Statistical & Mathematical Methods for Data Science

Project Guidelines

# Overview

This project will provide you an opportunity to explore the methods and techniques learnt in this course. Potential projects would tackle significant real world problems and newly surfacing social conundrums. Your goal is to draw meaningful insights and results from the data or provide innovative solutions to existing problems leveraging the data. You could also exploit machine learning and deep learning models for this goal.

# Project Deliverables

Project deliverables include data and a project report. Your performance will be evaluated on the timely submission of these deliverables.

## 1) Data Collection

Data collection constitutes 35% or 20% weightage of the project. The groups who construct their own questionnaire and data will get 35% credit of the project while the groups who ONLY collect/scrap data will get 20% credit of the project for its submission.

It is further divided into two milestones: **Questionnaire and Data**

### i) Questionnaire

**Deadline:** November 14, 2021

**Weightage:** 35% of Data Collection

You will construct a google form consisting of all the related questions. Number of questions in the form and number of people filling the form will depend on your project.

### ii) Data

**Deadline:** December 19, 2021

**Weightage:** 65% of Data Collection

You will prepare your own data through your questionnaire or scrapping. You will submit your data in appropriate format which can range from csv file to text files and images. The size of data will depend on your project and number of groupmates. You need to compress your data if it contains a large number of files.

## 2) Project Report

Project report constitutes a **research paper** written in IEEE journal/conference format. You are advised to use [Overleaf](https://www.overleaf.com/) for the writeup. You can choose from any IEEE journal/conference format but are encouraged to work on [IEEE Access](https://www.overleaf.com/latex/templates/ieee-access-latex-template/cdxrhtbjgszv) format which can be easily imported in Overleaf. You are required to mention the roll numbers of all group members under their names.

Project report carries 65% weightage for the groups who will construct their own questionnaire and data, while 80% weightage for the groups who will ONLY collect/scrap data. It should comprise 2600-3000 words.

Report is further divided into 5 sections: **Literature review, Data Collection, Introduction, Methodology & Experiments, Results & Conclusion**. Each section will be submitted in PDF format.

### i) Literature Review

**Deadline:** November 21, 2021

**Weightage:** 15% of project report

**Length:** 320-450 words

**Template:** Any IEEE Journal/Conference ([IEEE Access](https://www.overleaf.com/latex/templates/ieee-access-latex-template/cdxrhtbjgszv) is recommended)

In this section you will briefly describe the previously published work related to your project. You can use [Google scholar](https://scholar.google.com/) to browse through different papers and select **at least 10** of them related to your project. You may use [Sci Hub](https://sci-hub.se) to access the PDF of a paper if it’s not available on open access, [arxiv](https://arxiv.org/) or author’s website but be aware that Sci hub is not a legit host.

For each selected paper, you will concisely highlight the method, results and its difference from previous works. You can refer to the literature review section of your selected papers for further guidance.

### ii) Data Collection

**Deadline:** December 22, 2021

**Weightage:** 5% of project report

**Length:** 330-500 words

**Template:** Any IEEE Journal/Conference ([IEEE Access](https://www.overleaf.com/latex/templates/ieee-access-latex-template/cdxrhtbjgszv) is recommended)

This section will contain the specifications and procedure of data collection. It should have an image/pdf of the questionnaire if your project involves one. You should mention the scrapping procedure, scrapped websites, preprocessing and other details in case you scrape your data. The procedure of data collection should be illustrated in such a way that one can collect the data by exactly following it.

### iii) Introduction

**Deadline:** December 26, 2021

**Weightage:** 10% of project report

**Length:** 300-380 words

**Template:** Any IEEE Journal/Conference ([IEEE Access](https://www.overleaf.com/latex/templates/ieee-access-latex-template/cdxrhtbjgszv) is recommended)

In this section you will describe the motivation and problem statement of your project, its significance and sufficient theoretical background to understand your project. You will highlight the gap in current literature and the contribution of your project to reduce it. You may refer to the Introduction section of your selected papers for further details.

