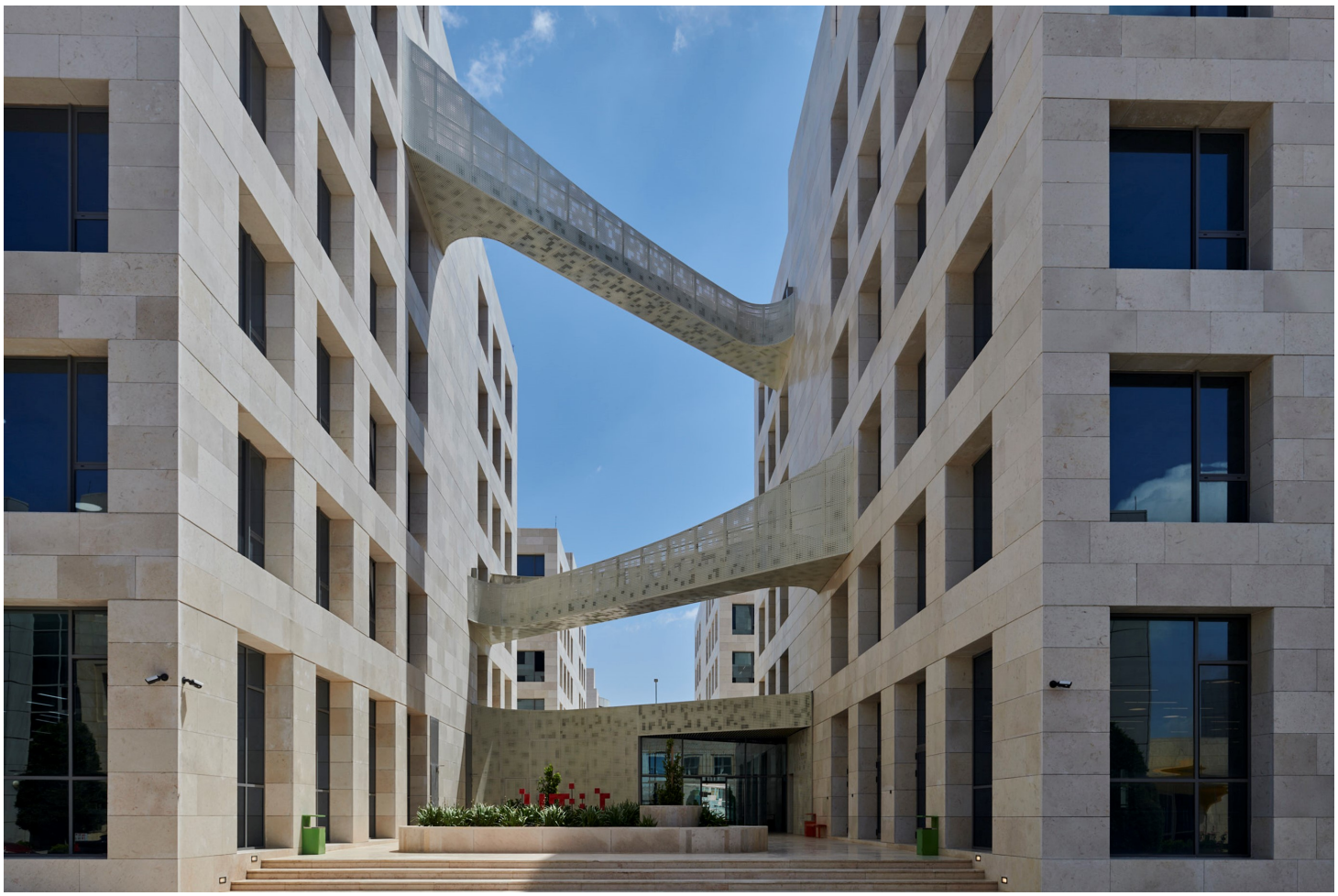
جامعة الحسين التقنية
Al Hussein Technical University

ASSIGNMENT BRIEF

HTU Course No: 30202232	**HTU Course Name:** Data Mining
BTEC Unit No: H/615/1653	**BTEC UNIT Name:** Data Mining

Version: 3



Student Name/ID Number/Section	
HTU Course Number and Title	30202232 Data Mining
BTEC Unit Number and Title	H/615/1653 Data Mining
Academic Year	2023-2024 Fall
Assignment Author	Raneem Qaddoura
Course Tutor	Raneem Qaddoura - Bassam Al-Kasasbeh
Assignment Title	Analyze and Implement text and graph mining application
Assignment Ref No	1
Issue Date	26/11/2023
Formative Assessment dates	From 26/11/2023 to 14/12/2023
Submission Date	23/12/2023
IV Name & Date	Aisha Al-Sadi 25/11/2023

