

**SCHOOL OF COMPUTING AND INFORMATION TECHNOLOGY**

**BACHELOR OF SOFTWARE ENGINEEERING**

**BACHELOR OF COMPUTER SCIENCE**

**GROUP ASSIGNMENT GUIDELINE**

## MODULE TITLE : Statistics and Operational Research / Statistical

**Inference and Modeling**

## MODULE CODE : MTH60904/ITS66904

## WEIGHTAGE : 30%

## DATE OF COMMENCING : 20/08/2023 (08:00AM)

## END DATE : 20/09/2023 (12:00PM)

**Note:**

**\* PLAGIARISM IS A SERIOUS OFFENCE AND PLAGIARIZED WORK WILL RESULT IN AN F GRADE. 0 mark and a barring from sitting final examination will be implemented for those who does not submit any assignments.**

**ASSIGNMENT DETAILS**

## A: Objective

The objective of this assignment is to use the Statistical techniques learnt to analyse a data set and present the output in an interesting critical manner using Python or R-studio using a downloaded dataset.

## B: Instruction

1. This is a group assignment (3-5 members).

1. You will need to do this assignment on a computer with Python or R-studio.

## C: General Guidelines

The idea of the group project is to give you some experience trying to do a piece of original research and writing up your results in a paper style format. What we expect to see is an idea/task that you describe clearly, relate to existing work, implement and test on a dataset. To do this you will need to write code, run it on some data, make some figures, read a few background papers, collect some references, and write a few pages describing your task, the method(s) you used and the results you obtained.

The grade will depend on the ideas, how well you present them in the report, how clearly you position your work relative to existing literature, how illuminating your experiments are, and wellsupported your conclusions are.

The idea is that this project report should be a manageable amount of work, but that if you want to turn your project into a paper, everything in the project report will need to be done anyway. If you feel that your project won’t fit into this rubric, please talk to your instructor. There are many ways to make contributions to a field!

Your project should consist of an implementation of one or more statistical methods, and their application to a dataset. Your project may be a comparison of several existing techniques which should then be compared to at least one machine learning. As a under graduate student, you are free to pick a project of your own design.

You are free to use any third-party ideas or code that you wish as long as it is publicly available. You must properly provide references to any work that is not your own in the write-up. The project is not intended to be a stressful exercise; instead it is a chance for you to experiment, to think, to play and to hopefully have fun! Start with simple methods that work more or less out of the box and go from there.

## D. Specific Requirements

Length: 10 to 15 pages, not including appendices. Don’t be afraid to keep the text short and to the point, and to include large illustrative figures.

1. Abstract : a summary of the main idea of the project and its contributions.
   1. Should be understandable to anyone in the course.
   2. You don’t need to say everything you did, just say what the main idea was and what were one or two takeaways.
2. Introduction: a statement of the problem being addressed and why we might want to solve it.
3. Related work :A section describing related works and the bibliography.
   1. If your project builds on previous work, clearly distinguish what they did from what your new contribution is.
   2. Include a 1-2 sentence summary of other closely related papers. You might not know about all related papers (or have time to carefully read all related papers), and that’s OK for this project. A rough guide is that you should be able to find 3-4 closely related papers, and another 3-4 papers that all those papers cite as foundational work.

These foundational papers are often cited in the introduction.

1. Data
   1. Provide brief description of the data and its context
   2. Include the statistical analysis
   3. Presentation, visualization as well as quantification of the data, images etc d. Source of the data
2. Methods
   1. Explanatory data analysis (EDA)
   2. Statistical Methods used to analyse the data and its brief description

1. Results
   1. Results of the EDA
   2. Results of the Statistical Analysis
   3. Discuss the results obtained
   4. Compare the results with other models or techniques
   5. Compare the results with other studies
2. Limitations and future study :A section describing the limitations of your approach.
   1. Describe some settings in which we’d expect your approach to perform poorly, or where all existing models fail.
   2. Try to guess or explain why these limitations are the way they are.
   3. Give some examples of possible extensions, ways to address these limitations, or open problems.
3. Conclusions :A section describing your conclusions and ideas for future work.
   1. Write a sentence about the topic
   2. State the results achieved in relation to the problem described in the introduction.
   3. Repeat the main takeaways from your paper.

## E: Format of the Written Report

1. You are required to use the R-studio or R-studio.

1. You are required to compile your report in the format below:

* 1. Students are required to submit the softcopy of the report (PDF form).

