

# Coding Group: Project Outline

Max Carlson

## **Abstract**

It is my intent with this project to explore a number of topics in statistical data processing to both increase my potential employability and satisfy my educational curiosity. The project will be partitioned into a number of sub-projects, each exploring a single concept or tool that will be necessary to understand and implement a machine learning algorithm for use on a to be determined data set. Since real data sets can and will be arbitrarily large, scalability will be the core focus of this project with each sub-project having both a sequential and parallel/distributed implementation.

## **1 Introduction**

## **2 Sub-Projects**

### **2.1 Principal Component Analysis**

### **2.2 Clustering ( $k$ -means)**