Term Project- Submission and grading guidelines

Please read and follow these guidelines.

Contents

The term project has been divided into seven phases (Phase 1 – Phase 7) as you learned from the project description. Please start working on the term project phases as soon as possible. There is no late work that will be accepted for submission and deadlines provided are very firm. You'll submit on UBLearns in the folder provided for this purpose. Please make sure you are able to locate it.

For each phase, there will be separate submission links on **UBLearns -> Course Documents -> Term Project Submission**

NOTE: Save every PDF as <Phase#_Report_person#.pdf>.

Each phase will be graded in a digital grading method for 15 points.

0 – no submission by deadline
15 – completed submission with all the requirements
10 – there are missing items, or unsatisfactory in some respects
5 – submission present but many required items missing or code not working

Once the deadline is past, you get a zero for the phase if you did not submit anything, even though you need to complete the phase to work on the next phase.

Phase 1: Define the problem: (Dates: 9/23-10/9)

Due Date: 10/09 11:59pm

Research and identify the problem you plan to solve. Choose an area you are familiar with. Discuss it with the TA and get approval.

- Identify a major area,
- a title for your project,
- identify your clients (and sponsors),
- potential data sources and
- the issue(s) you are trying to address: This will be a 100-word abstract in the form of a problem statement.

Submission guideline:

1. Prepare a pdf document with

- Header: Course name, phase#, term project name, ubit, date
- Body: Name, Area, Title of the project, Data sources, Issues addressed; as listed in the bullets above.

  Organize them legibly (expect not more than 1 page).

2. Save the PDF as <Phase1_Report_person#.pdf>.
3. Upload the PDF and Submit it on ublearns Phase1 submission link.

Grading guideline: Each report will be graded as follows

0 point: No submission by deadline.

5 points: PDF with only name, area, and title.

10 points: PDF with all the sections but incomplete/vague abstract.

15 points: PDF with all the sections. A clear, and complete abstract.

Phase 2: Data collection: (Dates: 9/23-10/23)

Due Date: 10/23 11:59pm

Submission guideline:

1. Your data will consist of multiple tables and sufficient number of observations (rows) and features (columns). This is the raw data that will be used by you.

   Try to provide as many tables as possible to support your analysis and modeling satisfactorily.

2. Prepare a pdf document that **lists** the data sources, names of the data file (with the type .csv, .rds, etc.) and a sentence explaining the contents and the relevance to the project.

Hypothetical Ex:

DataSource 1:

Name of the file: ClinicalLaboratories.csv

Source: CDC https://www.cdc.gov/flu/weekly/

Details of the content: Data from clinical laboratories (the percentage of specimens tested that are positive for influenza). This is used to monitor whether influenza activity is increasing or decreasing.

DataSource 2:
Name of the file: PublicHealthLaboratories.csv
Source: CDC https://www.cdc.gov/flu/weekly/
Details of the content: Data from public health laboratories are used to monitor the proportion of circulating viruses that belong to each influenza subtype/lineage.

3. Zip or compress the **data** and this **pdf** file.
4. Upload the zipped file on ublearns phase2 submission link and Submit.

Grading guideline: Each submission will be graded as follows
0 point: No submission by deadline.
5 points: Only 1 table (and, or) missing PDF.
10 points: Insufficient data (and, or) PDF with missing details.
15 points: Sufficient data and complete PDF.

Phase 3: Data cleaning and preprocessing (Dates:9/25-10/30)

Due Date: 10/30 11:59pm

Submission guideline:

1. Your ***R code*** to perform data cleaning should incorporate at least 5 data cleaning steps:
- Remove NA (null) values
- Zeroing
- Factoring
- Dropping redundant, irrelevant columns
- Dropping missing data rows
- Normalization
- Categorization (like age, income)

Refer Chapter 4 Phase 3.

2. Prepare a pdf document that **lists** the cleaning steps involved.
Against each cleaning step,
- mention how it was used to clean your data.
- paste screenshot of the line of the code involved for that data cleaning.

3. Zip the R code and the PDF.
4. Upload the zipped file on ublearns phase3 submission link and Submit.

Grading guideline: Each submission will be graded as follows
0 point: No submission by deadline.
5 points: R code only. Missing PDF.
10 points: R Code + PDF with fewer than 5 cleaning steps, missing explanation/ screenshots.
15 points: R Code + PDF with minimum 5 cleaning steps; proper explanation and code-screenshots.