* 1. Students are required to submit softcopy (via ***submission link***) of the report.

* 1. Due Date: **November 2, 2023 before 12:00pm.**

***No late submission is accepted (except for strong valid reason and prior approval granted by the lecturer).***

## D: Assessment Marking Criteria

Content

1. Abstract : 5
2. Introduction : 10 :
3. Related Work : 20
4. Data : 5
5. Methods : 20
6. Results : 20
7. Limitation and Future Study : 10
8. Conclusion : 5
9. Format : 5

**Total : 100**

**Marking Rubric for Written Assignment (Group):**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No** | **Criteria** |  | |  | |  | |  | |  | |  |
| 1. | Abstract (5 Points) | **5 marks**  A clear and concise abstract that gives the reader a clear idea of what the project is about and why it is interesting. The following components need to be included   1. Purpose and   motivation of  this research   1. Problem you   are addressing   1. Methods and   materials   1. Results 2. Conclusion | | **4 marks**  A clear abstract that gives the reader a clear idea of what the project is about.  Four of the following components are included   1. Purpose and   motivation of  this research   1. Problem you   are addressing   1. Methods and   materials   1. Results 2. Conclusion | | **3 marks**  The abstract is difficult to read and/or is very vague and/or doesn’t sell the project as well as it might have. Three of the following components are included   1. Purpose and   motivation of  this research   1. Problem you   are addressing   1. Methods and   materials   1. Results 2. Conclusion | | **2 marks**  Unable to read the abstract and/or is very vague and/or doesn’t sell the project as well as it might have. Only two of the following components are included   1. Purpose and   motivation of  this research   1. Problem you   are addressing   1. Methods and   materials   1. Results 2. Conclusion | | **1 mark**  Unable to read the abstract. Only one of the following components is included   1. Purpose and   motivation of  this research   1. Problem you   are addressing   1. Methods and   materials   1. Results 2. Conclusion | |  |
| 2. | Introduction (10 Points) | **9-10 marks**  A readable write-up that explains what the problem is and why it is of interest. The following components need to be included   1. Problem 2. Negative impact of the problem iii. Parties   affected iv. Benefit of  solving the problem | | **7-8 marks**  A readable write-up that explains what the problem is. Three of the following components are included.   1. Problem 2. Negative impact of the problem iii. Parties   affected iv. Benefit of  solving the  problem | | **5-6 marks**  The write-up is difficult to read, somewhat vague, or doesn’t make a really good case for why the problem is of interest. Two of the following components are included.   1. Problem 2. Negative impact of the problem iii. Parties   affected iv. Benefit of  solving the  problem | | **3-4 marks**  Unable to read the write-up and/or is very vague. Only one of the following components are included.   1. Problem 2. Negative impact of the problem iii. Parties   affected iv. Benefit of  solving the  problem | | **1-2 marks**  Unable to read the write-up. None of the following components are included.   1. Problem 2. Negative impact of the problem iii. Parties   affected iv. Benefit of  solving the problem | |  |
| 3. | Related work | **19-20 marks** | | **16-18 marks** | | **13-15 marks** | | **9-12 marks** | | **5-8 marks** | | **1-4 marks** |
|  | (Literature  Review)  (20 points) | An outstanding overview, with an insightful analysis of prior work and a clear connection between prior work and the proposed method. The following components are given.   1. Introduction   of the topic   1. Taxonomy   Mapping iii. Paragraphs  for each  branch of the taxonomy tree   1. Conclusion 2. Critical   Review | | A comprehensive overview of prior work that gives the reader a clear idea of what’s out there and how the proposed method is different. Four of the following components are given.   1. Introduction   of the topic   1. Taxonomy   Mapping iii. Paragraphs  for each  branch of the taxonomy tree   1. Conclusion 2. Critical   Review | | A fairly good overview of prior work, and some connection is made to the proposed method. Three of the following components are given.   1. Introduction   of the tropic   1. Taxonomy   Mapping iii. Paragraphs  for each  branch of the  taxonomy tree   1. Conclusion 2. Critical   Review | | An overview of several papers related to the proposed method, and some attempt is made to connect the prior work to the current method. Two of the following components are given.   1. Introduction   of the topic   1. Taxonomy   Mapping iii. Paragraphs  for each  branch of the  taxonomy tree   1. Conclusion 2. Critical   Review | | within a coherent conceptual frame One of the following componen  i.  ii. iii.  iv.  v. | An overview of several related papers, but not - work.  ts are given. Introduction of the topic  Taxonomy  Mapping Paragraphs for each  branch of the  taxonomy tree Conclusion  Critical  Review | Bad attempt at describing prior work. None of the following components are given.  i.  Introduction of the topic  ii. Taxonomy  Mapping iii. Paragraphs  for each branch of  the taxonomy tree   1. Conclusion 2. Critical   Review |
