Predicting avalanche risk in the Bavarian Alps

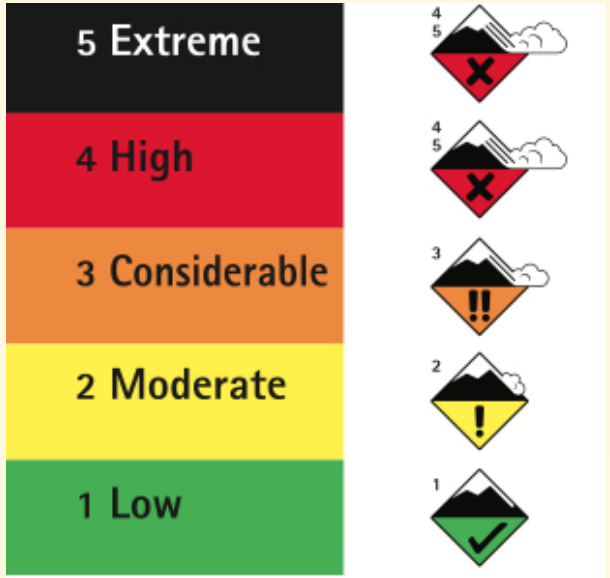
Daniel Wiegand

19.02.2021


Eingang

Avalanches

- Fatalities every year in the Bavarian alps
- Six risk monitoring regions from west (Allgäu) to east (Berchtesgaden)
- International avalanche danger scale with five levels



Project goals

- Predict regional avalanche danger level from weather data with more than 50% accuracy
- Identify the most important variables determining risk

Project structure

Data	EDA	Feature Engineering	Modeling
Webscraping	Data overview	Aggregation	Training
Data wrangling	Statistics	Imputation	Evaluation
Train-test-split	Plotting	Scaling	Hyperparameter optimization
		RFE	Feature importance
		SMOTE	

