 <p style="text-align: center;"><b>አዲስ አበባ ሳይንስና ቴክኖሎጂ ዩኒቨርሲቲ</b> <b>ADDIS ABABA SCIENCE &amp; TECHNOLOGY UNIVERSITY</b></p>	Document No. <b>VPAA/DPT/OF/032</b>	
	<b>FINAL EXAMINATION</b>	Issue No. <b>1</b>

[ **Software Engineering Department, College of Engineering** ]

**Final Examination for the Course Business Intelligence**

**Examination Date: 15 June, 2026**

**Academic Year: 2025/26**

**Maximum score: 50%**

**Semester: II**

**Time Allowed: 2 hours**

**Target Group: 1<sup>st</sup> Year**

**Tick an ‘X’ mark in the box where it applies you.**

Freshman	Applied Science	<input type="checkbox"/>	Engineering	<input type="checkbox"/>
Regular (2 <sup>nd</sup> & above):	Undergraduate	<input type="checkbox"/>	Postgraduate	<input type="checkbox"/>
Continuing Education	Undergraduate	<input type="checkbox"/>	Postgraduate	<input type="checkbox"/>

**Course instructor(s) name(s):**  Dr. Tilahun Melak


**Student Name:** \_\_\_\_\_ **ID. No:** \_\_\_\_\_

**General Directions**

- Write your name (including your grandfather’s name) and ID No. both on the question paper and answer sheet (if any).
- The examination booklet constitutes two types of questions, **25 Multiple Choice, 7 essay** questions.
- Make sure that all the parts are included in your booklet, which has **6 page(s)** including the cover page and answer sheet page(s), if any.
- Please read the instructions carefully and attempt all the questions.
- It is strictly forbidden to bring mobile phones and any electronic devices into the examination room
- Do not use pencil and red pen.
- Do not detach the answer sheet from the question paper.
- Cheating automatically nullifies your total scores.

**For Instructor’s Use Only**

Question Type/ No.	Mark Allotted	Mark Scored
Multiple Choice (1 – 25)	25 pts	
Work out/Essay	25pts	
<b>Total</b>	<b>50 pts</b>	

 <p style="text-align: center;"><b>አዲስ አበባ ሳይንስና ቴክኖሎጂ ዩኒቨርሲቲ</b> <b>ADDIS ABABA SCIENCE &amp; TECHNOLOGY UNIVERSITY</b></p>	<p>Document No. <b>VPAA/DPT/OF/032</b></p>	
	<p style="text-align: center;"><b>FINAL EXAMINATION</b></p>	<p>Issue No. <b>1</b></p>

**PART I: MULTIPLE CHOICE (25%)**

**Instruction:** For questions 1 to 25, write the letter of the correct answer on the answer sheet provided. Don't mark or circle the answer on the question paper. (1 points each)

1. Which one of the following best describes Business Intelligence (BI)?
  - A. A collection of programming languages used for data mining
  - B. A set of technologies and processes used to transform data into actionable information for decision-making
  - C. A database management system for storing transactional data
  - D. A statistical technique used for forecasting

**Answer: B**

2. The primary objective of Business Intelligence is to:
  - A. Replace human decision-makers
  - B. Increase data storage capacity
  - C. Support strategic, tactical, and operational decision-making
  - D. Eliminate uncertainty from business environments

**Answer: C**

3. Which type of data is estimated to constitute the largest proportion of organizational data today?
  - A. Structured Data
  - B. Semi-Structured Data
  - C. Unstructured Data
  - D. Transactional Data

**Answer: C**


4. Which type of analytics focuses on understanding what has happened in the past?
  - A. Predictive Analytics
  - B. Prescriptive Analytics
  - C. Descriptive Analytics
  - D. Diagnostic Analytics

**Answer: C**

5. Predictive analytics is primarily concerned with:
  - A. Summarizing historical events
  - B. Explaining past performance
  - C. Forecasting future outcomes
  - D. Automating transactions

**Answer: C**

6. Which one of the following is used to generate descriptive summary of the data to have general property of the data and understand which data values should be considered as noise or outliers.
  - A. measure of central tendency of the data
  - B. measure of dispersion of the data

