

## **Project Proposal**

### **ENSF 611**

Tobin Eberle

Jeff Wheeler

Tom Wilson

## 1. Proposed Client

---

The proposed clients for this project are avalanche agencies and professionals, mountain guides, and winter back-country recreationists as all these parties use avalanche forecasts to inform their back country decisions.

## 2. Topic of Investigation

---

In winter months, daily avalanche forecasts are released for regions of Canadian national parks by Avalanche Canada. These ratings indicate the likelihood of avalanches on a scale of one to five (one being unlikely 5 being very likely) for different regions of mountainous terrain (alpine, treeline, below treeline). Avalanche forecasters use a combination of objective and subjective factors including seasonal weather patterns, field observations, and snow pack tests when determining the ratings they release. This project will investigate which features are most important in predicting these danger ratings and whether the ratings can be accurately determined with only quantitative data (i.e. weather data).

## 3. Method

---

The data required for this analysis has different origins and so will need to be aggregated before the models can be developed. Weather data can be sourced from the Government of Canada data portal and avalanche ratings, observations, and observed avalanche issues can be obtained from Avalanche Canada. Preprocessing will be required to clean the data and then the models can be trained. We will train three different types of classification models and use grid search hyperparameter tuning to find the best versions of each, and we will compute feature importance/feature weights to determine which aspects of our data are most relevant. Additionally, we can train new models on an additional dataset which includes metrics obtained from field evaluations to explore how much of an impact they have on determining ratings.

## 4. Components

---

Datasets:

- Weather Data : obtained from Government of Canada for Yoho National Park weather station from 2013-2023
- Avalanche Data: obtained from Avalanche Canada webAPI for 2023-2023

Models:

- Logistic Regression
- Ensemble Classification
- Support Vector Classification

## Framework/Components:

1. Aggregate data from different sources
  - a. Weather data pulled using mingw scripts
  - b. Avalanche data pulled using python script for API
2. Preprocess data
  - a. Combine datasets
  - b. Perform encoding on text based columns
  - c. Fill/remove missing data, curate based on winter months
3. Split data
  - a. Split for target vectors for each level of terrain (alpine, treeline, below treeline)
4. Tune hyperparameters
  - a. Use grid search to find best parameters
5. Train models
  - a. Will need to train each model for each target vector (alpine, treeline, below treeline)
6. Predict and validate models
7. Visualize results
  - a. Compute confusion matrix
  - b. Compute feature importances
  - c. Compare results between each level of terrain predicted (e.g. were any better classified than others)