FINAL EXAM ODD SEMESTER ACADEMIC YEAR 2022/2023
INFORMATION SYSTEMS STUDY PROGRAM
FACULTY OF ENGINEERING AND INFORMATICS
UNIVERSITAS MULTIMEDIA NUSANTARA

Subject	:	**IS388 Data Analysis (Theory)**	**Date**	:	
Lecturer(s)	:	**Samuel Ady Sanjaya** **Raymond Sunardi Oetama** **Ahmad Faza** **Rudi Sutomo**	**Time**	:	
Form	:	**~~Essay~~ / Project**	**Type**	:	**~~Onsite~~ / Takehome**

EXAM CONDITIONS / INSTRUCTIONS:

1. The following **must be submitted** via e-learning in **ZIPPED** format (*.zip)
   a. Individual report in MS Word format. Template is available from https://ejournals.umn.ac.id/index.php/IJNMT
   b. Summary of the exam in *.txt format that contains:
      - Data source link or reference
      - Dataset information (information of column and other attributes
   c. Dataset in format: *.csv, *.xlsx, OR *.json
   d. Python script with code and markdown in *.ipynb format
2. Submit via e-learning with the following format: Class Code_IS388_Theory_Full Name_Student ID.zip
3. If any student is caught copy-paste from friends, the internet, or another resource the Final Score will be F.
4. If the submission does not follow these rules then the maximum score is 60. So before collecting please make sure that it is in accordance with the collection format.

COURSE SUB LEARNING OUTCOMES (SUB-CLO):

SUB LEARNING OUTCOMES (SUB-CLO)		ELO
Code	**Description**	
SUB-CLO-1	Students can evaluate results of data analysis using Python (C5).	ELO A
SUB-CLO-1	Students can analyze data using appropriate techniques using Python (C4).	ELO B
SUB-CLO-1	Students can gather data and import data using various sources using Python (C2).	ELO C
SUB-CLO-1	Students can present data analysis result in groups or individual (C6).	ELO H

Every student writes an individual project report set for data analysis group project based on the following rules

1. Analysis of data and evaluation of results of data analysis are based on application of 2 or more algorithms (1 individual algorithm and 1 algorithm from group algorithms).
2. Complete report set (report, Python programs, data set, references, format and structure)

PROBLEM/QUESTIONS:

1. **Question 1: Sub-CLO 3 & 4 Weight (37.5%)**

Gather open-source datasets to acquire the data with the following criteria:

a. Datasets should have a minimum of 1000 rows and 6 columns, and should not be older than 5 years.
b. Datasets should contain both numeric and categorical variable
c. Datasets should in CSV, XLSX, or JSON format

Do a visualization for:

a. Exploratory Data Analysis
b. Machine Learning Result (Classification, Clustering, Regression or Association Rules)
c. Evaluation comparison between Algorithm #1 and Algorithm #2

ASSESSMENT RUBRIC (per question):

Rated aspect	Assessment criteria				
	Very Poor	Poor	Satisfactory	Good	Excellent
	< 45	45-54	55-69	70-84	(Score ≥ 85)
Datasets (12.5%)	• Do not have minimum 1000 rows and 6 columns • Do not contains both numeric and categorical variable Not in suitable format	• Do not have minimum 1000 rows and 6 columns • Do not contains both numeric and categorical variable In CSV, XLSX or JSON format	• Have minimum 1000 rows and 6 columns • Do not contains both numeric and categorical variable In CSV, XLSX or JSON format	• Have minimum 1000 rows but do not have 6 columns • Contains both numeric and categorical variable In CSV, XLSX or JSON format	• Have minimum 1000 rows and 6 columns • Contains both numeric and categorical variable In CSV, XLSX or JSON format
Data Visualization (25%)	• 1 type of visualization for EDA • Not suitable visualization • Can not perform visualization for Clustering/ Classification/Association Rules	• 2 type of visualization for EDA • Not suitable visualization • Can perform visualization for Clustering/Classification/ Association Rules, but not suitable	• 2 type of visualization for EDA • Suitable visualization • Can perform visualization for Clustering/Classification/ Association Rules, but not suitable	• 3 type of visualization for EDA • 2 visualization are suitable • Can perform visualization for Clustering/Classification/ Association Rules well	• 3 type of visualization for EDA • All visualization are suitable • Can perform visualization for Clustering/Classification/ Association Rules well

2. Question 2: Sub-CLO 1 & 2, Weight (62.5%)

Do Data Preprocessing, Modeling, Model Evaluation, and compare the performance of algorithm #1 and algorithm #2 for the chosen dataset. (Clustering/Classification/Regression/Association Rules)

ASSESSMENT RUBRIC (per question):

Rated aspect	Assessment criteria				
	Very Poor	Poor	Satisfactory	Good	Excellent
	< 45	45-54	55-69	70-84	(Score ≥ 85)
Data Preprocessing and Preparation (25%)	• Cannot perform normalization • Cannot perform binning or encoding	• Perform normalization incorrectly • Perform binning or encoding incorrectly	• Perform normalization with suitable technique • Perform binning or encoding incorrectly	• Perform normalization incorrectly • Perform binning or encoding with suitable technique	• Perform normalization with suitable technique • Perform binning or encoding with suitable technique
Modeling and Model Evaluation (in a paper format) (37.5%)	• Major mistakes, incoherent/ no analysis • No clear structure in introduction. • Conclusion is inconsistent to thesis statement. • No/ irrelevant review of algorithms. • No description of methods/ data. • Data file/ program is missing • Programs cannot run. • Inadequate references, fail to cite references. • Numerous grammatical and/ or typing errors. • Incorrect template/ format or	• Evaluate some aspect of performances • Logical flow of evaluation • Submit all required components of report set, • Programs produce most components used in report. • Paper cites in correct format. • Some structure in introduction. • Basic review of relevant algorithms • Conclusion is consistent to data analysis. • Describe most methodology and data. • Grammatical and typing errors.	• Evaluate some aspect of performances • Logical flow of evaluation • Submit all required components of report set, • Programs produce most components used in report. • Use specified template. • Paper cites at least 5 references within 5 years of publication in correct format. • Some structure in introduction. • Basic review of relevant algorithms • Conclusion is consistent to data analysis. • Describe most	• Well-organized, evaluate and compare performance of each algorithm using appropriate tools and visuals. • Submit all required components of report set, • Programs produce all components used in report. • Paper cites 7 or more references within 5 years of publication in correct format. • Introduction is engaging, lays out main points and structure of paper clearly. • Clear review of relevant algorithms.	• Well-organized, evaluate and compare performance of each algorithm using appropriate tools and visuals. • Submit all required components of report set, • Programs produce all components used in report. • Paper cites 10 or more references within 5 years of publication in correct format. • Introduction is engaging, lays out main points and structure of paper clearly. • Clear review of relevant algorithms.

Rated aspect	Assessment criteria				
	Very Poor	Poor	Satisfactory	Good	Excellent
	< 45	45-54	55-69	70-84	(Score ≥ 85)
	• incomplete set automatically • result in failing grade.		methodology and data. • No grammatical and typing errors.	• Conclusion addresses the thesis statement. • Logical and coherent description of methodology and data. • No grammatical and typing errors.	• Conclusion addresses the thesis statement. • Logical and coherent description of methodology and data. • No grammatical and typing errors.

References:	Created by:	Approved by:
	on behalf of the Lecturer Team (Samuel Ady Sanjaya S.T., M.T.) Course Coordinator	 (Ririn Ikana Desanti, S.Kom, M.Kom) Head of Study Program