 <p style="text-align: center;"><b>አዲስ አበባ ሳይንስና ቴክኖሎጂ ዩኒቨርሲቲ</b></p> <p style="text-align: center;"><b>ADDIS ABABA SCIENCE &amp; TECHNOLOGY UNIVERSITY</b></p>	Document No. <b>VPAA/DPT/OF/032</b>	
	<b>FINAL EXAMINATION</b>	Issue No. <b>1</b>

- C. Visualizations
- D. All
- E. None

**Answer: D**

7. **Exploratory Data Analysis** or EDA is a critical first step in analyzing a new dataset. Which one of the following **cannot** be the primary objective of EDA
- A. To enable us to direct specific testing of the hypothesis.
  - B. To detect outliers
  - C. To detect anomalies
  - D. To identifying patterns and visualizations
  - E. None

**Answer: E**

8. Which one of the following statements is **not** correct about the data cleaning?
- A. It is a kind of pre-process in which the given set of data is prepared for the data warehouse.
  - B. It refers to the transformation of wrong data into correct data
  - C. It refers to correcting inconsistent data
  - D. All of the above
  - E. None

**Answer: E**

9. Which of the following also used as the first step in the data preprocessing?
- A. Data selection
  - B. Data cleaning
  - C. Data transformation
  - D. Data integration

**Answer: B**

10. Which one of the following missing value handling technique is usually applied when class label is missing
- A. Fill in the missing value manually
  - B. Use a global constant to fill in the missing value
  - C. Ignore the tuple
  - D. Using the attribute mean

**Answer: C**

11. Which one of the following is **not** true about data transformation through attribute construction
- A. Attribute construction can be important to improve performance
  - B. Enable to discover missing information or information hidden within the dataset
  - C. It is the process of driving new attributes from the existing attributes.
  - D. All
  - E. None

**Answer: E**



12. Which one of the following **cannot** be used for data reduction?

- A. Log-linear model
- B. Histograms
- C. Clustering
- D. Multiple regression
- E. None

**Answer: E**

13. Which one of the following is True about frequent pattern and association rule mining

- A. If an itemset is frequent, each of its subsets is frequent as well.
- B. An itemset X is closed itemset in a data set S if there exists proper super-itemset Y such that Y has the same support count as X in S.
- C. An itemset X is maximal frequent itemset (or max-itemset) in set S if X is frequent, and there exists super-itemset Y such that  $X \subset Y$  and Y is frequent in S.
- D. all closed frequent itemset is member of maximal frequent item set
- E. None

**Answer: A**

14. The Apriori algorithm uses prior knowledge of subset support properties. Which one of the following is true about these properties

- A. All nonempty subsets of a frequent itemset must also be frequent.
- B. The support of any nonempty subset  $s'$  of itemset  $s$  must have at least the same support count as  $s$ .
- C. Given frequent itemset  $l$  and subset  $s$  of  $l$ , the confidence of the rule " $s' \Rightarrow (l-s)$ " cannot be more than the confidence of " $s \Rightarrow (l-s)$ ", where  $s'$  is a subset of  $s$ .
- D. All
- E. None

**Answer: D**

15. If we don't have predefined classes in the dataset for training our data mining model which data mining technique is more appropriate

- A. Association rule mining
- B. Classification and prediction
- C. Clustering
- D. All
- E. None

**Answer: C**

16. Concurrent access to shared data may result in\_\_\_\_\_

- A. Data consistency
- B. Data insecurity
- C. Data inconsistency
- D. None

**Answer: C**

17. Which one of the following is not true about big data

- A. Traditional RDBMS queries are not sufficient to gain useful information out of huge volume of data



FINAL EXAMINATION

- B. Searching with traditional tools to find out if a particular topic is trending can be timely.
- C. It is difficult to process big data with on hand database management tools.
- D. All
- E. None

**Answer: B**

18. A flight generates 240 terabytes of data in a flight of 6 to 8 hours. This data needs to be analyzed to make decisions and guide actions. Which one of the big data characteristics more appropriately describes the stated challenge?
- A. Volume
  - B. Variety
  - C. Velocity
  - D. All
  - E. None

**Answer: C**

19. One of the big data characteristics is variability where the meaning of data changes over time. Which one of the following cannot be a reason for the meaning of data change over time?
- A. Language evolution
  - B. Changes in characteristics of the data source
  - C. Data availability
  - D. Sampling process
  - E. None