### iv) Methodology & Experiments

**Deadline:** January 28, 2022

**Weightage:** 40% of project report

**Length:** 1100-1300 words

**Template:** Any IEEE Journal/Conference ([IEEE Access](https://www.overleaf.com/latex/templates/ieee-access-latex-template/cdxrhtbjgszv) is recommended)

Theoretical and practical methods/models of your project should be added in this section. It should be detailed enough that someone can replicate your project. At least **one diagram describing your methodology** must be added in this section. Implementation details of the experiments should also be mentioned. You can divide this section into two separate sections if your models require thorough explanation or your experiments are extensive. You are advised to take guidance from the methodology & experiment section of your referenced papers. You are also required to submit project code if your project involves one.

### v) Results & Conclusion

**Deadline:** January 31, 2022

**Weightage:** 25% of project report

**Length:** 750-900 words

**Template:** Any IEEE Journal/Conference ([IEEE Access](https://www.overleaf.com/latex/templates/ieee-access-latex-template/cdxrhtbjgszv) is recommended)

You will describe the outcomes and insights gained from your experiments in this section. You can summarize your findings in tabular form. You need to construct the reasons and intuitions behind your results in this section. Finally you will conclude your report by summarizing the achievements/results of the project, its significance, any shortcomings and expected future work.

You can divide this section into two separate sections in case your results require strenuous explanation. Please consult your referenced papers for more guidance.

### vi) Abstract

**Deadline:** January 31, 2021

**Weightage:** 5% of project report

**Length:** 125-175 words

**Template:** Any IEEE Journal/Conference ([IEEE Access](https://www.overleaf.com/latex/templates/ieee-access-latex-template/cdxrhtbjgszv) is recommended)

An abstract comes at the beginning of the report and succinctly describes your project, its motivation and results. It briefly elucidates your problem, your implemented solution/technique and presents core findings. It summarizes each section of the project in 1-3 lines. Please consult your referenced papers for more guidance.

# Collaboration Policy

You can work in a **group of 1 to 4 members**. One student can ONLY work in one group. More data and work is expected from a group having more members and vice versa. Every student will be evaluated individually on the basis of his/her contribution to the project and we’ll ask to mention the %contribution of each group-member at the submission of every milestone. One member may contribute more in one milestone and less in another but his/her overall contribution to the project should be comparable to all his/her group mates.

# Plagiarism Policy

Intra-group discussions are encouraged but Inter-group collaborations are not allowed. You should build your own models and techniques but can refer to literature for guidance. Plagiarism is strictly prohibited in the project report and **maximum allowed limit is 20%** as calculated by Turnitin. Plagiarism (similarity score) from a single source (website, paper, book etc) must not exceed 5%. **Your report will be rejected if it exceeds the plagiarism limit**. You can check plagiarism of your final report before submission by sending a plagiarism check request to **reference.librarian@itu.edu.pk** with your report as an attachment or view it yourself by using the following credentials. Please refer to [this](https://www.turnitin.com/static/plagiarism-spectrum/) link to know about what Turnitin considers plagiarism and [this](https://help.turnitin.com/feedback-studio/canvas/plagiarism-framework/teacher/the-similarity-report/interpreting-the-similarity-report.htm) link to know about how we will evaluate the report that exceeds the maximum allowed limit.

# Late Submission Policy

Timely submissions of the milestones will be an important factor in the assessment of your project and late submissions without pre-approved extensions will be penalized. **10% marks available from the achieved marks will be deducted for each subsequent day** after the deadline. Submission will not be accepted after 5 days of the deadline.

# Summary

Max members per group: **4**

Max plagiarism in report: **20%** (email to [reference.librarian@itu.edu.pk](mailto:reference.librarian@itu.edu.pk) for plagiarism check)

Late submission: **10% deduction** foreach day

Milestones in chronological order

|  |  |  |  |
| --- | --- | --- | --- |
| **Milestone** | **Specifications** | **Deadline** | **Credit** |
| Questionnaire | May or may not be required | November 14, 2021 | 0% or 35% of data collection |
| Literature Review | 320-450 words (Min 10 references) | November 21, 2021 | 15% of project report |
| Data | Data size depends on project and group size | December 19, 2021 | 65% or 100% of Data Collection |
| Data collection (report) | 330-500 words | December 22, 2021 | 5% of project report |
| Introduction | 300-380 words | December 26, 2021 | 10% of project report |
| Methodology & Experiments | 1200-1350 words  (Min 1 diagram) | January 28, 2022 | 40% of project report |
| Results & Conclusion | 750-900 words | January 31, 2022 | 25% of project report |
| Abstract | 125-175 words | January 31, 2022 | 5% of project report |