Beautiful Soup

seaborn

scikit learn

Keras

NumPy

pandas

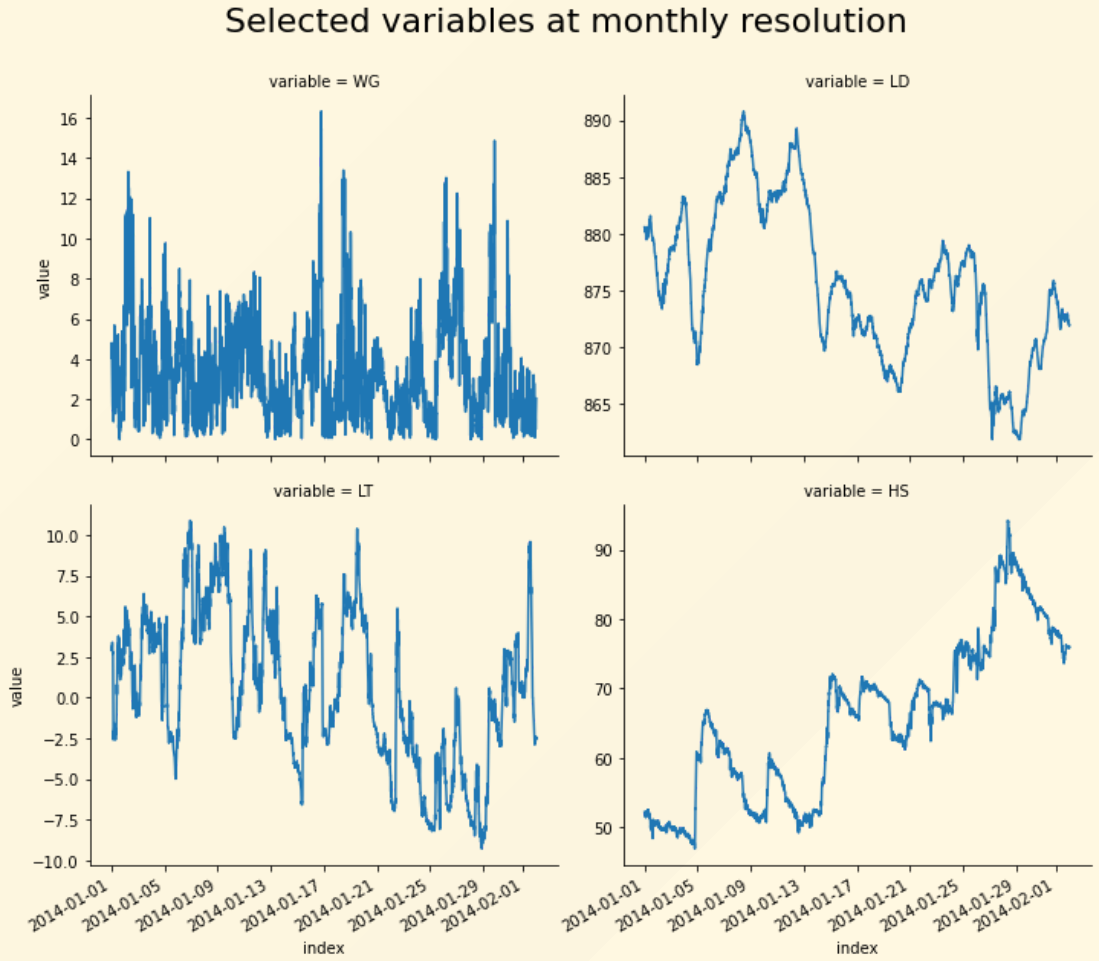
matplotlib

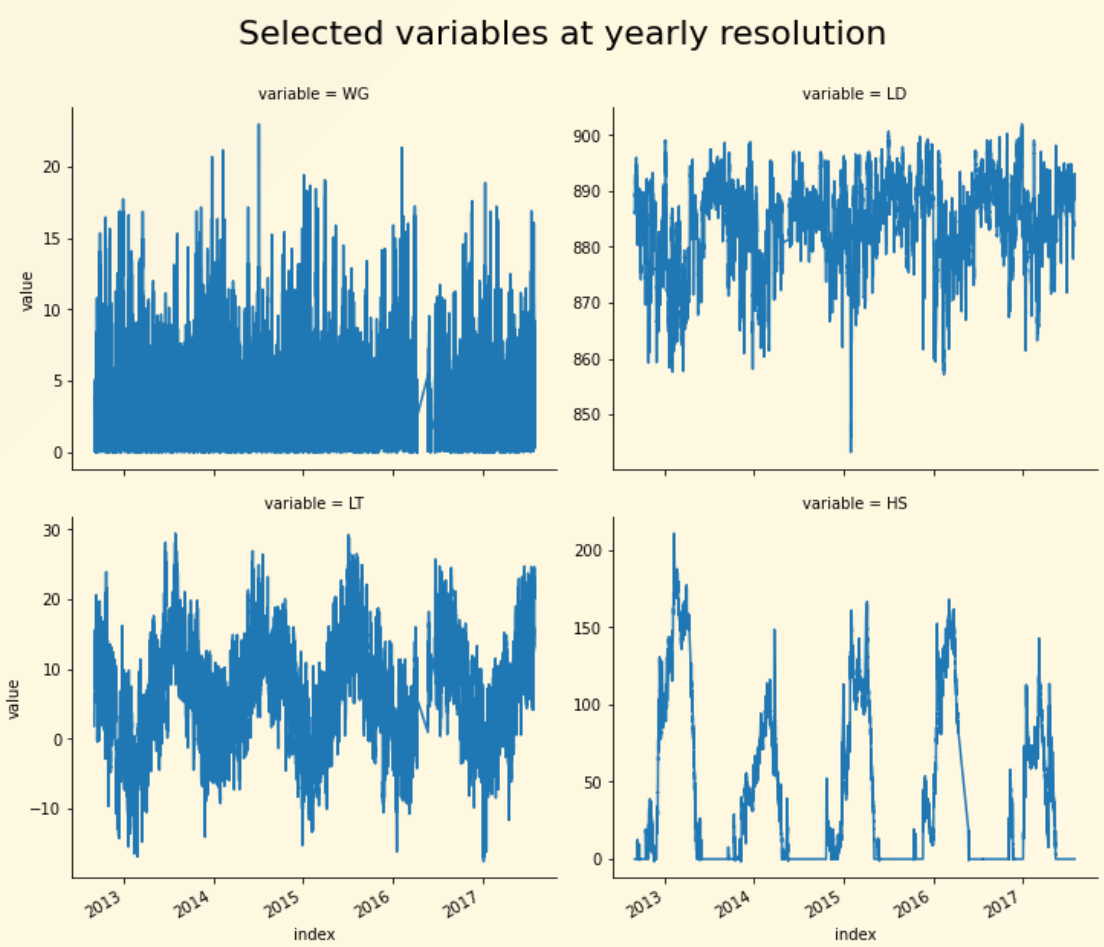
CatBoost

Data

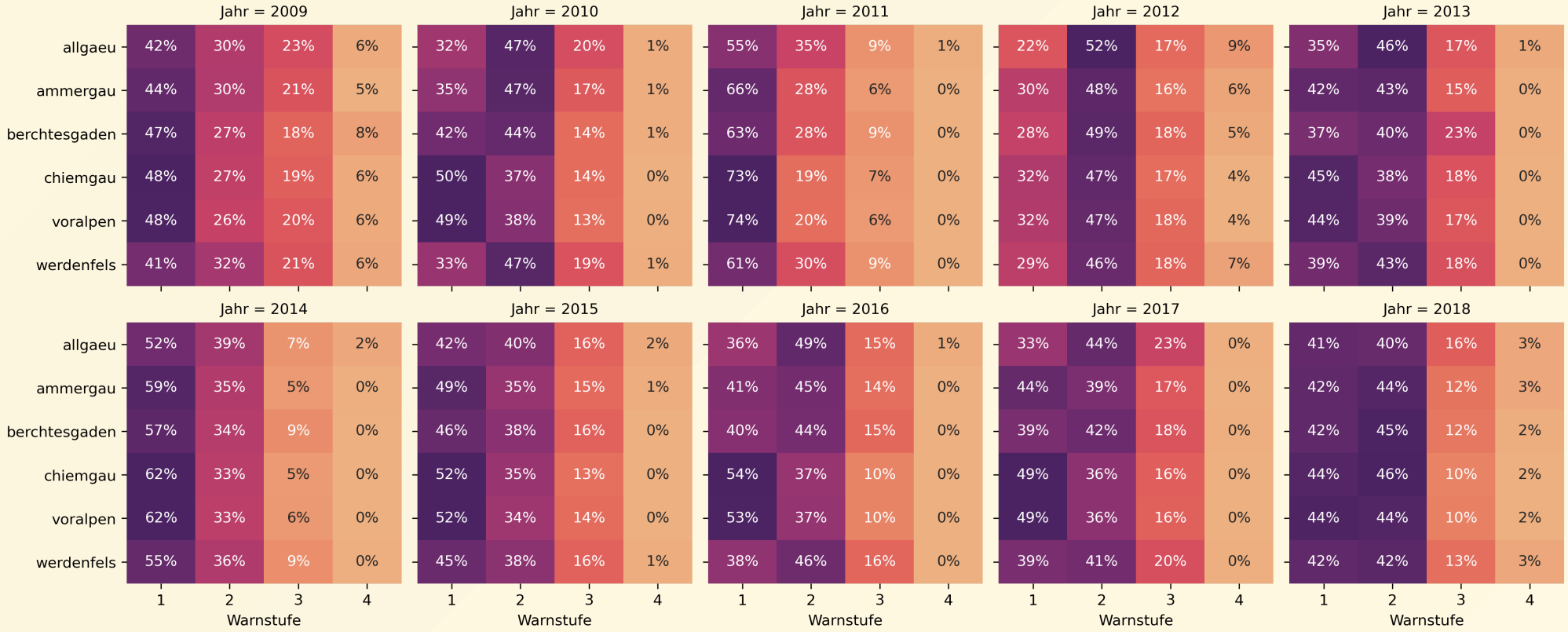
- Avalance danger levels per day from 2008 - 2018
- Original weather data from about 20 Alpine weather stations
  - Timespan: 2012 - 2018
  - Time resolution: 10 minutes
  - More than 25 variables