**Answer: E**

20. Which model is the current model for big data
- A. Few companies are generating data and all others are consuming data
  - B. All of us are generating data and all of us are consuming data
  - C. All companies are generating data and few others are consuming data
  - D. All of us are generating data and few of us are consuming data

**Answer: B**

21. What are the main components of Hadoop Ecosystem?
- A. MapReduce, HDFS, YARN
  - B. Gelly, Table, CEP
  - C. MLlib, GraphX ,
  - D. None

**Answer: A**


22. \_\_\_\_\_ is a resource management platform responsible for managing compute resources in the cluster and using them in order to schedule users and applications.
- A. Hadoop Common
  - B. Hadoop YARN
  - C. Hadoop Distributed File System (HDFS)
  - D. Hadoop MapReduce

**Answer: B**

23. In spark, a \_\_\_\_\_ is a read-only collection of objects partitioned across a set of machines that can be rebuilt if a partition is lost.
- A. Spark Streaming
  - B. FlatMap
  - C. Resilient Distributed Dataset (RDD)
  - D. Driver

**Answer: C**

24. Which one of the following is a Transformation?
- A. collect()
  - B. reduce()
  - C. count()
  - D. map()

 <p style="text-align: center;"><b>አዲስ አበባ ሳይንስና ቴክኖሎጂ ዩኒቨርሲቲ</b> <b>ADDIS ABABA SCIENCE &amp; TECHNOLOGY UNIVERSITY</b></p>	Document No. <b>VPAA/DPT/OF/032</b>	
	Issue No. <b>1</b>	Page No. <b>Page 6 of 6</b>
<b>FINAL EXAMINATION</b>		

**Answer: D**

25. The primary objective of data protection is to:
- Increase data redundancy
  - Safeguard data from unauthorized access, alteration, or loss
  - Reduce database size
  - Improve network bandwidth

**Answer: B**

### **PART II: ESSAY (25%)**

*Write appropriate answers to the following questions*

- Define Business Intelligence and explain its major components. (4pts)

**Answer:**

Business Intelligence (BI) refers to a set of technologies, processes, and tools used to collect, integrate, analyze, and visualize business data in order to support better strategic and operational decision making.

The major components of Business intelligent Comprises of;

Data Sources → ETL Process → Data Warehouse → Analytics / OLAP → Dashboards & Reports → Decision Making

- Describe how BI contributes to organizational competitiveness and strategic planning. (3pts)

**Answer:**

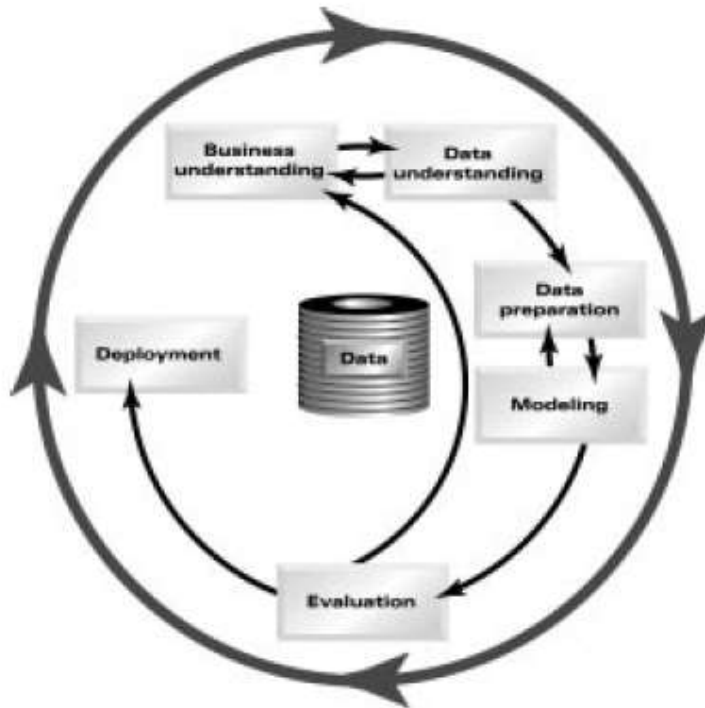
Business Intelligence (BI) provides organizations with **data-driven insights** that improve efficiency, reduce uncertainty, and give a competitive advantage.

