Multi-Task Learning for Ski Injury Predictions

Sandro Radovanović, Boris Delibašić, Milija Suknović
University of Belgrade – Faculty of Organizational Sciences
Jove Ilića 154, 11000 Belgrade, Serbia
{Sandro.radovanovic, boris.delibasic, milijas}@fon.bg.ac.rs

Abstract*. Predicting ski injuries is a very hard classification problem. This is due to the high class imbalance of injured vs. non-injured skiers and the lack of demographic information about skiers. Additional problems are the intrinsic properties of the ski lifts. Ski lifts differ in width, the difficulty degree, geographical position on the mountain etc. which results in different patterns for ski injury. In most researches, this information is not included. Aim of this paper is to develop multi-task classification models, which account for the uniqueness of ski lifts, taking into consideration information from other ski lifts. The proposed models were created on Mt. Kopaonik, Serbia ski resort and they show that ski injury in the following hour can be predicted with AUC ~0.64, or 3-4% better compared to the classical approaches.*

Keywords. Ski injury, logistic regression, multi-task learning

1 Introduction

Ski sports and leisure industry is considered as a big industry with steady 60 million average skier visits since 2002/2003 in the United States only. However, the United States is estimated to have 15% of worldwide skiers. Therefore, it is expected to have 400 million skier visits worldwide (The National Ski Areas Association, 2018). Additionally, every country with mountain regions tends to have ski resorts because they will be a major source of income and sustainability not only for the mountain region but for the whole country as well.

Although skiing is very popular, especially in the winter, the decision-making process in the ski industry is in most cases not data driven, and ski resorts often face problems with sustainability and profitability. This means that decisions are made based on views, opinions and experience of top management. Without adequate support of data-driven decision making it is nowadays not possible to reach the KPIs of an organization.

One way for improvement of the decision-making process in the ski industry is by inspecting the data about skiing behavior. This data is already available in most ski resorts. Namely, most ski resorts utilize RFID ski passes for entering ski lift gates. Based on this, a huge amount of data about skiers are generated which can be used for informed decision making.

One problem which is often analyzed in ski resorts is ski injury. This problem is considered as a highly challenging one (Ruedl et al., 2014; Bianchi et al., 2017) because incidence rate is very low. Namely, ski injuries are very rare events with 0.2% or fewer injuries per skier day. This number may seem low, but it is expected to have yearly 800,000 ski injuries worldwide, which is a major cost for insurance companies, and a global public health problem. Additionally, ski injuries come with a high cost, i.e. broken arm or leg, temporary or permanent movement disability and sometimes even death.

In this paper, we created a prediction model which predicts whether an injury will occur in the following hour on the ski lift on Mt. Kopaonik, Serbia. This research setup is important, as it is relevant to find-out real-time predictors of ski injury occurrence which could help in real-time prevention of injuries, and therefore reduce the occurrence of ski injuries. From a machine learning perspective, this problem presents a binary classification problem which we evaluated using the area under the ROC curve (AUC). An additional problem which is present in the data at hand is the fact that ski lifts are very different among themselves. Therefore, using all data for classification model could generate a too general model. On the other side, creating a classification model for each ski lift would require creating as many classification models as there are ski lifts (in this case 14). Additionally, we might not have enough data for some ski lifts as some ski lifts are less utilized, meaning that fewer data are available. As a solution, we propose multi-task classification (Pan & Yang, 2010) which generate multiple classification models at the same time, while exploiting similarities and differences between models, in this case, ski lifts. In other words, we want to extract and utilize shared information from all available ski lifts but also account for the differences due to the uniqueness of ski lifts. Since each ski lift is learned independently it is expected to have better performances compared to a model using all data. Additionally, since information is shared with other ski lifts it is expected to have better performance compared to independent classification models for each ski lift. We will train three different multi-task logistic regression models. Namely, Lasso regularized multi-task logistic regression, Trace norm regularized multi-task logistic regression and $l_{2,1}$ norm regularized multi-task logistic regression.

The aim of this paper is a proposal of multi-task logistic regression model for prediction of ski injuries. We argue that data mining and machine learning techniques could be beneficial for the whole ski industry, which means that effects could be seen not only by the ski resort but also by the final users of the ski resort. Also, information about possible ski injury could be of great help especially for mountain rescue service which can be used to reduce the time needed for intervention, or even act preventively.

The remainder of the paper is structured as follows. Section 2 provides a literature review on ski injury predictions. Section 3 provides a methodology of the research providing a brief description of the data, multi-task learning, and experimental setup. Section 4 presents results and discussion of the results, while Section 5 concludes the paper.

