

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	Methods	3
3.1	Feature selection	3
3.1.1	Decision Tree	4
3.2	Feature extraction	4
3.3	Machine Learning Models	4
3.3.1	SVM (Support Vector Machine)	5
3.3.2	MDA (multivariate discriminant analysis)	5
3.3.3	Neuronal Network	5
4	Data	5
4.1	Origin of the data	5
4.1.1	Data owner	5
4.1.2	Recording area	5
4.2	Meteorological Data	5
4.3	Avalanche related data	6
4.4	Topographical Data	6
4.5	Data composition	6
	References	9

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 Methods

3.1 Feature selection

It is a challenging task in the field of data science to create machine learning models from high dimensional data sets. Machine learning research has long assumed that too many columns of data lead to a reduction in prediction quality. Therefore, it can help to focus on a small number of features [8]. This can be achieved by removing redundant and unimportant data columns to get enhance performance in learning efficiency and avoid overfitting. A good Feature Selection can help to get much better predictions from the machine learning models [8].

Depending on how the training set is labeled, supervised (fully labeled), unsupervised (unlabeled) and semi-supervised (partially labeled) feature selection methods are used [8]. Good feature selection methods should have a small time and space complexity and do not generate a lot of overhead, but must also have a high learning accuracy [8].

How important parameters were considered to be for the avalanches in each study seems to be strongly related to what parameters were available and wich machine learning models have

been used for the study. For example, in the study in Iran, which is described in 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 reported in the paper "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].

3.1.1 Decision Tree

"Decision Trees (DTs) are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A tree can be seen as a piecewise constant approximation." [9] Decision trees are a representation method based on knowledge about the features of a dataset to represent classification rules [10]. Decision Trees use a set of if-then-else rules to decide which value to predict. These models are good to understand, interpret and visualisable because they use white box models in which every step is a boolean logic and easy explainable [9]. A standard decision tree starts with a root node, does have some branches as well as child nodes and leaves [10]. The root node splits the set by a rule on the features which provides the best classification of the instance. This goes recursive till the max depth is reached or the classification is completed. So a branch is the path from the root node to the leaf. The leaf at the end of an branch, which represent the class labels of the feature to predict. [10]

They can also handle categorical data.

It's also possible to use them for multiple values to predict at the same time, which is a typical problem in supervised machine learning called the Multi-output problem [9]. In the case of this work it is to predict lots of parameters in context of avalanches by the use of topographical and Meteorological data. Because Decision Trees are likely to overfit if used on high dimensional datasets, they are no option to be used as an alternative prediction method for this work. But if used with a low tree depth, they can give a use a good understanding about the importance of some individual features for the prediction of multiple or specific parameters [9]. This advantages of decision trees make them also useful for feature selection.

3.2 Feature extraction

3.3 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] [11].

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].

3.3.1 SVM (Support Vector Machine)

3.3.2 MDA (multivariate discriminant analysis)

3.3.3 Neuronal Network

4 Data

4.1 Origin of the data

4.1.1 Data owner

The avalanche warning service of the energy company VERBUND AG, headquartered in Kaprun, Salzburg, provides the data required to carry out this work. VERBUND represents Austria's leading energy company and is Europe's largest producer of electricity from hydropower. The company states that it obtains almost 100% of its electricity from renewable sources. The electricity is mainly generated from hydro, wind and photovoltaic power plants. In addition, this is supported by gas-fired thermal power plants [12].

4.1.2 Recording area

The data were recorded in the vicinity of the Mooserboden storage power plant for 39 avalanche strokes. The background for the exceptionally accurate recordings of avalanches is the need for the most accurate possible prediction of avalanches in the 39 avalanche lines around the power plant. Accurate forecasting is of great importance to ensure the safety of the employees working in the areas of the avalanche lines. The power plant is part of the Kaprun power plant group, which includes both pumped and storage power plants. The power plant group is operated by Verbund Hydro Power and is located in Salzburg on the edge of the Hohe Tauern at 2040 meters above sea level and is surrounded by the over 3000 meter high mountains of the Glockner group.[13]

4.2 Meteorological Data

The Data Table Mooser_Wetter_Daten includes meteorological data for each day in the months of november to may in the period from 1953 to 2022. The table contains data about the air temperatures at the times 7:00, 14:00 and 19:00 as well as the snow temperature, the wind direction, the wind force, the snow sinking depth, the day weather as well as the weather from the day before, the snow depth, the precipitation, as well as the avalanche degree.

4.3 Avalanche related data

The Table Allg_Lawinen_Katalog represents general data on all recorded avalanches in the 39 avalanche lines of the area for the same period as the meteorological data from the table Mooser_Wetter_Data. More precisely, it contains data such as the type of avalanche, the old ID of the avalanche line where the avalanche went down, the time and date when the avalanche was recorded, the volume of the avalanche, the general weather conditions at the time of the avalanche, as well as general data on the snowpack and the danger level on the day of the avalanche. The Meteorological Data shown in this Table are not used in this study, because the data from the table Mooser_Wetter_Daten are homogeneous and available for each day of the winter season. Another table of the database named kaplawstr contains additional information about the avalanche lines. For this work, the old and the new code of the avalanche lines are used from this table

4.4 Topographical Data

The Topographical Data is recorded in the Database Table TOPP. This table contains several rows for each avalanche line, which can be identified by the new avalanche line code. The rows record minimum, maximum and mean slope, as well as the orientation and height of the slope. The table also contains various other data columns. These are not used in the further course of the work, since they cannot be assigned to the individual avalanche lines in general, but are connected with individual avalanches, which are not allocated to them in the context of this work.