Weather data

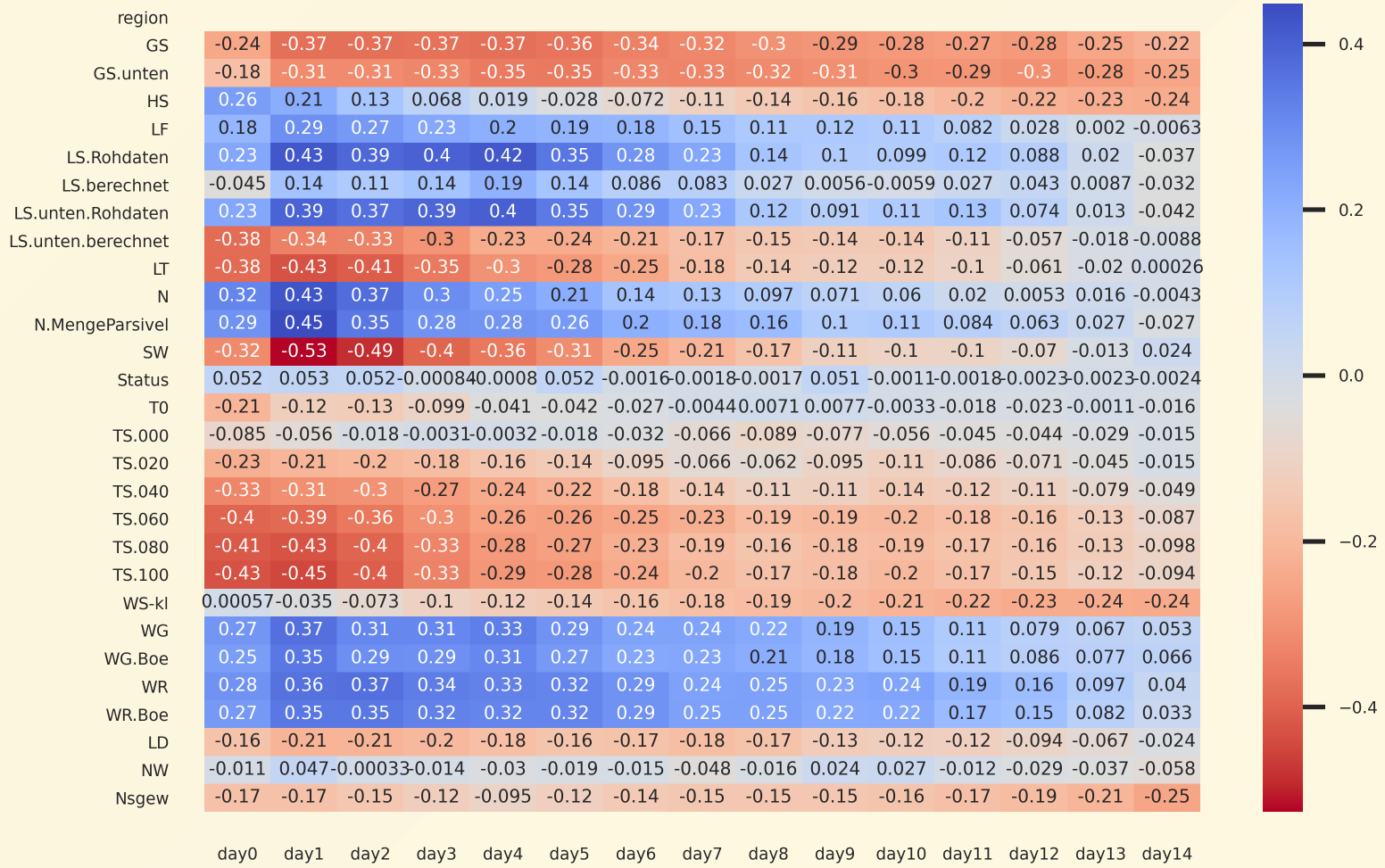




Warning levels



Relations to target variable



Baseline model

Model	Features	TimeShift	Param	Acc	ValAcc
NaiveBayes	all	1	S = 1e-9	0.50	0.50

```
class   precision  recall   f1-score   support

1.0        0.44      0.89       0.59       157
2.0        0.64      0.42       0.51       314
3.0        0.82      0.22       0.35       147
4.0        0.20      0.93       0.32        14
```