| 4. | Data  (5 Points) | **5 marks**  The data are comprehensive and clearly described. At least 6 of the following components are given.   1. Source of the data 2. Description of   the data and its context   1. Statistics of the data iv. Presentation,   visualization and quantification of the data and images  v. Conclusion | | **4 marks**  The data are fairly explained. At least 5 of the following components are given.   1. Source of the data 2. Description of   the data and its context   1. Statistics of the data iv. Presentation,   visualization and quantification of the data and images  v. Conclusion | | **3 marks**  The data are not comprehensive and/or there is a flaw in the explanation. At least 4 of the following components are given.   1. Source of the data 2. Description of   the data and its context   1. Statistics of the data iv. Presentation,   visualization and quantification of the data and images  v. Conclusion | | **2 marks**  The explanations are significantly flawed. At least 3 of the following components are given.   1. Source of the data 2. Description of   the data and its context   1. Statistics of the data iv. Presentation,   visualization and quantification of the data and images  v. Conclusion | | **1 mark**  The explanations are flawed. At least 2 of the following components are given.   1. Source of the data 2. Description of   the data and its context   1. Statistics of the data iv. Presentation,   visualization and quantification of the data and images  v. Conclusion | |  |
| 5. | Methods  (20 Points) | **17-20 marks** | | **13-16 marks** | | **9-12 marks** | | **5-8 marks** | | **1-4 marks** | |  |
|  |  | The methods of analysis are comprehensive and clearly described. At least 6 of the following components are given.  i. Explanatory  data analysis ii. Statistical  data analysis methods iii. Appropriate  data analysis iv. Statistical  methods address the research  objective   1. Information   on data analysis  process   1. Clear   relationship between methods | | The methods of analysis are fairly explained. At least 5 of the following components are given.  i. Explanatory  data analysis ii. Statistical  data analysis methods iii. Appropriate  data analysis iv. Statistical  methods address the research  objective   1. Information   on data analysis  process   1. Clear   relationship between methods | | The methods of analysis are not comprehensive and/or there is a flaw in the explanation. At least 4 of the following components are given.  i. Explanatory  data analysis ii. Statistical  data analysis methods iii. Appropriate  data analysis iv. Statistical  methods address the research  objective   1. Information   on data analysis  process   1. Clear   relationship between methods | | The methods of analysis are significantly flawed. At least 3 of the following components are given.  i. Explanatory  data analysis ii. Statistical  data analysis methods iii. Appropriate  data analysis iv. Statistical  methods address the research  objective   1. Information   on data analysis  process   1. Clear   relationship between methods | | The methods of analysis are flawed. At least 2 of the following components are given.  i. Explanatory  data analysis ii. Statistical  data analysis methods iii. Appropriate  data analysis iv. Statistical  methods address the research  objective   1. Information   on data analysis  process   1. Clear   relationship between methods | |  |
| 6. | Results  (20 Points) | **17-20 marks**  The results are comprehensive and clearly described. At least 6 of the following components are given.   1. Subheadings   are included and are clear and informative   1. Figures and   tables are supported by  text iii. Correct interpretation of the results | | **13-16 marks**  The results are fairly explained. At least 5 of the following components are given.   1. Subheadings   are included and are clear and informative   1. Figures and   tables are supported by  text iii. Correct interpretation  of the results iv. Results with  tables and diagrams | | **9-12 marks**  The results are not comprehensive and/or there is a flaw in the explanation. At least 4 of the following components are given.   1. Subheadings   are included and are clear and informative   1. Figures and   tables are supported by  text iii. Correct interpretation of the results | | **5-8 marks** The results are significantly flawed. At least 3 of the following components are given.   1. Subheadings   are included and are clear and informative   1. Figures and   tables are supported by  text iii. Correct interpretation  of the results iv. Results with  tables and diagrams | | **1-4 marks**  The results are flawed. At least 2 of the following components are given.   1. Subheadings   are included and are clear and informative   1. Figures and   tables are supported by  text iii. Correct interpretation  of the results iv. Results with  tables and diagrams | |  |
