

# Data Warehousing and Data Mining Assignment

MIS SE 20.2/20.3

## Preliminaries

This is a group assignment. Size of the group is 3 to 6 members. You need to prepare a document and hand it over on or before the deadline. Discussions between groups are allowed but **collaborating on the actual solution or copying from other groups are consider as cheating and will be reported** (This will caused both teams to suffer with 0 marks). Your document should not less than 3 pages and should not exceed 10 pages.

## Due Date

You will get 4 weeks to complete your assignment. So final version should be submitted **on or before 1<sup>st</sup> May 2024**

## Part 1 – How to apply data warehousing and mining techniques to solve a real business problem?

### Overview

You have to select a Sri Lankan company (better if it is an IT company) which you think that you can use data mining and warehousing techniques in order to solve their serious problems. You need to prepare a document which highlights the below points.

1. **Company Profile:** This should have an introduction to company which covers (but not limited to) what they are doing, customer base, how they operates (multinational, Sri Lankan, Sri Lankan with overseas branches, etc.), what kind of products they have, what kind of technologies they are using, and anything that you think which suitable to your write-up.
2. **Problem Statement:** You need to think of problem/s which they are currently experiencing (Most common problems are how to increase customer base, time to market of products, reduce bug counts, attaching new customers in to business, etc.) and brief those.
3. **General Solution:** How you can use Data mining and warehousing techniques to solve the identified problem. This should include what kind of data you should collect, what kind of modeling to be used, what kind of dimensions you need and any concept hierarchies, what kind of mining algorithms to be used, etc. Pictorial representation of data using Cuboids, Cubes, Graphs, Tables, Schemas, etc. will add plus points to your document.
4. **Evaluate Model:** This should include how you evaluate the created models, what kind of queries operation user can be executed (better if you can add sample queries and possible output of

those). Also this should clearly specify how you solve the identified problems using suggested mining and warehousing approach.

5. **Conclusion and drawbacks to implement suggested approach:** If you see any issues when introducing such model you can note down here.

Your report needs not to be limited into above scope but it should at least cover all above points.

## **Evaluation**

Evaluation will be based on how well you can apply the things that you have learned in Class to resolve a real-world problem, creativity and organization of the document.

## **Part 2 - Study about data analytics software tools**

### **Overview**

You need to identify and **evaluate** software tools you can use for data analytics. Some of the evaluation criteria can be (but not limited to) easy to use, cost, flexibility, advanced features, support for multiple model creation, support for multiple problem areas, etc. You should provide advantages and disadvantages of these tools. Also, you need to identify a tool you will be using if you have some business analytics issue to resolve (You must provide a brief about the problem you are going to address). Also, you should indicate a clear justification about your selection.

At the end of the report, you should provide a reference list of resources you have used. Your report need not be limited to the scope defined here.

### **Evaluation**

Evaluation will be based on how critically you analyzed the differences and similarities of the tools and the justification for selecting a tool to resolve a business problem.

Please note that your report will be subjected to plagiarism checks.