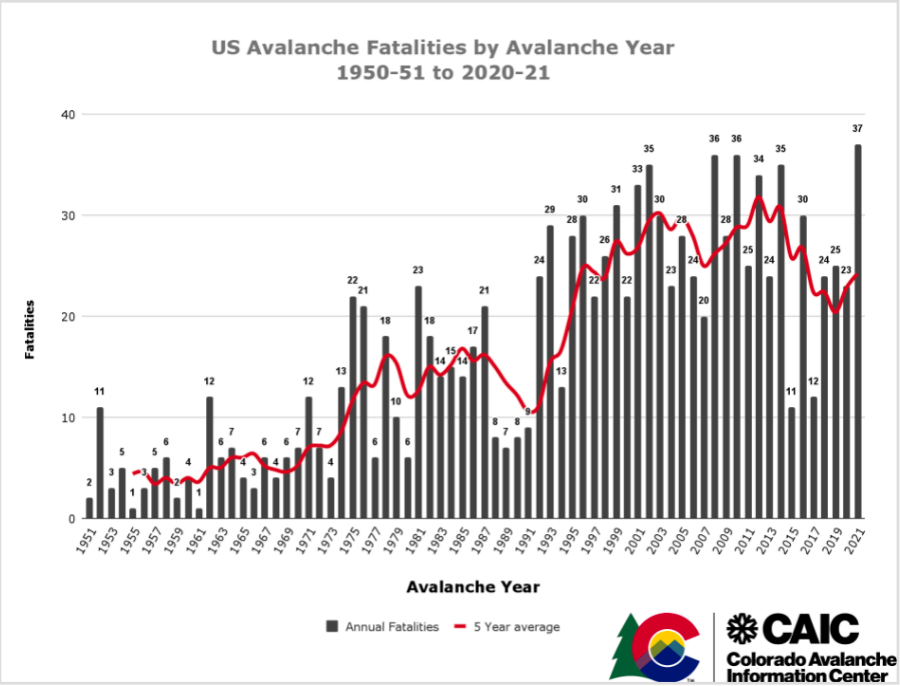
ISyE 6740 – Spring 2021
Bayesian Backcountry Boarder - Project Proposal

Team Member Names: **Evan Anderson** (eanderson42)

Project Title: **Colorado's Front Range Avalanche Prediction**

Problem Statement

In the American West, particularly in the state of Colorado with easy access to the rocky mountain range, winter gravity sports are an incredibly popular activity. In particular, backcountry skiing, snowboarding, and snowmobiling have been increasing in popularity of the past decade. With this rise in popularity, and with a lack of knowledge by these new sport participants, the inherent dangers of being in a backcountry environment in the winter are not adequately observed, in part because the information is not easily accessible to those who are untrained in interpreting the weather and avalanche safety reports being published. This has culminated in a significant increase in deaths in the Colorado Front Range region, and over the entire United Staes over the past decade (CAIC, 2021).



The goal of this project is to take the vast amount of information around weather observations and avalanche reporting, and provide a simple to interpret avalanche risk level for anyone wanting to recreate in backcountry, stating whether or not an avalanche is likely for a given day in a given area, and if so, on what aspects (direction a given mountain slope is facing) is that risk present.

Data Source

For data collection and model learning, two major sources will be used, with supplemental sources as necessary.

The first data source will come primarily from a snow fall reporting service, Snow Telemetry, via National Water and Climate Center (SNOTEL). This information is collected from hundreds of stations across the rocky mountain range (and others) and includes data such as daily snow fall, wind direction and speed, and high, low, median, and average temperatures, all on a daily basis. These observations can be supplemented with data from similar collection agencies like CoCoRaHS, and RAWS if more collection is necessary. SNOTEL data was specifically selected because it is the standard that the CAIC prefers to use for all avalanche forecasting, and it's data is readily accessible for large historical sets without data scrapping from websites being necessary.

