

Bachelor of Computer Science (Hons)

**Introduction to Data Science**

**XBDS2014 & XBDS2014N**

Prepared by Ts. Dr. Law Foong Li

Semester September 2022

A close up of text on a white background

Description automatically generated**ASSIGNMENT: Group Assignment**

Course Title : Bachelor of Computer Science (Hons)

Course Code : XBDS2014 & XBDS2014N

Course Lecturer : Ts. Dr. Law Foong Li

: fllaw@uow.edu.my

**BRIEF**

Machine learning is now an established approach to generalizing conclusions from a (large) dataset, where many characteristics (features) of the population of interest are considered simultaneously. The underlying assumption is that the dataset is a thorough representation of that population so that the space of features is widely and fairly explored. Though the approach and the very term “Machine learning” were coined in the computer science community, the field shares more than a common property with Statistics. In some cases, the two disciplines talk about the same thing, just with a twist of words: what machine learning people call a dataset would be called a sample by most statisticians. Moreover, the aim of machine learning is essentially the same as Statistics: to gain information about the population through the available data. And use those data to perform either unsupervised or supervised learning tasks.

You are encouraged to produce an application that focuses on novel ways of using statistical methods in any of the activities involved in a machine learning task.

Topics of particular interest include, but are not limited to:

1. Data cleaning and data pre-processing
2. Time series analysis
3. Association analysis
4. Clustering and classification
5. Regression analysis

**REQUIREMENTS**

* If you do not attend the walkthrough the maximum mark you can achieve for this assignment is **40%**.
* You will need to bring one hard copy of your product to provide the walkthrough. The walkthrough will take place in **week 6** and **week 12**.
* Please do not submit hand-drawn diagrams. Hand-drawn diagrams or handwritten reports will receive **zero (0) marks**.
* Your submission documentation’s content should include the following items:
  + Report
  + Teamwork Evaluation Form
  + Assessment rubric

**ASSESSMENT CRITERIA**

**Report: 20%**

1. **Methodology – 30% of marks**

It must be a complete, clear, unambiguous English description of the analysis you performed. This should be sufficient for someone else to write a Python or R program (or perform manual computations) that reproduces your results, without access to your source code, and without having to guess or make significant design choices.

1. **Implementation – 30% of marks**

The product you produce for this assignment must be produced in Python or R.

It should be built to a good standard and fit for the intended purpose. Code should be well laid out and commented, and should make use of appropriate techniques.

1. **Results & Discussion – 40% of marks**

You explain and discuss the outcome of your implementation using descriptive statistics and visualizations as a result of the methodology [or methodologies] you applied. The results should simply state the findings, without bias and be arranged in a logical sequence.

**Communication Skills: 10%**

1. **Presentation – 50% of marks**

The presentation defines as the ability to engage effectively in verbal, non-verbal, written, and/or symbolic communication. You should have effective communication of the product demonstration. Your explanation and discussion of the work carried out, handling of questions, and responses to markers.

1. **Teamwork - 50% of marks**

Teamwork assesses the effectiveness of a group to work as a team, not as an assessment of an individual student’s ability to work on a team.