Possible improvements

- More advanced algorithms
- Include more time lags
- Feature selection
- Data upsampling (imbalanced target variable)
- Use data from other stations
- Account for autocorrelation

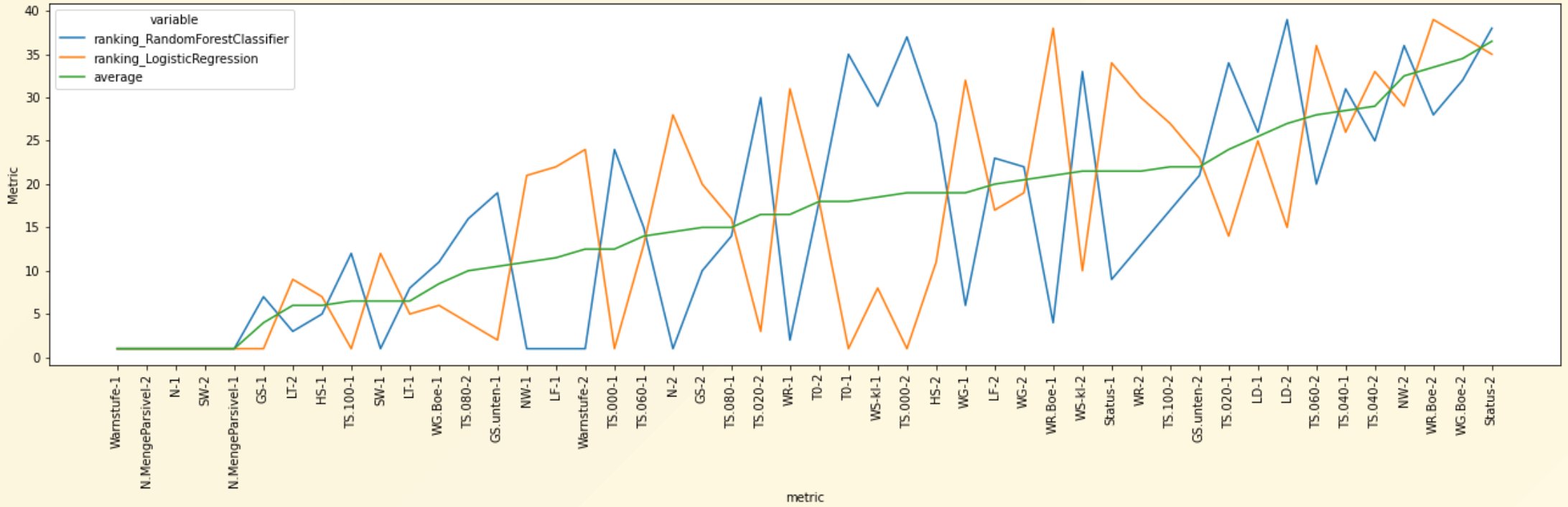
Improved models

Model	Features	TimeShift	SMOTE	Acc	ValAcc	F1_4
LogReg	all	1	no	0.67	0.63	0.55

Improved models

Model	Features	TimeShift	SMOTE	Acc	ValAcc	F1_4
LogReg	all	1	no	0.67	0.63	0.55
LogReg	all	1-2	no	0.71	0.60	0.70

