

Prediction of Avalanche risk in the alps by the use of Machine Learning Algorithms. A case study in Kaprun/Mooserboden

2010695005

FH Salzburg

02. January 2022

Contents

1	Introduction	2
2	Related Work	3
3	Concept	3
3.1	Avalanche Expert Survey	3
3.2	Implementation	4
3.2.1	Data	4
3.2.2	Feature Selection	4
3.2.3	Machine Learning Models	4
3.2.4	Results	5
3.2.5	Generalization	5
4	Supervisor	5
	References	7

1 Introduction

Avalanches are among the most dangerous natural disasters [1]. They threaten people and environment in seasonally snow-covered mountain regions such as the alps. In 2021, 265 recorded avalanches (recorded on Lawis) in Austria killed 17 people and injured 5 others. A total of 41 people were involved in the events [2]. The occurrence of snow avalanches is steadily increasing due to climate change [3] [4] [5]. In order to enable preventive measures to combat avalanches, a good assessment of the impending danger is necessary. The complexity of this task has already been described in several studies. It comes from the fact that there are many potentially influential parameters, of which in most scenarios only a few are available and they change significantly even for small geographical differences. For example, the estimation of avalanche danger levels is mostly based on weather data, which are also predicted, results of snowpack models and data on snow instability. In addition, information on the terrain is often used for these features [5]. Avalanches have been artificially triggered by explosions for a long time. For this preventive fight against snow avalanche accidents, automatic and semi-automatic snow avalanche detection systems have been developed. Of these, infrasound-based systems are the most promising, in the early detection of slow avalanches, for example ice avalanches [6] [1]. In addition, these systems are used to detect as many avalanches as possible. Many avalanches have not been recorded, which means that the accuracy of predictions made by automatic systems is compromised. Geographic Information Systems (GIS), Hierarchy Process Methods (AHP), and Remote Sensing (RS) are used together for this purpose to assess and assess avalanche hazards. with the help of these systems avalanche hazards can be detected, but they have a high error rate. Therefore, predicting the risk of avalanches on selected slopes would be a great improvement in the precision of preventing avalanches by blasting them. It has been revealed that some machine learning can achieve good results in predicting natural disasters and explicit avalanches [3] [4] [5]. Avalanche warning services could receive promising forecasts for the following days in advance and make them available to the public. This leads to a significant increase in safety for backcountry skiers and other winter mountain sports enthusiasts.

Backcountry skiing is a sport in which the athletes are skiing unmarked or unguarded areas within or outside the boundaries of a ski resort. For this, they either climb the mountain with touring skis or are brought up by helicopter or snowmobile. Having good predictions for the avalanche hazard, makes it easier and safer to plan these tours. This has led to the author's personal interest for this topic and the resulting possibilities to get high precision predictions of avalanches hazard for explicit Slopes by the use of machine learning models.

The aim of the master thesis is to predict the probability of avalanches from explicit slopes. These predictions should be made for a certain number of days in the future, for example for one week in the future. This is done by using terrain related data (e.g. slope, slope orientation by cardinal directions or proximity to rivers and streams) in conjunction with meteorological data for the day to be predicted (e.g. weather, temperature, snow depth, wind direction), which will be also used retrospective. Thus, the study is intended to provide an answer to the questions, "What data features are needed to predict a day-accurate avalanche probability for explicit slopes by the use of machine learning methods?" and "How many days in advance can these

predictions be determined?”.

2 Related Work

In the context of avalanche risk prediction, several studies have already been conducted. Machine learning methods have been applied on this purpose. Data sets of avalanche events from smaller mountain areas were used as case studies. In the alps a dataset of avalanche events from the area around Davos, Switzerland over the last 13 years is one Example for these case studies. The study about the data from Davos aimed to predict an entire winter season of avalanche days. To get the Meteorological Data for their study the Team from the SLF, Switzerland combined the data of snow avalanche events with those of an automatic weather station and the simulated snowpack properties, like new snow depth, liquid water content, Stability indices, critical crack length, and the hand hardness, of the SNOWPACK model. A random forest model was then trained on these merged data. One finding of the resulting study is that the predictions without the snowpack factors are just as good as those without this additional data. Furthermore, they concluded that their prediction attempts were severely limited by the use of inaccurate, biased avalanche-related data and by the fact that the spatial scale was too large for their models. They came to the result that forecasting on a small spatial scale using only one avalanche warning system could work well and could be a good aid for avalanche hazard forecasting services. [7]. In another study, avalanche hazard maps were attempted using predictions from machine learning methods. The space around the watershed of Karaj, Iran was used in this paper. The machine learning methods used in this study were SVM and MDA. Also for this purpose, various meteorological data were brought in and used to train the algorithms. From the study, it was found that avalanches seem to slide out mainly in the vicinity of streams. In addition, MDA performed better in predicting avalanche danger compared to SVM. Both methods produced results with an accuracy of 0.83 for the SVM and 0.85 for the MDA. [5].