**DUE DATE:** 29 November 2022

**ASSESSMENT RUBRIC**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **CRITERIA** | **MARKS** | | | | | | |
| **Methodology (30%)** | **16-20** | **13-15** | **10-12** | | **8-9** | **0-7** | **Comments** |
| Excellent in documenting the methodology.  • Generates complete, clear and unambiguous requirements specification.  • Identifies ambiguity in givens and states necessary assumptions.  • Uses appropriate diagrams to describe implementation clearly including the design decisions.  Overall contents comprehensively articulates all relevant and pertinent issues related to the overall solution.  All areas are at least good. May be outstanding is some areas and good in others and hence is on balance excellent. Good or above in all areas. Likely to contain minor errors, omissions or additions which prevent the methodology from being outstanding. Overall an excellent methodology. | Good in documenting the methodology.  • Generates requirements specification with minor residual ambiguity.  • Identifies ambiguity in givens however necessary assumptions are not fully stated.  • Uses appropriate diagrams to describe implementation however contain a small number of errors, omissions or additions.  It is possible that the methodology is outstanding or excellent in some areas and satisfactory in others but on balance is good. Satisfactory or above in all areas. Likely to contain a small number of errors, omissions or additions which prevent the methodology from being excellent. Overall a good methodology. | | Satisfactory in documenting the methodology.  • Generates requirements specification with some residual ambiguity.  • Omits ambiguity in givens and states necessary assumptions ambiguously.  • Uses appropriate diagrams to describe implementation however contain a number of errors, omissions or additions.  It is possible that the methodology is good or above in some areas and satisfactory in others. Weak in no more than two areas. Likely to contain a number of errors, omissions or additions which prevent the methodology from being good. Overall satisfactory. | Weak in documenting the methodology.  • Generates requirements specification with substantial ambiguity.  • Omits ambiguity in given and necessary assumptions.  • Uses inappropriate diagrams to describe implementation and contain large number of errors, omissions or additions.  It is possible that the methodology is satisfactory or above in some areas and unsatisfactory in others. Likely to be weak in more than three areas. It might be unsatisfactory in one area but no more. Likely to contain errors, omissions, additions, or misunderstandings which prevent the methodology from being satisfactory. Overall poor. | Unsatisfactory in documenting the methodology.  • Generates requirements specification with substantial ambiguity.  • Omits ambiguity in given and necessary assumptions.  • No diagrams to describe software architecture.  It is possible that the methodology is weak or above in some areas and unsatisfactory in others. Unsatisfactory in two or more areas. Likely to contain errors, omissions, additions, or misunderstandings which prevent the design model from being weak. May not be recognisable as a methodology, might have majors errors in content or a combination of the two. |  |
|  | **16-20** | **13-15** | | **10-12** | **8-9** | **0-7** |  |
| **Implementation (30%)** | Excellent in the implementation, comments, indentation, consistency and syntax. Correct following of object-oriented concepts. The work shows particular insight or originality in its approach. Excellent functionalities which identifies the underlying principles behind the problem. All areas are at least good. May be outstanding is some areas and good in others and hence is on balance excellent. Good or above in all areas. Likely to contain minor errors, omissions or additions which prevent the implementation from being outstanding. Overall an excellent implementation. | Good in the implementation, comments, indentation, consistency and syntax. It is possible that the application is outstanding or excellent in some areas and satisfactory in others but on balance is good. Satisfactory or above in all areas. Likely to contain a small number of errors, omissions or additions which prevent the implementation from being excellent. Overall a good implementation. | | Satisfactory in the implementation, comments, indentation, consistency and syntax. It is possible that the application is good or above in some areas and satisfactory in others. Likely to be weak in no more than two areas. Likely to contain a number of errors, omissions or additions which prevent the implementation from being good. Overall a satisfactory implementation. | Weak in the implementation, comments, indentation, consistency and syntax. It is possible that the application is satisfactory or above in some areas and unsatisfactory in others. Likely to be weak in more than three areas. It might be unsatisfactory in one area but no more. Likely to contain errors, omissions, additions, or misunderstandings which prevent the implementation from being satisfactory. Still recognisable as an object-oriented application of the problem in focus. Overall poor but satisfactory. | Unsatisfactory in the implementation, comments, indentation, consistency and syntax. It is possible that the application is weak or above in some areas and unsatisfactory in others. Unsatisfactory in two or more areas. Likely to contain errors, omissions, additions, or misunderstandings which prevent the implementation from being weak. May not be recognisable as an object-oriented application, might have majors errors in implementation or a combination of the two. |  |
| **Results & Discussion (40%)** | **16-20** | **13-15** | | **10-12** | **8-9** | **0-7** |  |