Phase 4: Exploratory data analysis (EDA) (Dates: 10/10-11/6)

Due Date: 11/6 11:59pm

Submission guideline:

1. Your ***R code*** to perform EDA:
   - Plots like box plot, scatter plot
   - Chart like bar chart, pareto chart
   - Graphs like line graph
   - Forming data frames
   - Simple stats like head, tail, shape, counts
   - Summary stats on the data like summary, info
   - Data engineering as the dplyr lecture: mutate, %>%
   - We want to see some plots/graphs in this phase
   - Graphics: make sure you use ggplot2 library
   - Include comments in the R code to indicate the type of EDA

2. Prepare a pdf document that **lists** the EDA steps involved.
   Against each EDA step,
   - mention what was the purpose.
   - paste screenshot of the line of the code
   - paste screenshot of the outputs:
     - visualization (if bars, charts, etc. involved) output
     - console (if statistics involved) output.

3. Zip the R code and the PDF.
4. Upload the zipped file on ublearns phase4 submission link and Submit.

Grading guideline: Each submission will be graded as follows
0 point: No submission by deadline.
5 points: R code only. Missing PDF.
10 points: R Code + PDF with fewer than 5 EDA steps, missing explanation/ screenshots.
15 points: R Code + PDF with minimum 5 EDA steps; proper explanation and code, output screenshots.

Phase 5: Modeling and analysis: (Dates: 10/17-11/13)

Due Date: 11/13 11:59pm

Submission guideline:

1. Your ***R code*** to develop a good model using algorithms like:

- KNN classifier
- KMeans clustering
- Hierarchical clustering
- Logistic Regression classifier
- Linear regression
- Graphics: make sure you use ggplot2 library
- We want to see some plots to depict your modeling outcomes
- Include comments in the R code to indicate the type of modeling used

And evaluate goodness of the model using p-value, or ROC, or odds ratio etc.

2. Prepare a pdf document that lists the algorithm involved for classification/ clustering/ prediction. Explain in detail:

- intent of the algorithm, goodness of the model.
- paste screenshot of the line of the code
- paste screenshot of the outputs

3. Zip the R code and the PDF.
4. Upload the zipped file on ublearns phase5 submission link and Submit.

Grading guideline: Each submission will be graded as follows

0 point: No submission by deadline.
5 points: R code only. Missing PDF.
10 points: R Code + PDF with unsatisfactory explanation, missing screenshots.
15 points: R Code + PDF with proper details and screenshots.

Phase 6: Build a data product: (Dates: 10/20-11/20)

Due Date: 11/20 11:59pm

Submission guideline:

1. Your ***RShiny code*** to display your product as a web app. Your code should be reproducible and reusable. It should work with other datasets (of similar kind), or for different parameters, and should work for different parameters.
2. Prepare a pdf document that explains:
    - your RShiny code
    - Web URL to access the web-application
    - Screenshot of the web-application.

3. This phase is like building a simple version of a dashboard. For example, if it worked August 2020 data, it should work for another data set from September. Or if it worked for New York state's data, it should work for Hawaii's data and so on. This is like a simple dashboard.
4. Zip the R code and the PDF.
5. Upload the zipped file on ublearns phase6 submission link and Submit.

Grading guideline: Each submission will be graded as follows
0 point: No submission by deadline.
5 points: Submitted Code doesn't work. Missing PDF.
10 points: Submitted Code works somewhat. PDF with unsatisfactory details.
15 points: Submitted Code works as expected. PDF with satisfactory details.

Phase 7: Report and Communication:(Dates: 11/20-12/4)

Due Date: 12/4 11:59pm

Submission guideline:

1. Prepare a pdf document with

- Problem statement
- Steps carried to solve the problem
- Block diagram detailing the steps involved in your project.
- Outcome/ result
- Instructions to run the R code.
- Details of your product.
- How to use your product. Ex: Rshiny web-app and how to use it.
- Challenges, issues faced in the phases, and how you resolved them.

Use diagrams like architecture or block wherever needed.

2. Complete ***R code.***
3. Zip the R code and PDF.
4. Upload the zipped file on ublearns phase7 submission link and Submit.

Grading guideline: Each report will be graded as follows
0 point: No submission by deadline.
5 points: Submitted Code doesn't work. Missing PDF.
10 points: Submitted Code works somewhat. PDF with unsatisfactory details.
15 points: Submitted Code works as expected. PDF with satisfactory details.