- Supports a **data-driven culture**
- Enhances **operational efficiency**
- Reduces **decision-making uncertainty**
- Provides **competitive advantage**

- The steps involved in data mining when viewed as a process of knowledge discovery from a huge collection of data source are usually referred to as DM models. CRISP-DM is one of the most common datamining model. Discuss the detail steps in CRISP-DM model. (5pts)



Answer:



4. Explain the difference between the following concepts/techniques/terms (4pts)  
A. Supervised and Unsupervised learning


Answer:

Supervised learning (classification)

- Supervision: The training data (observations, measurements, etc.) are accompanied by **labels** indicating the class of the observations
- New data is classified based on the training set

Unsupervised learning (clustering)

- The class labels of training data is unknown
- Given a set of measurements, observations, etc. with the aim of establishing the existence of classes or clusters in the data

 <p style="text-align: center;">አዲስ አበባ ሳይንስና ቴክኖሎጂ ዩኒቨርሲቲ</p> <p style="text-align: center;">ADDIS ABABA SCIENCE &amp; TECHNOLOGY UNIVERSITY</p>	Document No. <b>VPAA/DPT/OF/032</b>	
	Issue No. <b>1</b>	Page No. <b>Page 8 of 6</b>
<b>FINAL EXAMINATION</b>		

B. Classification and Prediction

**Answer:**

Classification

- predicts categorical class labels (discrete or nominal)
- classifies data (constructs a model) based on the training set and the values (class labels) in a classifying attribute and uses it in classifying new data

Numeric Prediction

- models continuous-valued functions, i.e., predicts unknown or missing values

5. There are a lot challenges in big data. Discuss the challenges of big data in detail with examples. (4pts)

**Answer:**

**Volume** – Massive amounts of data require large storage and processing capacity.

*Example:* Social media platforms generate terabytes of data daily.

**Velocity** – Data is generated rapidly and often requires real-time processing.

*Example:* Credit card fraud detection systems.

**Variety** – Data comes in different formats (structured, semi-structured, unstructured).

*Example:* Text, images, videos, and sensor data.

**Veracity** – Ensuring data quality, accuracy, and reliability is difficult.

*Example:* Fake or incomplete social media data.

**Storage and Processing** – Traditional systems cannot efficiently handle huge datasets.


*Example:* Large-scale video streaming services.

**Security and Privacy** – Protecting sensitive information from unauthorized access.

*Example:* Patient medical records in healthcare systems.

**Data Integration** – Combining data from multiple sources is complex.

*Example:* Merging online and offline customer data.

 <p style="text-align: center;"><b>አዲስ አበባ ሳይንስና ቴክኖሎጂ ዩኒቨርሲቲ</b></p> <p style="text-align: center;"><b>ADDIS ABABA SCIENCE &amp; TECHNOLOGY UNIVERSITY</b></p>	Document No. <b>VPAA/DPT/OF/032</b>	
	<b>FINAL EXAMINATION</b>	Issue No. <b>1</b>

**Scalability** – Systems must grow as data volume increases.

*Example:* Expanding cloud infrastructure for growing businesses.

**Visualization** – Presenting large datasets in an understandable format.

*Example:* Business intelligence dashboards.

**Skill Gap** – Shortage of professionals with big data expertise.

*Example:* Difficulty finding skilled data scientists and engineers.

6. Discuss at least two reasons of how spark is better than Apache Hadoop. (2pts)

**Answer:**

In contrast to Hadoop's two-stage, disk-based MapReduce paradigm, Spark's multi-stage in-memory primitives provide performance improvements of up to 100 times faster for certain applications.

Spark allows user programs to load data into a cluster's memory and query it repeatedly. Spark is particularly well suited for machine learning applications, which often involve iterative computations and repeated in-memory processing.

7. Critically examine the importance of data privacy and data protection in modern Business Intelligence systems. Discuss the challenges organizations face in balancing data utilization and privacy preservation. (3pts)

**Answer:**

Data privacy and protection are essential in Business Intelligence (BI) to safeguard sensitive information, maintain customer trust, prevent data breaches, and ensure compliance with legal regulations.

**Challenges**

- **Data Access vs. Privacy:** BI needs large amounts of data, but excessive access can expose sensitive information.
- **Regulatory Compliance:** Privacy laws may restrict data collection and usage.
- **Security Threats:** Increased risk of hacking, data leaks, and unauthorized access.
- **Data Sharing:** Sharing data across departments can compromise privacy.
- **Cost and Complexity:** Implementing security and privacy measures requires resources and expertise.