|  | Excellent in the results discussion. The explanation and justification of how it meets specified requirements shows outstanding insight into the issues involved and alternatives available. The presentation clearly and concisely demonstrates a deep understanding of the project implementation. May be outstanding is some areas and good in others and hence is on balance excellent. Good or above in all areas. Likely to contain minor errors, omissions or additions which prevent the results discussion from being outstanding. Overall an excellent results discussion. | Good in the results discussion. The explanation and justification of how it meets specified requirements shows good understanding into the issues involved and alternatives available. The presentation demonstrates a good understanding of the project implementation. It is possible that the presentation is outstanding or excellent in some areas and satisfactory in others but on balance is good. Satisfactory or above in all areas. Likely to contain a small number of errors, omissions or additions which prevent the results discussion from being excellent. Overall a good results discussion. | | Satisfactory in the results discussion. The explanation and justification of how it meets specified requirements covers relevant aspects into the issues involved and alternatives available, but is not outstanding in any respect. It is possible that the presentation is good or above in some areas and satisfactory in others. Likely to be weak in no more than two areas. Likely to contain a number of errors, omissions or additions which prevent the results discussion from being good. Overall a satisfactory results discussion. | Weak in the areas of the results discussion. The explanation and justification of how it meets specified required is inadequate. It is possible that the presentation is satisfactory or above in some areas and unsatisfactory in others. Likely to be weak in more than three areas. It might be unsatisfactory in one area but no more. Likely to contain errors, omissions, additions, or misunderstandings which prevent the results discussion from being satisfactory. Overall poor but satisfactory. | Unsatisfactory in the areas of the results discussion. It conveys little understanding of solution of the problem or no useful explanation and justification of how it meets specified requirements. It is possible that the presentation is weak or above in some areas and unsatisfactory in others. Unsatisfactory in two or more areas. Likely to contain errors, omissions, additions, or misunderstandings which prevent the results discussion from being weak. |  |
| **Presentation (50%)** | **16-20** | **13-15** | | **10-12** | **8-9** | **0-7** | **Comments** |
|  | Excellent, well-directed presentation, logically and coherently structured. It is free or almost free grammatical errors. The format is clear and consistent with appropriate use of headings and paragraphs. English usage is easily understandable. References and quotations are utilized appropriately to indicate sources.  Excellent in the presentation and handling all questions. The explanation and justification of how it meets specified requirements shows outstanding insight into the issues involved and alternatives available. The presentation clearly and concisely demonstrates a deep understanding of the project implementation. May be outstanding is some areas and good in others and hence is on balance excellent. Good or above in all areas. Likely to contain minor errors, omissions or additions which prevent the presentation from being outstanding. Overall an excellent presentation. | Good presentation, logically structured. There are occasional spelling and grammatical errors, but the reader does not struggle to interpret the writer’s intended meaning. The writing would benefit from the use of organizational tools (e.g. headings, paragraphs) and more consistent use of references to sources.  Good in the presentation and handling all questions with some minor inaccurate answers. The explanation and justification of how it meets specified requirements shows good understanding into the issues involved and alternatives available. The presentation demonstrates a good understanding of the project implementation. It is possible that the presentation is outstanding or excellent in some areas and satisfactory in others but on balance is good. Satisfactory or above in all areas. Likely to contain a small number of errors, omissions or additions which prevent the presentation from being excellent. Overall a good presentation. | | Satisfactory presentation, well structured. It contains number of spelling and grammatical errors, but the reader does not struggle to interpret the writer’s intended meaning. It is possible that the use of organizational tools (e.g. headings and paragraphs) are good or above in some areas and satisfactory in others.  Satisfactory in the presentation and handling most of the questions with some inaccurate answers. The explanation and justification of how it meets specified requirements covers relevant aspects into the issues involved and alternatives available, but is not outstanding in any respect. It is possible that the presentation is good or above in some areas and satisfactory in others. Likely to be weak in no more than two areas. Likely to contain a number of errors, omissions or additions which prevent the presentation from being good. Overall a satisfactory presentation. | Weak presentation and structure. Spelling and grammatical errors force the reader to struggle to determine the intended meaning. Organizational tools such as headings, paragraphs are used inconsistently. References are not used properly to indicate the sources of material.  Weak in the areas of the presentation and handling some of the questions but most answers are inaccurate. The explanation and justification of how it meets specified required is inadequate. It is possible that the presentation is satisfactory or above in some areas and unsatisfactory in others. Likely to be weak in more than three areas. It might be unsatisfactory in one area but no more. Likely to contain errors, omissions, additions, or misunderstandings which prevent the presentation from being satisfactory. Overall poor but satisfactory. | Unsatisfactory presentation and structure. Numerous spelling and grammatical errors and a lack of clear consistent organization interfere with the writer’s ability to communicate to the key points. The reader frequently cannot determine the intended meaning. There are no references to indicate material taken from other sources.  Unsatisfactory in the areas of the presentation and handling all questions and not able to answer any questions posed. It conveys little understanding of solution of the problem or no useful explanation and justification of how it meets specified requirements. It is possible that the presentation is weak or above in some areas and unsatisfactory in others. Unsatisfactory in two or more areas. Likely to contain errors, omissions, additions, or misunderstandings which prevent the presentation from being weak. |  |