Recursive Feature Elimination



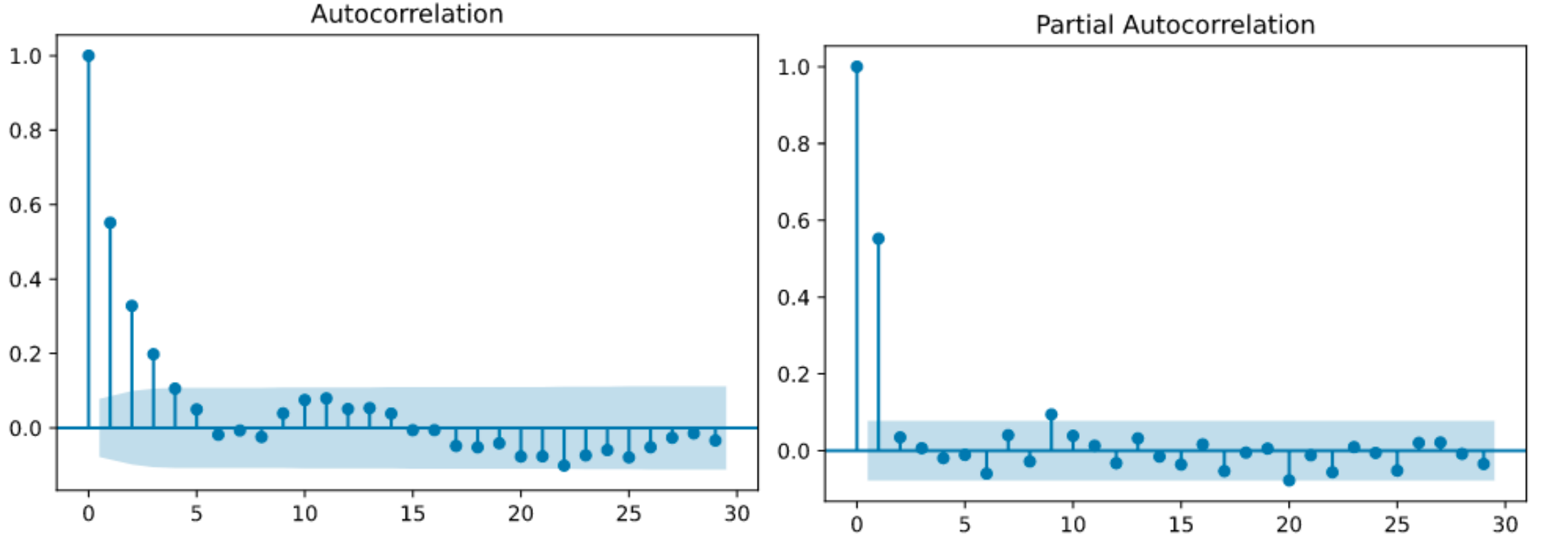
Improved models

Model	Feat	TimeShift	SMOTE	Autoc	Acc	ValAcc	F1_4
LogReg	all	1	no	no	0.67	0.63	0.55
LogReg	all	1-2	no	no	0.71	0.60	0.70
LogReg	rfe	1-2	no	no	0.68	0.60	0.70

Improved models

Model	Feat	TimeShift	SMOTE	Autoc	Acc	ValAcc	F1_4
LogReg	all	1	no	no	0.67	0.63	0.55
LogReg	all	1-2	no	no	0.71	0.60	0.70
LogReg	rfe	1-2	no	no	0.68	0.60	0.70
LogReg	rfe	1-2	yes	no	0.75	0.72	0.90

Autocorrelation of residuals



This looks like a AR(1) process!