Submission Format

The assignment should be submitted to the university's eLearning system within the deadline specified above from the link: https://elearning.htu.edu.jo. The assignment is in the form of:

1. A PowerPoint slide in .pptx format (**PPT Document**).
   a. You must ensure consistency and readability: Use the same style, font, and color scheme throughout.
   b. You must effectively use headings, bullet points, and subsections as appropriate. Your research should be referenced using the **Harvard referencing system**.
   c. You must use visuals: Include simple, clean graphs and charts to present data effectively.
2. Code file(s).

Signed declaration Form (**Word Document**).

Unit Learning Outcomes

LO1 Discuss the historical and theoretical foundation of data mining, its scope, techniques, and processes.

LO3 Illustrate how a data mining algorithm performs text mining to identify relationships within text.

LO4 Evaluate a range of graph data mining techniques that recognise patterns and relationships in graph-based technologies.

Assignment Brief and Guidance

You have a position as a Junior Data Mining Specialist for a telecom company with systems that the IT department implements. These systems have generated massive amounts of data, and the company wants to extend its work to include data science applications.

The company wants to know the possibilities of applying data mining techniques, including text and graph mining, to its applications. It wants to start with the complaint data provided in this dataset. The company also wants to evaluate your ability to do the work. You are assigned to implement a complete text and graph mining application by allowing the employees in the company to search for specific keywords in the complaints and represent your results using a graph. You then need to communicate your work to the management by producing a comprehensive presentation to show some basic technical concepts and the application obtained from your work.

Part I (Text and Graph Mining Application):

Your company asked you to generate a complete text mining and graph mining application that must include the following:

- Show how text mining works using a programming language for the employees in the company to allow them to search for specific keywords from the complaints and show at least five most relevant complaints to the query entered by the employee.
- Develop a complete text mining application for this real world issue by showing the results using a graphical user interface.
- Demonstrate how graph data mining works using a programming language by showing a graph for the results generated by the query to show how each word is connected by prefix and suffix. Show the centrality measures for the words in the graph.
- Develop a complete graph mining application for this real world issue by generating a graph that shows the co-occurrence of the different words in at least five complaints. You need to show the weights on the edges of the graph that represent the number of co-occurrences for each two words. Show both graphs on the graphical user interface.

Part II (Presentation)

Your presentation must include the following:

- Investigate the historical background of data mining.
- Analyse the theoretical background of data mining.
- Identify the data mining tools used in the industry.
- Select an organization from the industry and meet with a representative to understand the traditional and modern approaches they have used in the organization. You should be able to do the following after meeting with the representative:
    a. Evaluate the traditional and modern approaches to data mining used in the organization and show the building blocks of both approaches.
    b. Review how this organization benefits from data mining by meeting one of the employees.
- Discuss what is meant by text mining, explain it with examples, and show where it is applied in the real world.
- Analyse how text mining algorithms, techniques, methods and approaches work.
- Discuss what is meant by graph data mining and explain with appropriate examples.
- Assess how graph mining algorithms work.
- Identify appropriate programming languages and tools used by the industry for graph data mining.
- Show the text and graph mining application that you have implemented for a real world issue.

Learning Outcomes and Assessment Criteria			
Learning Outcome	**Pass**	**Merit**	**Distinction**
LO1 Discuss the historical and theoretical foundation of data mining, its scope, techniques, and processes.	**P1** Investigate the historical background of data mining. **P2** Analyse the theoretical background of data mining and identify data mining tools used in industry.	**M1** Evaluate traditional and modern approaches to data mining and show the building blocks of both approaches.	**D1** Review how an organisation benefits from data mining.
LO3 Illustrate how a data mining algorithm performs text mining to identify relationships within text.	**P5** Discuss what is meant by text mining and explain with appropriate examples. **P6** Analyse how data mining algorithms, techniques, methods and approaches work.	**M4** Show how text mining works using a tool or programming language	**D3** Develop a complete text mining application for a real world issue.
LO4 Evaluate a range of graph data mining techniques that recognise patterns and relationships in graph-based technologies.	**P7** Discuss what is meant by graph data mining and explain with appropriate examples. **P8** Assess how graph mining algorithms work and identify appropriate programming languages and tools used by industry for graph data mining.	**M5** Demonstrate how graph data mining works using a tool or programming language.	**D4** Develop a complete graph data mining application for a real world scenario