2 Literature Review

Ski injuries are most often analyzed in small-scale, case-control studies. The goal of such analyses is to compare injured skiers population to a small subsample of the non-injured population in order to discover what the differences were between subpopulations. As a result, one could see odds ratios or risk ratios for different properties of skiing. Those properties could be physical, i.e. gender (Ruedl et al., 2016a) and age (Ruedl et al., 2016b, Chamarro & Fernández-Castro, 2009). Properties could refer to skiing behavior, i.e. the speed of skiing (Ruedl et al., 2016a) and skiing experience (Hume et al., 2015). Also, properties could be ski lift related, i.e. quality of ski lifts (Chamarro & Fernández-Castro, 2009) and snow condition (Ruedl et al., 2013). Finally, properties could represent weather (Hume et al., 2015).

Although information about odds ratio or risk ratio are of interest and could be useful for injury prevention and education they could hardly be used in real-time analysis and decision making. To the best of our knowledge, the first data mining model for ski injury prediction is presented in (Bohanec & Delibašić, 2015). In that paper Decision Expert (DEX) model was combined with data mining model to predict global daily prediction of ski injuries. Namely, it combined domain knowledge with data to improve predictive performance. Although it dealt with regression problem remark that combination of expert modeling and data mining does improve the predictive performance of the learning algorithm. Further improvement using expert knowledge and data mining is presented in (Delibašić et al., 2018a), with a framework that allows extension of logistic regression models with DEX hierarchies. The framework is based on a stacking like approach to logistic regression. The proposed framework can be seen as a feature extraction model (resembling neural networks) where DEX model provides the structure. It has been shown that utilizing DEX models in a hierarchical manner in combination with the logistic regression improved performance of the predictive model. Both abovementioned papers present an introduction of knowledge into data mining and machine learning algorithms. The goal of knowledge is to enhance evidence available in the data. In this paper, we would like to enhance data mining models by sharing knowledge between classification models. This can be done using multi-task learning.

Another example of the application of data mining models for prediction of ski injuries is presented in (Delibašić et al., 2017a). The idea of the paper was that classical analysis is not suitable and that interaction of attributes are of great importance for the prediction model, namely for prediction of ski injuries. Therefore, logistic regression models are not suitable because they assume a linear dependency between attribute and ski injury. Therefore, the CHAID algorithm was used. It has been shown that performance of the data mining models was significantly better compared to univariate analysis and that performances of logistic regression and CHAID algorithm were comparable. However, CHAID decision tree model could be useful for identification of injury risk subpopulations because decision tree is much more interpretable compared to logistic regression.

Prediction whether a ski injury will occur or not is an information of high value. However, mountain ski rescue service could use information about an injury, i.e. what type of injury and what part of the body was injured in order to better allocate resources before the intervention. This further complicates classification model because instead of one label (whether an injury occurred or not) multiple are available (a type of the injury and part of the body). Therefore, the multi-label prediction must be applied. This is presented in (Radovanović et al., 2018). It has been shown that the performance of multi-label models could be utilized. However, some types of injuries and some part of the body are hard for prediction models. Namely, those are types of injuries and part of the body which are seldom injured.

Another interesting approach to ski injuries can be found in (Delibašić et al., 2017b). Instead of using classification models to predict whether an injury will occur or not we might use different data mining methodology which will return the same output (whether an injury will occur or not), namely recommender systems. Although this seems less intuitive it has been shown that predictive performance of recommender systems was comparable or better than data mining and machine learning algorithms.

Analysis of ski industry is not limited to ski injuries. One can find analysis and recommendations for ski lift pricing tickets, clustering of skiers etc. Interested readers in the analysis of the ski industry related researches are referred to (Delibašić et al., 2018a).

Authors observed that missing part of the ski injury prediction models is the account for unique characteristics of the ski lift, but also share information between ski lifts. Namely, every above-mentioned paper creates a classification model using data from all ski lifts. Because of that, we propose a classification model which is able to use information from other ski lift if available or data is missing, but also utilize the uniqueness of ski lift.

3 Methodology

In this section, we will present data, short explanation and motivation for multi-task and logistic regression, and experimental setup.

3.1 Data

The data used in this research was obtained from Mt. Kopaonik, Serbia ski resort. Mt. Kopaonik has 20 ski lifts with different degrees of difficulty, from which 14 are used in this research due to the lack of observed injury data on a specific lift on a specific hour. Data include all ski lift gate entrances from season 2005/2006 to season 2011/2012. Ski lift entrances are obtained using RFID checkouts of ski tickets which are needed in order to start skiing on the lift. In order to prevent confusion, term *skier* is used for all ski participants, i.e. skiers, snowboarders. Ski injuries are available in the other data source, namely, ski mountain rescue service database. These data sources are joined using ski ticket number and ski lift gate checkout. Finally, data about the weather is obtained from Republic hydro-meteorological service of Serbia. The whole dataset has over 20,000,000 observations.