Improved models

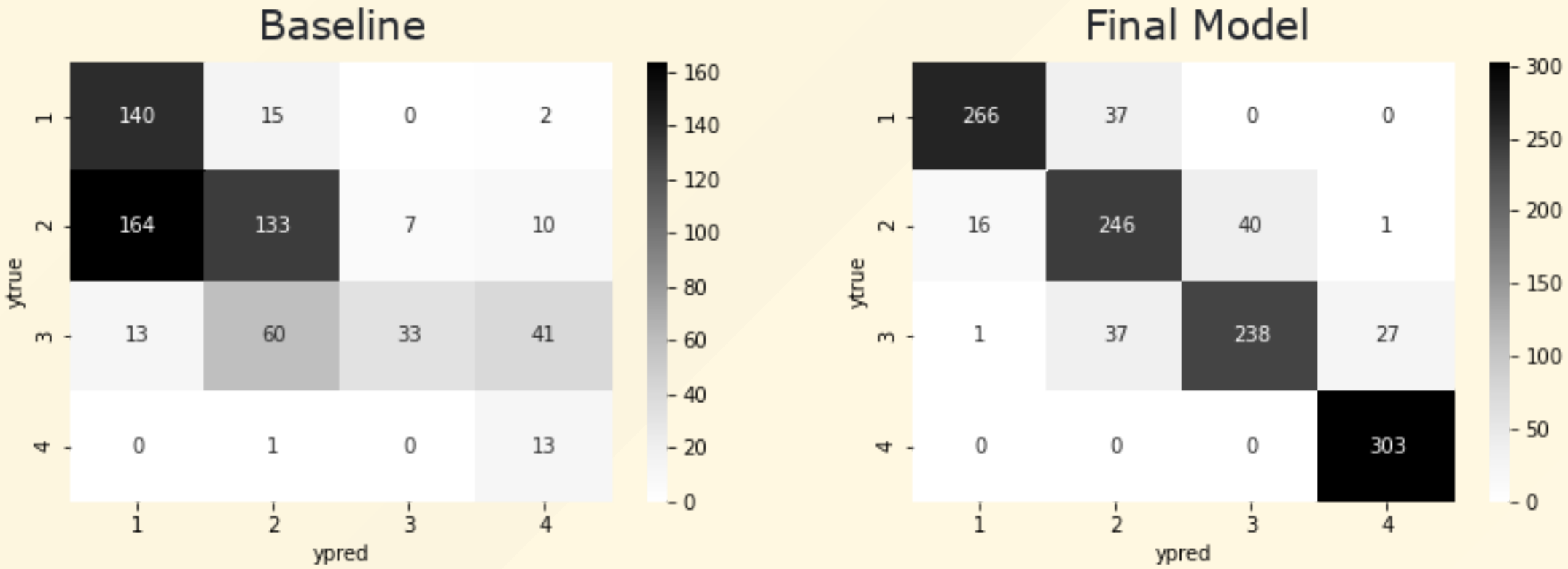
Model	Feat	TimeShift	SMOTE	Autoc	Acc	ValAcc	F1_4
LogReg	all	1	no	no	0.67	0.63	0.55
LogReg	all	1-2	no	no	0.71	0.60	0.70
LogReg	rfe	1-2	no	no	0.68	0.60	0.70
LogReg	rfe	1-2	yes	no	0.75	0.72	0.90
LogReg	rfe	1-2	yes	yes	0.85	0.80	0.96

Improved models

Model	Feat	TimeShift	SMOTE	Autoc	Acc	ValAcc	F1_4
LogReg	all	1	no	no	0.67	0.63	0.55
LogReg	all	1-2	no	no	0.71	0.60	0.70
LogReg	rfe	1-2	no	no	0.68	0.60	0.70
LogReg	rfe	1-2	yes	no	0.75	0.72	0.90
LogReg	rfe	1-2	yes	yes	0.85	0.80	0.96
RF	rfe	1-2	yes	yes	0.87	0.84	0.96