|  |  | 1. Results with   tables and diagrams   1. Additional   insight to the  content vi. Critical analysis of  the results vii. Clearly addresses the research  question | | v. Additional  insight to the  content vi. Critical analysis of  the results vii. Clearly addresses the research  question | | 1. Results with   tables and diagrams   1. Additional   insight to the  content vi. Critical analysis of  the results vii. Clearly addresses the research  question | | v. Additional  insight to the  content vi. Critical analysis of  the results vii. Clearly addresses the research  question | | v. Additional  insight to the  content vi. Critical analysis of  the results vii. Clearly addresses the research  question | |  |
| 7. | Limitations and Future Study  (10 points) | **9-10 marks**  analysis.  i.  ii.  iii.  iv.  v. | An insightful and correct  The following  components are given. Discussion addresses the major finding of the study Results are interpreted with respect to outside sources Identify the limitation or limitations    Explain these limitations in detail  Propose a future direction for future studies | **7-8 marks**  One of he following  i.  ii.  iii.  iv.  v. | A correct analysis that could be more complete and is not very insightful.  components is missing. Discussion addresses the major finding of the study Results are interpreted with respect to outside sources Identify the limitation or  limitations    Explain these limitations in detail  Propose a future direction for future studies | **5-6 marks** An incomplete or somewhat incorrect analysis.  are missing.  i.  ii.  iii.  iv.  v. | Two of the  following components    Discussion addresses the major finding of the study Results are interpreted with respect to outside sources Identify the limitation or  limitations    Explain these limitations in detail  Propose a future direction for future studies | **3-4 marks**  An incorrect analysis. One of the following  i.  ii.  iii.  iv.  v. | components are given. Discussion addresses the major finding of the study Results are interpreted with respect to outside sources Identify the limitation or limitations    Explain these limitations in detail  Propose a future direction for future studies | **1-2 marks** No analysis  are given.  i.  ii.  iii.  iv.  v. | . None of the  following components  Discussion addresses the major finding of the study Results are interpreted with respect to outside sources Identify the limitation or limitations    Explain these limitations in detail  Propose a future direction for future studies |  |
| 8. | Conclusions (5 points) | **5 marks**  A clear and insightful summary of the paper, perhaps with interesting ideas for future work. | | **4 marks**  A summary of the experiments is given, but | | **3 marks**  A flawed conclusion. Two of the following components are missing. | | **2 marks**  An incorrect conclusion. Three of the following components are missing. | | **1 marks**  No conclusion. One of the following components is given. | |  |
|  |  | The following components are given.   1. Restate your   research  topic   1. Restate the objective 2. Summarize   the main  topics iv. Significance  of results  v. Conclude the  thoughts | | the conclusion is a mere summary. The ideas for future work are not interesting. One of he following components is missing.   1. Restate your   research  topic   1. Restate the objective 2. Summarize   the main  topics iv. Significance  of results  v. Conclude the thoughts | | i.  ii. iii.  iv.  v. | Restate your research topic Restate the objective Summarize the main topics Significance of results  Conclude the thoughts | i.  ii. iii.  iv.  v. | Restate your research topic Restate the objective Summarize the main topics Significance of results  Conclude the thoughts | i.  ii. iii.  iv.  v. | Restate your research topic Restate the objective Summarize the main topics Significance of results  Conclude the thoughts |  |
| 9. | Formatting of the paper | **5 marks**  A clear and correct formatting. The following components are given.   1. Number of   pages 10 -25   1. Use the   correct template iii. Similarity index less than 20%   1. All the sections given in proper   order   1. Readable pdf   file | | **4 marks**  A clear and correct formatting. One of the following components is missing.   1. Number of   pages 10 -25   1. Use the   correct template iii. Similarity index less than 20%   1. All the sections given in proper   order   1. Readable pdf   file | | **3 marks**  Two of the following components are missing.   1. Number of   pages 10 -25   1. Use the   correct template iii. Similarity index less than 20%   1. All the sections given in proper   order   1. Readable pdf   file | | **2 marks**  Three of the following components are missing.   1. Number of   pages 10 -25   1. Use the   correct template iii. Similarity index less than 20%   1. All the sections given in proper   order   1. Readable pdf   file | | **1 marks**  One of the following components is given.   1. Number of   pages 10 -25   1. Use the   correct template iii. Similarity index less than 20%   1. All the sections given in proper   order   1. Readable pdf   file | |  |

# END