The goal of the paper is to create a prediction model for ski injuries. Namely, we want to predict whether an injury will occur in the following hour on a ski lift. Because of that, data was aggregated on the hour level. Due to aggregation of the data instead of 20,000,000 observations we have 44,941 observations across 14 ski lifts and for each ski lift we have 19 input attributes, and one output attribute (label) which present binary signal whether an injury occurred in the following hour or not. Input attributes can be roughly divided into three categories. First set of attribute present ski lift crowd. Attributes which represent this group are the *hour* of observation, *number of ski lift checkouts* in one hour and *number of skiers* in one hour. The second group would present skier behavior on a ski lift. Attributes in this group are *average time on tracks* skier spent on ski resort up to that hour, *number of local maxima in average time on tracks* (Delibašić et al., 2017a), *average vertical distance* skier spent on ski resort up to that hour, *number of local maxima for vertical distance*, *average number of lifts* skier skied up to that hour, *number of local maxima of number of lifts*, *number of distinct lifts* skier skied up to that hour and *number of local maxima of distinct lifts*. The third group represents weather attributes. Those are *temperature*, *dew point*, *humidity*, *wind speed*, *visibility*, *fog*, *rain*, and *snow*.

It is worth to mention that class imbalance is present in data. This means that the majority of observation are non-injured. The overall class imbalance is 3.73%.

3.2 Multi-task logistic regression

Many data mining and machine learning applications are related to classification tasks. This means that same input attributes and same output attributes are used but with different observations. One instance of observation is called task. Most often task are related between themselves. The simplest approach is to solve these tasks independently, ignoring the relations between them. However, we would like to utilize relations and create a classification model simultaneously. This is called multi-task model learning. Difference between traditional and multi-task learning is presented in Figure 1.

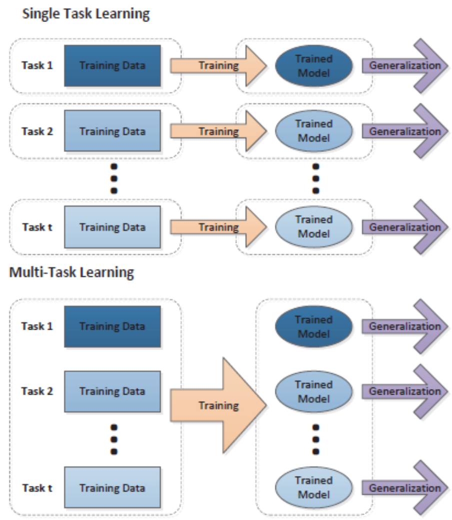

Figure 1. Illustration of traditional data mining model learning and multi-task model learning (Zhou et al., 2011)

A common mathematical model in data mining and machine learning is to minimize loss function and a regularization term. Namely, we want to:

$$\min_{W} L(W) + \lambda(W) \quad (1)$$

where W is the parameter to be estimated from the data, namely coefficients of logistic regression, $L(W)$ empirical loss from the data, namely logistic loss and $\lambda(W)$ regularization term with is used to restrict overfitting and in case of multi-task learning share knowledge between classification models. In this paper we will utilize three multi-task approaches.

The first multi-task approach is called Lasso regularized multi-task learning. Lasso regularization is one of the most popular method for reducing generalized linear model complexity (Tibshirani, 2011). Complexity is reduced by forcing some coefficients of logistic regression to be zero. In this setting zero coefficients will be distributed across the tasks. Namely, some tasks might have several zero coefficients while other tasks would have none. It is easily extended into multi-task formulation by providing a matrix of logistic regression coefficients instead of a vector (Zhou et al., 2011) providing the following mathematical model.

$$\min_{W,\ c} \sum_{i=1}^{t} \sum_{j=1}^{n_i} \log(1 + \exp(-Y_{i,j}(W_j^T X_{i,j} + c_i)) + \lambda ||W||_1 \qquad (2)$$

where $X_{i,j}$ present observation j of the task i, $Y_{i,j}$ output attribute j of the task i, W_j and c_i represent the logistic regression models. Parameter λ controls the strength of the regularization.

The second approach used in this paper is called trace norm regularized multi-task learning (Grave et al., 2011). The mathematical model can be presented as:

$$\min_{W,\ c} \sum_{i=1}^{t} \sum_{j=1}^{n_i} \log(1 + \exp(-Y_{i,j}(W_j^T X_{i,j} + c_i)) + \lambda ||W||_* \qquad (3)$$

where $Y_{i,j}$ present observation j of output attribute for task i, $X_{i,j}$ input attributes vector of a j-th row of task i, while W_j and c_i represent the logistic regression models. Finally, the trace norm is presented as $||W||_* = \sum_i \sigma_i(W)$. Parameter λ controls the strength of the regularization. This means that trace norm tries to capture the relationship between tasks by assuming that every classification model share a common low-dimensional subspace.