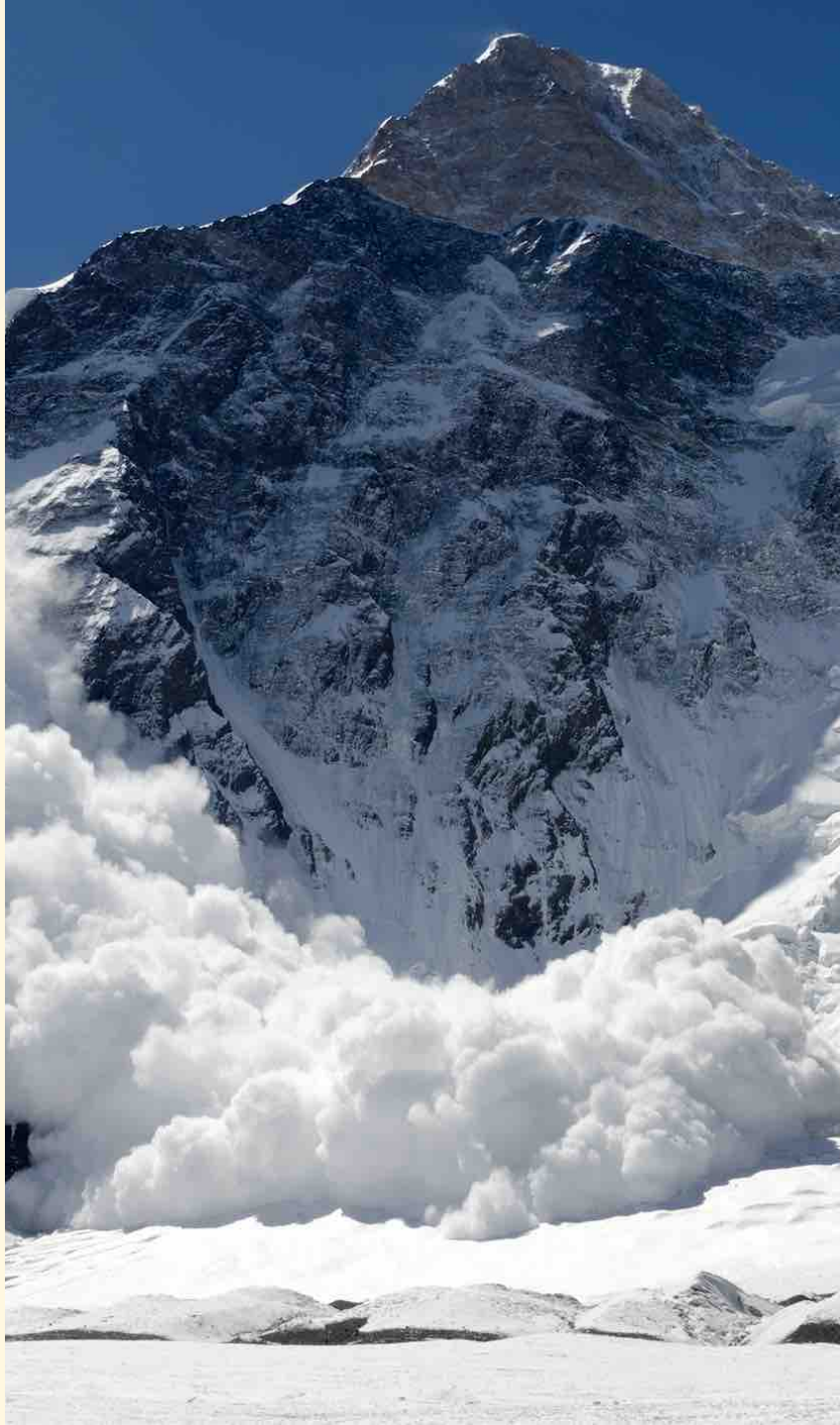
- Other models tried: SVM, FF-NN, Voting Classifier, Grad Boosting

Final model


Baseline
ytrue
ypred
1
2
3
4
140
15
0
2
164
133
7
10
13
60
33
41
0
1
0
13
160
140
120
100
80
60
40
20
0
Final Model
ytrue
ypred
1
2
3
4
266
37
0
0
16
246
40
1
1
37
238
27
0
0
0
303
300
250
200
150
100
50
0

Conclusion

- The avalanche warning level can be modeled from weather data alone with about 70% accuracy
- Taking into account autocorrelation, accuracy can be improved to 84%
- Most important variables seem to be precipitation of prior days, air / snow temperature, snow height, wind velocity



Thank you! :)