4.5 Data composition

The database tables Allg_Lawinen_Katalog (Avalanche related data), kaplawstr (contains the old as well as the new avalanche line IDs), TOPP (Topographical Data) and Mooser_Wetter_Daten (Meteorological data) which are already described in the previous chapters were merged into a homogeneous data set in the context of this master thesis. This process is explained in detail in this chapter.

The tables include data from 1952 to 2022. the quality of data increased with the years. This can be shown above all by the fact that in the years before 1992 there are more empty data columns in some tables. In order to increase the homogeneity of the data, all data outside the period from 1992 to 2022 were removed from the database tables. Subsequently to this measure the kaplawstr table has been merged to the Allg_Lawinen_Katalog table using the old avalanche line ID. This adds to each avalanche the associated new avalanche code and avalanche name, which are used as additional ID. The connection is necessary because the TOPP table, which represents the topographic data for the avalanche routes, does not contain the old avalanche route ID. In the course of this step, all lines that were labeled with the avalanche line name "all avalanches" also have been removed. These are not included in the kaplawstr table, since this does not represent an exact departure of an avalanche in one of the avalanche lines, but only states that in many of the avalanche lines small avalanches have departed.

```

1 TOPP = kaplawstr['Code_neu'].apply(lambda x: TOPP.loc[TOPP['Lawinencode']
== x].mean())

```

Listing 1: calculation of TOPP data for every avalanche line

In the second step, the average values for all columns from the associated avalanches were calculated from the TOPP table for each avalanche line. The python code shown in Listing 1 demonstrates this process. By this measure, one row is created for each avalanche line. The table contained without this procedure a total of 1905 rows. In the default state, the table could not have been connected to the other data tables. Another way to get only one row per avalanche stroke would be to select a random value for the respective stroke. The reason for taking the average value is that there are not the same number of lines for all avalanche lines and the values of the individual lines per avalanche line do not differ greatly from each other. Thus, the average value represents the entirety of the lines per stroke consistently. The topographic data from the newly assembled TOPP table was then merged to the entire dataset using the new avalanche ID. Subsequently, these avalanches were assigned to the daily recorded meteorological data of the Mooser_Wetter_Data table by an outer join, so as result there is at least one row per day in the dataset. In cases where several large avalanches have occurred at the same day, the dataset contains one row per avalanche and each Includes the meteorological data for this day.

In order to train a machine learning algorithm for the prediction of avalanches for topographically defined slopes in conjunction with the meteorological data available for this work, the topographical data must also be mapped onto the days without avalanches. The algorithm needs this information, as the data set would otherwise only contain topographic data directly related to avalanches. This would mean, for example, that the slope inclination could not become a feature for the prediction.

```

1 for i in gesamtdf.index:
2     if(pd.isnull(gesamtdf['meanExpo'][i])):
3         sample = TOPP.sample(1)
4         gesamtdf['meanExpo'][i] = sample['meanExpo'].values[0]
5         gesamtdf['meanSlope'][i] = sample['meanSlope'].values[0]
6         gesamtdf['stdDevSlope'][i] = sample['stdDevSlope'].values[0]
7         gesamtdf['MinSlope'][i] = sample['MinSlope'].values[0]
8         gesamtdf['MaxSlope'][i] = sample['MaxSlope'].values[0]
9         gesamtdf['Altitude'][i] = sample['Altitude'].values[0]

```

Listing 2: mapping random sample lines of topographical data onto the rows of non avalanche days

The consequence of this is that the topographic data must also be mapped to the days without avalanches. Because these days are not connected to an avalanche line ID and an even distribution on the slopes on these days is required, random lines from the calculated mean values of the TOPP table were mapped to them. Listing 2 shows this process in form of the corresponding Python code.

The resulting dataset maps 6547 rows and 139 columns. Some of this columns are redundant, empty, sparsely filled or contain information which can not be used to train a machine learning algorithm. This requires the measure to remove all columns with these characteristics.

After dropping this features the Dataset includes 46 columns.

In addition, a new column was added to the dataset, which contains either a 1 in case of an avalanche or a 0 in case no avalanche has occurred. This column is added to make it possible to predict whether an avalanche will occur or not. To train a machine learning algorithm on predicting whether an avalanche will go down from a particular slope, all features that can be used to directly and without any other features determine whether an avalanche will descend must be removed from the dataset.

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] (2022), [Online]. Available: <https://scikit-learn.org/stable/modules/tree.html> (visited on 04/13/2022).
- [10] V. Sugumaran, V. Muralidharan, and K. Ramachandran, „Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing“, *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 930–942, 2007, ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymssp.2006.05.004>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0888327006001142>.
- [11] A. Pozdnoukhov, R. Purves, and M. Kanevski, „Applying machine learning methods to avalanche forecasting“, *Annals of Glaciology*, vol. 49, 107–113, 2008. DOI: 10.3189/172756408787814870.

- [12] (2022), [Online]. Available: <https://www.verbund.com/de-at/ueber-verbund> (visited on 04/01/2022).
- [13] (2022), [Online]. Available: <https://www.verbund.com/de-at/ueber-verbund/besucherzentren/kaprun> (visited on 04/05/2022).