**CSCI 4314/5314 Benchmark - Algorithms for Molecular Biology**

**Members:**

Monte Anderson, Christine Samson, Han Ngo, Byron Bearden, Rasmi Lamichhane, and Sushma Colanukudhuru

**Dataset Generation:**

Description of how the dataset was generated, and how it was then used as an input to the application for benchmarking purposes. (Using Generated\_Data.py) to get FASTA file output of random sequences.

**BioProspector:**

How to run BioProspector and how we were getting the benchmarking results. The commands, maybe a little on how BioProspector works, and the length of motifs we were generating on a sequence of what length and how big the file was.

**Benchmark Results:**

Will probably attach this in a zip file later (containing each of the benchmarking results).

**Benchmark Interpretation:**

Tally up what BioProspector got right, and what it got wrong. Include positives, negatives, false positives, and false negatives.

|  |  |
| --- | --- |
| *Type of Result:* | *Count:* |
| Positive | Number of positives from BioProspector |
| Negative | Number of negatives from BioProspector |
| False Positive | Number of false positives from BioProspector |
| False Negative | Number of false negatives from BioProspector |

Runtime of BioProspector on certain files:

* Files with about ~75 sequences took about 7 seconds
* Files with more than 5,000 sequences took about 25 minutes and counting

**Accuracy:**

Percentage of how accurate the benchmark was after comparing with the expected results from our own dataset, and any other accuracy factors that can contribute to it.