Data Structures (CS-2001)

*Fall 2022*

**Project Guidelines**

As a requirement, you will need to submit a course project using a combination of the concepts learned in the theory classes and labs. Following are some ideas/guidelines that might help you decide on your project ideas.

# Projects related to Scheduling

Many of the data structures find their classic usage in scheduling tasks. This scheduling can either be in the form of algorithms that work in industries for the conveyer belts or more commonly in Operating Systems. Usually Queues, Priority Queues, Linked Lists and Graphs are ideally suited to problems related to resource allocation, process scheduling and interrupt management in the Operating Systems and are used in many modern algorithms.

# Projects related to Natural Language Processing (NLP)

## Text Classification using Sparse Matrix:

The data structure Sparse Matrix is simply a matrix filled mostly with zeroes. The nature of this data structures made it ideal for representing text (sentences) as a Sparse Matrix and then applying machine learning algorithms on it for some useful tasks e.g. Sentiment Analysis, Text Summarization, Topic Identification, and Analysis on Social Media posts.

## Parser & Semantic Analyzer for Programming Language:

A combination of data structures can be used for creating a system i.e. compiler for analyzing the lexical structure, syntax, and semantics of programs written in C.

## Search engine utilizing core Data Structures

A search engine can be implemented by using concepts like Stacks (for page navigation) and Sparse Matrix (for query text).

## PrivRank for social media

Social media sites mine their users’ preferences from their online activities to offer personalized recommendations. However, user activity data contains information which

Can be used to infer private details about an individual (for example, gender, age, etc.) And any leak or release of such user-specified data can increase the risk of interference attacks.

# Projects related to Pattern/Data Mining

Application domains like bioinformatics, social networks, and privacy enforcement often encounter uncertainty due to the presence of interrelated, real-life data archives. This uncertainty permeates the graph data as well. This problem calls for innovative data mining projects that can catch the transitive interactions between graph nodes.

# Projects based on Deep Learning / Machine Learning theme

In simplest of terms, Deep Learning is a field that deals with algorithms used for predicting things, and Neural Networks are graph like structures (networks) used to visualize the prediction process. The process of classification is simply predicting whether your input belongs to one of the given classes e.g. is it a duck or not (binary classification), is it a duck, whale, frog or turtle (multi-class classification).

You can learn more about it here:

*htt*[*ps://w*](http://www.kdnuggets.com/2015/01/deep-learning-explanation-what-how-why.html)*ww.k*[*dnuggets.com/2015/01/dee*](http://www.kdnuggets.com/2015/01/deep-learning-explanation-what-how-why.html)*p*[*-learning-explanation-what-how-why.html*](http://www.kdnuggets.com/2015/01/deep-learning-explanation-what-how-why.html)

For projects based on this field, you can take inspiration from any of the following:

## Binary Tree for multi-class classification:

In classification problems where we have more than one classes, we can mimic binary classification with the help of a binary tree by recursively splitting classes into two groups.

## Model Stacking

In this problem, many different models are used for predicting the class (classification process), and afterwards their individual predictions are *stack*ed to make a final prediction.

## Using linked list for linear regression model

For a model capable of doing linear regression (a neural network with linear activation functions), linked lists can be used for an efficient implementation. Its effects or usefulness can also be analyzed afterwards.