleaning the dataset as well as to understand the correlation among attributes.
CO3	**Apply** different statistical learning models for classification of datasets to solve real-life problems and also **analyze** and **compare** their performance.
CO4	**Choose** and **justify** appropriate algorithms and tools for exploratory data analysis; **perform** and **demonstrate** skills and **write** report to design and implement statistical learning models using realistic data sets.

Course Topics, Teaching-Learning Method, and Assessment Scheme

Course Topic	Teaching-Learning Method	CO	Mark of Cognitive Learning Levels			CO Mark	Assessment (Mark)
			C3	C4	C5		
Introduction to Data Science and Basic Statistical Concepts	Lectures and discussions inside and outside the class	CO1					**Midterm Exam I (20)**
Exploratory Data Analysis	Do						
Different Data Distributions and Sampling Methods	Do						
Inferential Statistics and Hypothesis Testing	Do	CO1					**Midterm Exam II (20)**
Linear Algebra Basics, Different Matrix and Vector Operation	Do	CO2					
Linear Regression and its variants	Do						
Predictive analysis with Logistic Regression	Do	CO3					**Final Exam (20)**
Concepts of Bias and Variance, Support Vector Machine	Do						
Model Validation and Evaluation Metrics	Do						
Dimensionality Reduction using Principal Component Analysis	Do						

Laboratory Experiments and Assessment Scheme

Experiment	Teaching-Learning Method	CO	Mark of Cognitive Learning Levels			Mark of Psychomo-tor Learning Levels		CO Mark
			C3	C4	C6	P2	P3	
Introduction to Python Programming	Lab Experiment and Result Analysis	CO4						
Intermediate Python Programming	Do	CO4						
Pandas for Data Analysis	Do	CO4						
Exploratory Data Analysis and Data Visualization using Pandas	Do	CO4						
Introducing Numpy and Matplotlib Libraries for Data Wrangling and Visualization	Do	CO4						
Intermediate Data Visualization using Matplotlib Libraries	Do	CO4						
Linear Algebra in Numpy	Do	C04						
Linear Regression using Python	Do	CO4						
Logistic Regression and Support Vector Machine using Python								

Mini Projects and Presentations

Item	Teaching-Learning Method	CO	Mark of Cognitive Learning Level		Mark of Psychomotor Learning Levels		Mark of Affective Learning Level	CO Mark
			C3	C4	P3	P4	A2	
Presentations	Group-based, Relevant topics on	CO4	**2**	**1**			**2**	**5**

	Regression Analysis							
Lab-based Mini Project including Report and Presentation	Group-based moderately complex design project with report writing and oral/poster presentation	CO4	**2**	**3**	**2**	**2**	**1**	**10**

Overall Assessment Scheme

	CO				**Assessment Area Mark**
Assessment Area	**CO1**	**CO2**	**CO3**	**CO4**	
Class Participation					5
Class Test/Quizzes					10
Midterm Exam – I	20.0				20
Midterm Exam – II		20.0			20
Final Exam			20.0		20
Lab Performance, Lab Assignments or Lab Exam				10.0	10
Presentations, Mini Project				15.0	15
Total Mark	**20.0**	**20.0**	**20.0**	**25.0**	**100**

Teaching Materials/Equipment

Books:

- Data Science from Scratch (DSS) by Joel Grus
- Python for Data Analysis (PDA) by Wes McKinney
- Think Stats (TS) by Allen B. Downey.
- Optional: Python Data Science Handbook (PDSH) by Jake VanderPlas

Software/Tools:

- Anaconda Distribution https://www.anaconda.com/products/individual
- Google Colab https://colab.research.google.com/
- Other appropriate tools for data wrangling

** Lecture Slides and Lab Manuals will be made available to the students during the class in electronic form.*

Grading System

Marks (%)	Letter Grade	Grade Point	Marks (%)	Letter Grade	Grade Point
97-100	A+	4.00	73-76	C+	2.30
90-96	A	4.00	70-72	C	2.00
87-89	A-	3.70	67-69	C-	1.70
83-86	B+	3.30	63-66	D+	1.30
80-82	B	3.00	60-62	D	1.00
77-79	B-	2.70	Below 60	F	0.00

Exam Dates

As per the schedule provided by the university.

Academic Code of Conduct

Academic Integrity:

Any form of cheating (physical/online), plagiarism, personification, falsification of a document as well as any other form of dishonest behavior related to obtaining academic gain or the avoidance of evaluative exercises committed by a student is an academic offence under the Academic Code of Conduct and **may lead to severe penalties as decided by the Disciplinary Committee of the university.**

Special Instructions:

- Students are expected to attend all classes and examinations. A student MUST have at least 80% class attendance to sit for the final exam.
- Students will not be allowed to enter into the classroom after 10 minutes of the starting time.
- For plagiarism, the grade will automatically become zero for that exam/assignment.
- Normally there will be **NO make-up exam**. However, in case of **severe illness, death of any family member, any family emergency, or any humanitarian ground**, if a student miss any exam, the student MUST get approval of makeup exam by written application to the Chairperson through the Course Instructor **within 48 hours** of the exam time. Proper supporting documents in favor of the reason of missing the exam have to be presented with the application.
- For **final exam**, there will be NO makeup exam. However, in case of **severe illness, death of any family member, any family emergency, or any humanitarian ground**, if a student miss the final exam, the student MUST get approval of **Incomplete Grade** by written application to the Chairperson through the Course Instructor **within 48 hours** of the final exam time. Proper supporting documents in favor of the reason of missing the final exam have to be presented with the application. **It is the responsibility of the student to arrange an Incomplete Exam within the deadline mentioned in the Academic Calendar in consultation with the Course Instructor**.
- All mobile phones MUST be turned to silent mode during class and exam period.
- There is **zero tolerance for cheating** in exam. Students caught with cheat sheets in their possession, whether used or not; writing on the palm of hand, back of calculators, chairs or nearby walls; copying from cheat sheets or other cheat sources; copying from other examinee, etc. would be treated as cheating in the exam hall. The only penalty for cheating is **expulsion for several semesters as decided by the Disciplinary Committee of the university**.