3 Concept

3.1 Avalanche Expert Survey

Determining avalanches is a complex task, and for this purpose avalanche awareness levels have been developed. These are determined on the basis of various factors at the respective time and for the respective location. For this reason the first empirical part of the work, a survey of avalanche experts is conducted in the form of a questionnaire. The purpose of this survey is to obtain information about the current approach of the experts in the determination of avalanche danger levels. The evaluation can also provide initial indications of good features in the data and be helpful in feature selection. In the course of the Master Thesis, the results of this survey will also serve as an additional evaluation and comparison for the results of the machine learning models.

3.2 Implementation

3.2.1 Data

Recorded avalanche events are rare and a whole data set of homogeneous data is even more. If you want to predict avalanches you also need Meteorological Factors over the entire period of the study, which can lead to the formation of wet and dry avalanches [7]. The study area for this case study is the AHP Verbund Kaprun/Mooserboden. The homogeneous long-term data set of the mentioned region was provided by the Avalanche Warning Service Salzburg. The data consists of a combination of avalanche related data (avalanche type, avalanche size, and others), daily breakdown of meteorological data (weather, temperature, wind direction, wind strength, etc.), and terrain data. This information is combined in a daily resolution in a homogeneous data set over the years.

3.2.2 Feature Selection

Analyzing high-dimensional data sets has always been a major challenge for data scientists and machine learning engineers. Machine learning research has long assumed that too many columns of data lead to a reduction in prediction quality. Therefore, it is even more important to focus on a small number of features [8]. In the past, several studies have been done to find good parameters for the complex prediction of avalanches. How important parameters were considered to be for the avalanches in each study seems to be strongly related to what parameters were available for the study. For example, in the study in Iran from the article "Snow avalanche hazard prediction using machine learning methods" [5], elevation was not ranked as particularly important for prediction, whereas in a study in India from the article "Parameter importance assessment improves efficacy of machine learning methods for predicting snow avalanche sites in Leh-Manali Highway, India" [4], it has been ranked as the second most important feature. In the First Study, more additional meteorological and geographic parameters were available, which appear to be more important than the elevation [5] [4]. From the knowledge gained through the research that a good feature selection naturally depends on the available features and the feature to be predicted, the survey of avalanche experts on their methods and the parameters, which are used by them to calculate the avalanche risk, are used in this study to get knowledge about how important each parameter is for the creation of avalanche risk reports. These Informations can also be important for the Feature Selection. To get a good comparison and as well as good features, in addition typical methods from data science for sensitivity analysis are performed to find a set of suitable data columns for building the machine learning models. In addition, the relevance of the characteristics identified in other studies is evaluated and compared with the results from the study prepared for this purpose.

3.2.3 Machine Learning Models

In order to achieve adequate results, a series of machine learning models will be trained in the context of the thesis. In the past, some models have already proven their worth in predicting natural disasters. For example, the SVM (support vector machine) and the MDA (multivariate discriminant analysis) models. They are useful for detecting subtle patterns in complex data sets

and Flexible in handling data of different dimensions. SVM models are designed to deal with high dimensional data. That's one aspect why they have already been used to predict natural disasters, such as earthquakes, floods, typhoons, drought, landslides and avalanches [5] [4] [9]. MDA forms efficient linear combinations of independent variables. MDAs have not been used that often to predict natural disasters, but shows superior performance compared to SVM in the case study in the Karaj water conservation area in predicting avalanche risk levels [4]. Because SVM and MDA models have proven themselves in the prediction of natural disasters, they will be used for the studies of this Master thesis.

