**Capstone Project 1:**

**Predicting Alzheimer’s Disease**

**Chuck Tucker**

**Introduction:**

Alzheimer’s disease is a particular form of dementia that is eventually terminal and has specific pathology within the brain. Both conditions have similar symptoms at onset. Alzheimer’s disease is the most expensive disease in America according to the Alzheimer’s Association ([alz.org](http://alz.org)). Caring for someone with Alzheimer’s or dementia is very demanding, with the Alzheimer’s Association estimating that around 15 million family members are providing 18.2 billion hours of unpaid care. This care would have an economic value of more than $230 billion.

These impacts alone are enough to warrant early detection of the disease and models that are able to predict those at higher risk. There is also a correlation that suggests Alzheimer’s caregivers experience increased health decline compared with caregivers for other diseases. Early diagnosis and prediction of high risk individuals can aid in more advanced planning by families, medical care facilities, insurance companies, and companies that provide in-home care for these patients. Additionally, there are many activities and treatments such as art and music that have been demonstrated to preserve brain connections and establish new ones very early on or prior to the onset of Alzheimer’s. Thus, identifying those at risk or in the early stages of the disease is an important task for diagnosis and treatment.

This particular dataset could possibly be used to identify patterns in cognitive tests, brain scans, and medical history that are good predictors of an Alzheimer’s diagnosis. This diagnosis typically involves analyzing medical history, a physical exam, lab tests, and monitoring changes over time. It would be beneficial to medical providers and patients to be able to discern who is likely to need the additional testing and who is most likely to develop Alzheimer’s.

The goal of this project is to examine certain biomarkers in conjunction with basic medical history to train a model based on patients diagnosed with Alzheimer’s disease, then apply this model to a test dataset in an attempt to predict future clinical status in terms of a diagnosis, later iterations of Alzheimer’s Disease Assessment Scale (ADAS) cognitive screening tests, and ventricle volume in the brain. With the data provided and available background information on Alzheimer’s, the prediction is that the clinical factors within this dataset that correlate the highest with dementia are sex, age, education level. However, brain scan information will be used to follow the progression of the disease to provide predictive information associated with the progression of the disease. This information would allow health care providers to give an accurate assessment of a patient including what to expect in terms of symptoms and patient outlook.

**Dataset:**

The data for this project was acquired from the Alzheimer’s Disease Neuroimaging Initiative ([ADNI website](http://adni.loni.usc.edu/data-samples/access-data/)). The investigators participated in the design and implementation of ADNI and/or provided data but are not involved with this project. The exact data used was part of the [Tadpole Grand Challenge](https://tadpole.grand-challenge.org/Home/). A complete listing of ADNI investigators can be found at <http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf>

The data include results from paired longitudinal and cross section MRI brain scans from the same patient over time, with the longitudinal scans separated by at least one year for a patient. The cross sectional scans were produced from a single visit for each patient, and a reliability dataset produced from patients without dementia.

The data from the MRI scans includes various longitudinal and cross-sectional scans. These scans provided information on many different brain measures, and this project will focus on a few included ventricle volume. Main cognitive tests include ADAS, Mini-Mental State Examination (MMSE), Rey Auditory Verbal Learning Test (RAVLT), Montreal Cognitive Assessment (Moca), and ElectroCochleoGraphy (Ecog). Also included in the data are gender, age, and education level.

**Project:**

The approach to this project will be to analyze the correlation between the reported factors and occurrence of Alzheimer’s. Because it is already established that there is no one factor that causes this disease, multivariate statistics will be used to look for patterns of correlation. Appropriate measures of model fit will be used to assess how accurate the model is. Machine learning algorithms will be applied to find the measures that best predict the progression of the disease in terms of cognitive decline, ventricular volume, and diagnosis of Alzheimer’s.

**Audience**:

The primary goal of this project is to identify those specific variables from this dataset that best predict the progression of Alzheimer’s. These results could potentially aid medical professionals in the diagnosis of Alzheimer’s by providing a list of measures capable of producing an early diagnosis prior to extreme onset of symptoms. This is extremely important for the treatment of Alzheimer’s disease, as early diagnosis and treatment is highly beneficial to patient outcomes. This project also has some potential to provide doctors and patients with information on which patients should be referred to MRI scanning and those that have a lower risk, possibly saving both patients and medical professionals time and money.