The second data source that will supply the pre-classified values will come from the Colorado Avalanche Information Center (CAIC). This organization collects avalanche observation information and forecasts avalanche types for large scale areas across Colorado. These forecasts are manually intensive, with human reported observations and expert forecasts analyzing data each evening for the given forecast the next day. The data utilized from the CAIC will be confirmed avalanche events for the past 10-20 years, with the events filtered only to "large" scale avalanches, meaning the snow slide has the potential to bury or harm humans in its path. These avalanche reports include information on location, face aspect, avalanche size, avalanche type, and trigger events.

Merge the two datasets will require some manual effort, specifically mapping the large number of weather data collection locations to the reported avalanche observation locations. This process will be somewhat manual, but once created the mappings will be easily utilized.

Will split models into human vs natural cause.
Will filter to only >R3 or >D2

Will need to map SNOTEL weather station data to various landscapes. Not sure how granular we can get with these, if it'll be "entire front range" or "Loveland Pass". Might need to choose one SNOTEL station and apply to an entire BC zone. Or take averages.

Methodology

The goal of avalanche prediction utilizing the data described above will be split in to two distinct steps. The first is a binary classification of “avalanche likely” or “avalanche unlikely” given the current weather conditions. The second step will do further classification to specify what aspects are most at risk for avalanche danger given that avalanche conditions are presently and likely from the initial classification. These two pieces of information will provided the most critical and most useful data points for backcountry recreations when making the decision of if and where to ski in avalanche terrain.

The initial binary classification will utilize the weather data primarily for doing the classification. Avalanche activity is huge dependent on recent snowfall information (the past 2-3 days) but can also be significantly impacted by the first snow date and depth of the season, previous weeks and months snow fall and temperature swings, and other weather events occurring before the current days observation. As such, all weather data will be manipulated to create features of the data for pervious weather data. Specifically we will have prior 7 days, and prior week, two week, and 3 week aggregations, along with a given season’s first snow(s) and first snow depth information. With this data, using a classification method such as logistic regression, naive Bayes, or K-nearest neighbor classification, a binary high or low risk will be assigned using the CAIC avalanche observation data classified trading set.As mentioned above, the days in a given area classified as having an avalanche event will be limited to only avalanche events that are considered “large”, or at risk of burying a human.

Once this high-low assignment has been made, given that the risk for avalanche is high, a second classification for the aspect(s) most at risk for avalanche will be made. This will be don by utilizing a combination of the weather data (with all historical context) as will as the aspect and size information present in the avalanche observation data. Typically only a single aspect in a given day is at risk for avalanche given the prior weather behavior. As such, any day that is at risk for avalanche will be assigned a single aspect of high risk, with the directions above and below being classified as lower but present risk (ie. North West aspects are at high risk on Jan 23rd, and North and Westerly aspects are at low but present risk). This classification will be slightly more complex, as the primary components contributing to aspect risk include more detailed weather information that are less relevant for over all risk, such as window direction and speed, and temperature fluctuation throughout the day. With this in mind, while a more simple model like KNN or Bayes

classifier may suffice, GMM and even ANNs will be explored and selected based on overall performance.

Evaluation and Final Results

For this problem statement, it is fortunate that we have a large amount of data for up to 20 years of history, for both daily weather and avalanche observation. From initial exploration, there are roughly 30 to 40 avalanche observations annually, and daily weather observations. With this large volume of data, reasonable training and test data sets can be created and used to evaluate the various proposed classification methods, with methods like k-fold cross validation not necessarily being needed. The models will be evaluated using the typical methods, mis-classification error, precision, recall, and the results of a confusion matrix. The goal of the evaluation is to ensure that whatever modal is selected is as accurate as possible, where, with the assumption that an avalanche is unlikely, we ensure that the accuracy leans to the side of a type 2 error, or predicting that avalanche risk is high when it may not be, so as to ensure that information provided is airing on the side of caution.

The final results of this project will be, at a minimum, a simple classification output for a given day, whether the avalanche risk is high or low, and if it is high, what aspect is most likely to be at risk. If time allows, implementing the model in to be utilized by the open internet would be ideal, as it is one that (assuming good overall results) I will be utilizing daily in the winter months to supplement my preparations for any backcountry expedition.