3.2.4 Results

The prediction quality of the machine learning models is evaluated using various statistics such as accuracy, false alarm rate, and others. All evaluations of the MDA and SVM will be compared to each other. One finding of the study will be which of the two methods has the best performance in predicting avalanches with the available data. In addition, the effectiveness of each feature used to train the ML models is also determined using statistical values such as the decrease in correlation and the AUC statistic. The values determined are compared with those of the other models in the course of the evaluation and the differences are recorded. All findings from the evaluations of the machine learning models are also compared with the results from the survey of avalanche experts.

3.2.5 Generalization

In Data Science and the topic of machine learning it has always been interesting if the trained models can be used for similar scenarios but with different conditions. For the approach to test the generalizability of the trained machine learning models of this thesis, an additional data set was provided by the Swiss Institute for Snow and Avalanche Research (SLF). This dataset contains records of avalanche events in the area of Davos, Switzerland, as well as the meteorological data of the entire period. In the course of the general contact with all avalanche warning services in the alpine region, while the research for high quality and homogeneous data in connection with avalanche events of the last years. It was pointed out to me that the data set has already been used several times to train machine learning models. With this data, a random forest model was trained in a study by Harvey, Alec van Herwijnen and Bettina Richter.

4 Supervisor

In the context of the master thesis I am in contact with my potential supervisor Professor Doctor Simon Ginzinger. He gave me the feedback that the topic is very interesting and does have the potential and scope for a master thesis. Such as most data science projects this topic is very dependent on the necessary data. In the course of this I contacted various avalanche warning services in the alpine region. Some of them were very interested in my project and the avalanche warning service Salzburg offered me the support of Ing. Harald Etzer who is the head of the Verbund AHP Kaprun/Mooserboden and supports the project. He also arranged that I receive

the data used for the master thesis. In addition, the Avalanche Warning Service Salzburg will provide me with the necessary data.

References

- [1] V. by Hermann Maurer, version 3, 2019. [Online]. Available: <https://austria-forum.org/af/AustriaWiki/Lawine?version=3> (visited on 01/15/2022).
- [2] L. S.L.O.-L.V.L.K.U. W. Lawinenwarndienst Tirol Lawinenwarndienst Steiermark. (2022). Lawinen Ereignisse, [Online]. Available: <https://lawis.at/incident/> (visited on 01/10/2022).
- [3] Y. L.G. B. Eric Martin Gérald Giraud, „Impact of a climate change on avalanche hazard“, *Annals of Glaciology*, vol. 32, pp. 163–167, 2001. [Online]. Available: <https://doi.org/10.3189/172756401781819292>.
- [4] B. D. V. Anuj Tiwari Arun G., „Parameter importance assessment improves efficacy of machine learning methods for predicting snow avalanche sites in Leh-Manali Highway, India“, *Science of The Total Environment*, vol. 794, 2021. [Online]. Available: <https://doi.org/10.1016/j.scitotenv.2021.148738>.
- [5] A. M.F.S.-H.V.P.S. S. Bahram Choubin Moslem Borji, „Snow avalanche hazard prediction using machine learning methods“, *Journal of Hydrology*, vol. 577, 2019. [Online]. Available: <https://doi.org/10.1016/j.jhydrol.2019.123929>.
- [6] T. Thüring, M. Schoch, A. van Herwijnen, and J. Schweizer, „Robust snow avalanche detection using supervised machine learning with infrasonic sensor arrays“, *Cold Regions Science and Technology*, vol. 111, pp. 60–66, 2015, ISSN: 0165-232X. DOI: <https://doi.org/10.1016/j.coldregions.2014.12.014>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0165232X14002419>.
- [7] B. R. Stephan Harvey Alec van Herwijnen, „Statistical Nowcast of Avalanche Activity at the Regional Scale“, 2016. [Online]. Available: <https://arc.lib.montana.edu/snow-science/item/2437>.
- [8] J. Cai, J. Luo, S. Wang, and S. Yang, „Feature selection in machine learning: A new perspective“, *Neurocomputing*, vol. 300, pp. 70–79, 2018, ISSN: 0925-2312. DOI: <https://doi.org/10.1016/j.neucom.2017.11.077>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231218302911>.
- [9] A. Pozdnoukhov, R. Purves, and M. Kanevski, „Applying machine learning methods to avalanche forecasting“, *Annals of Glaciology*, vol. 49, pp. 107–113, 2008. DOI: 10.3189/172756408787814870.