**CMPSC 463 Project 1 – Time Series Clustering with Divide & Conquer**

**Jack Demtshuk**

**Professor Janghoon Yang**

**Description**

This project tasked us to design a system that performs unsupervised clustering of time-series segments extracted from a database using divide and conquer strategies and algorithmic reasoning. The goal was to group similar data signals together and analyze their internal structures using closest-pair algorithms and maximum subarray analysis. While initially asked to use PulseDB as the database source and to collect 1000 segments from 10 different sources, I opted to use VitalDB, and only collected from 1 source, due to the large size of the PulseDB dataset.

**Installation and Usage**

The project has two different Python files that must be used in order. The first file to be run is segmentScraper.py, which is used to collect the 1000 segments that will be used in our algorithms. I collected this data from VitalDB\_CalBased\_Test\_Subset.mat, which can be downloaded from Kaggle using the link I provided in Dataset Link. This first program will run through the mat file and scrape out the 1000 data segments of ABP signals, before saving these segments into a new npz file titled VitalDB\_Subset\_Processed.npz.

After this, you can run the second Python file, algorithmApplication.py. This file will use VitalDB\_Subset\_Processed.npz to pass the 1000 segments collected in segmentScraper.py through the algorithms. This npz file is accessed correctly as long as it is saved in the same folder as the project. The second file first performs recursive clustering on the segments, before then passing the next function, which finds closest pairs. After closest pairs, Kadane’s algorithm is applied to the data, and then visualizations are created for all algorithms and their results.

**Structure of Code**

**Description of Algorithms**

**Verification**

**Execution Results**

**Discussion**

**Conclusion**