**Introduction**

Health care is a passionate topic. It’s a daily news headline. In fact, in the November 2017 elections, CNBC reported that 70% of Virginia voters reported that health care was the most important issue influencing their vote (Mangan, 2017). Given that large percentage, one might find it shocking to learn that “no show” appointments are one of the largest healthcare costs to medical practices. In fact, it’s estimated that no-shows cost the U.S. health-care system more than $150 billion a year (Toland, 2013). But the “costs” are not limited to monetary damages. In a country with over 230,000 physician practices in the U.S. (Martin, 2017); we also have extended wait times to see physicians. In a survey of physician practices in 15 regional areas, it was estimated that the average wait time for a new patient to see a physician is 18 days (Hawkins, 2014). While no-show appointments are not the only reason for this extensive delay in getting treatment, they do attribute to the issue, as every no-show missed by one person is a actually an appointment missed by two- the person it was originally booked for and a person waiting for an appointment (Peres, 2016).

In this project, using 100k+ records of medical appointments, we analyze no-show appointments in order to determine the possibility of predicting likelihood of no-show appointments. This prediction is based on the analysis of the following variables: Gender, Age, Scholarship, Hypertension, Diabetes, Alcoholism, Handicap, and SMS received.

**Dataset Description:**

The analysis was conducted on a 100K+ dataset from Kaggle. The 100K+ records represent public healthcare medical appointments from Espirito Santo State, Victoria, Brazil (Hoppen, 2017). The dataset contains the following 15 variables:

· PatientId = ID of the patient

· AppointmentID = ID of the appointment

· Gender = Gender of patient

· ScheduledDay = The day which appointment scheduled

· AppointmentDay = The day which appointment planned to occur

· Age = Age of the patient

· Neighborhood = The place where hospital located. These hospitals are in neighborhoods within Victoria, a capital city, and receive dozens of patients from other cities.

· Scholarship = If the patient has “scholarship” or not. In this case, scholarship is a social welfare program of the Brazilian government ("Bolsa Família", 2017).

· Hypertension = If the patient has Hypertension or not

· Diabetes = If the patient has Diabetes or not

· Alcoholism = If the patient has Alcoholism or not.

· Handicap = If the patient has Handicap or not. This is not a binary measurement, but is measured in the total number of handicaps. For example, if the person is blind and can't walk the total is 2.

· SMS\_received = If the patient received an SMS for the appointment. This also is not a binary measurement, but measures the number of SMS texts the patient received.

· No.show = no show information. “Yes” means patient did not come to the appointment, “No” means patient came to appointment.