SECB3203 PROGRAMMING FOR BIOINFORMATICS – GROUP PROJECT (25%)

Project Description

This project provides you with hands-on experience in utilizing Python, Microsoft Azure, or AWS cloud computing. The project is designed to align with current trends in bioinformatics and apply various bioinformatics analyses to the collected dataset.

To accomplish this, you are going to form a three-person team for the project and select a specific bioinformatics trend as the project's focal point. You will be responsible for collecting sample data for analysis. You'll have the flexibility to choose between Python, Microsoft Azure, or AWS cloud computing for data analysis, with a comprehensive report documenting your findings as a mandatory component.

Throughout the project's timeline, you are required to update your project's progress and share your work on GitHub regularly. Also, we recommend looking for a mentor or client (who can be your academic advisor or any other related professional) to enhance the project's depth and guidance.

Note:

- Create a folder named your project title at https://github.com/sean-seah/SECB3203_25261 then the README.md of the folder must include showing your name and your member's name.
- Don't rely on your member to finish everything alone.

Examples of Trends in Bioinformatics

- **Focus**
  - Discovery of biomarkers: Schizophrenia disease, cancer (breast/lung/kidney/stomach/skin)
  - Classification of diseases: Alzheimer's disease, Parkinson's disease, cancer (breast/lung/kidney/stomach/skin)
  - Prediction of cancer prognosis: cancer (breast/lung/kidney/stomach/skin)
  - Protein structural class classification
  - Protein structure prediction
  - Gene-disease predictions
- **Bioinformatics analysis/approach**
  - Gene selection/feature selection/variable selection
  - Deep learning: Convolutional neural network
  - Network analysis/Pathway analysis/Gene expression analysis
- **Data**
  - Image/Video: GitHub/KAGGLE/research papers
  - Gene expression data: GEO
  - Pathways: KEGG
  - Protein-protein interaction network: STRING

PROJECT 2023

Project Activities and Due Dates

Task		Marks (%)	Start Date	Due Date	Duration
1	**Project Proposal and Group Formation.** - Students in groups of 3 - Choose a trend in bioinformatics Proposal 1.0 Introduction 1.1 Problem Background 1.2 Problem Statement 1.3 Objectives 1.4 Scopes 1.5 Conclusion	2.5	Week 2 15 Oct	Week 4 29 Oct, 0000 MYT submit to the e-learning	7 days
2	**Project Progress 1.** - Software and hardware requirements - Flowchart of the proposed approach	2.5	Week 5 5 Nov	Week 9 3 Dec, 0000 MYT submit to the e-learning	14 days
3	**Project Progress 2.** - Importing Dataset - Understanding the data - Importing and exporting data in Python - Getting started analyzing data in Python - Python packages for Data Science - Data Wrangling (Pandas/Numpy) - Identifying and handling missing values - Data formatting - Data normalization (centering/scaling) - Binning - Indicator variables	5.0			
4	**Project Progress 3** - Exploratory Data Analysis (Pandas/Numpy)	2.5	Week 7 19 Nov	Week 11 17 Dec, 0000 MYT	10 days

Task		Marks (%)	Start Date	Due Date	Duration
	- Descriptive statistics - Basic of grouping - ANOVA - Correlation			submit to the e-learning	
5	**Project Progress 4** - Model Development (Pandas/Keras/TensorFlow/Matplotlib/…) - Simple and multiple linear regression - Model evaluation using visualization - Polynomial regression and pipelines - R-squared and MSE for In-Sample Evaluation -Prediction and decision making **Note: make sure to submit the codes and flowchart for model development**	5.0	Week 9 3 Dec	Week 12 24 Dec, 0000 MYT submit to the e-learning	20 days
6	**Project Progress 5** - Model evaluation (sklearn) - Model evaluation - Over-fitting, under-fitting, and model selection - Ridge regression - Grid search - Model refinement	5.0	Week 11 17 Dec	Week 14 7 Jan, 0000 MYT submit to the e-learning	20 days
7	**Project Assessment and Demo** - Live presentation/Recorded video (5 minutes) - Individual and group assessment - shared the recorded video and your GitHub link with your client and put your client's comments in the report (section 4.0) - Report (follow the contents below) 1.0 Introduction 1.1 Problem Background 1.2 Problem Statement 1.3 Objectives	7.5	Week 12 24 Dec	Week 14 7 Jan, 0000 MYT submit to the e-learning submit to the github submit to the youtube	

PROJECT 2023

Task	Marks (%)	Start Date	Due Date	Duration
1.4 Scopes 2.0 Data collection and pre-processing 2.1 Importing dataset 2.2 Data wrangling 2.3 Software and hardware requirements 3.0 Flowchart of the proposed approach 3.1 Exploratory data analysis 3.2 Model development 3.3 Model evaluation 4.0 Testing and validation 5.0 Conclusions				

Note:

- Take photos of your meetings (including meetings with your client) and put them in GitHub to monitor your project progress.
- All progress must be tallied between GitHub and what you submitted to the e-learning.
- During e-learning submission, you just copy-paste or print in pdf what you updated in the GitHub, and make sure the submitted file is pdf.