Finally, we utilized $l_{2,1}$ norm regularization. This regularization forces all tasks to share a common set of features (Nie et al., 2010). This can be interpreted that each task would try to reduce same set of coefficients to be zero. In order to get the following properties, one needs to solve the following mathematical model.

$$\min_{W,\ c} \sum_{i=1}^{t} \sum_{j=1}^{n_i} \log(1 + \exp(-Y_{i,j}(W_j^T X_{i,j} + c_i)) + \lambda ||W||_{2,1} \qquad (4)$$

where $Y_{i,j}$ present row j of output attribute for task i, $X_{i,j}$ input attributes vector of a j-th row of task i, while W_j and c_i represent the logistic regression models. Parameter λ controls the strength of the regularization.

Finding global minima of such models can be done in the same manner as in for simple logistic regression by using gradient descent. We expect that multi-task formulation would yield better performing models compared to baseline methods which will be explained in the Experimental setup section. Additionally, we expect Lasso logistic regression to be the best performing one because it allows different representation of logistic regression coefficients, i.e. one attribute is selected by one ski lift model, but not by the other ones. Also, we expect that $l_{2,1}$ norm obtain results better than baselines but lower compared to other multi-task since it forces all models to select the same attributes for predictions, i.e. it forces every ski lift to use same attributes.

Finally, each of the presented models needs to provide a probability of injury occurrence. Since every task is trained at the same time, we must update formula for obtaining log-odds. Therefore, we would use:

$$\log\left(\frac{p}{1-p}\right) = X_t * W_t^T \qquad (5)$$

This will give us for each task t a vector of predictions which can be converted to probabilities using the sigmoid function.

3.3 Experimental setup

In order to test our hypothesis that sharing knowledge between classification models using multi-task logistic regression would yield better predictive performance compared to single classification model per ski lift and model which uses all the data, we will use split validation of proposed models. Namely, we will use random 70% for model training, and the remaining 30% for model testing. The procedure is repeated 10 times and the average value of performance measure is reported alongside the standard deviation.

We will train two baseline models. Those are single model per ski lift and model using all available data. We expect that model which uses only data from that ski lift would obtain good performances only for those ski lifts which have a high number of examples and a high number of injuries while performing badly for ski lifts with a small number of observations and huge class imbalance. Also, we expect the model which utilizes all available data to be too general.

As a performance metric, we selected the area under the ROC curve (AUC) since it is a common binary classification measure. Additionally, AUC is a decision threshold independent measure which means that the value of AUC present the overall goodness of the model. AUC is calculated by calculating the true positive rate and false positive rate for every possible decision threshold available in the data and calculating the area under the curve which is created by those two values. However, it can be calculated more easily using Mann-Whitney U test. AUC ranges from 0 to 1, where 1 present perfect classifier, while the value of 0.5 present classifier which is equal to the random classifier.

An additional challenge is a selection of the regularization parameter λ. In order to get a best possible estimate of the parameter λ, we performed 10-fold cross validation on a specified vector of possible values of λ. Vector of possible values is implemented such that biggest value is smallest possible λ which yield intercept model only (all coefficient of logistic regression are equal to zero). Value is reduced by ~8 until λ reaches zero (James et al., 2013). Value λ which obtained the best AUC on inner 10 fold cross validation is then selected. One additional challenge comes from multiple tasks. Since we have 14 models we would obtain 14 AUCs. We will present the micro and macro values of AUCs. Micro value of AUC would account for a number of observations in the dataset by presenting weighted average, while the macro value of AUC would take average values of AUCs regardless of the number of observations in each task. In process of inner cross-validation, we need to select what will be optimized. We will have two results, one with optimized macro AUC and other with optimized micro AUC.

4 Results and Discussion

The experimental results are shown in Table 1 and Table 2. We report AUCs for baseline methods, which are all data logistic regression (AD-LR) and a logistic regression model for each ski lift which we can call task independent logistic regression (TI-LR), and also for multi-task models. Those are lasso logistic regression (MT-LN-LR), trace norm logistic regression (MT-TN-LR) and $l_{2,1}$ norm logistic regression (MT-L21-LR). Since the experiment is repeated 10 times we report the average value of AUC on the test set with the standard deviation, but also the lowest value and the highest value.

Table 1. Macro AUCs of the ski injury prediction models

Method	Average +/- S.D.	Min	Max
AD-LR	0.601 +/- 0.008	0.587	0.611
TI-LR	0.612 +/- 0.023	0.579	0.647
MT-LN-LR	**0.644 +/- 0.006**	**0.632**	0.657
MT-TN-LR	0.643 +/- 0.011	0.630	**0.661**
MT-L21-LR	0.641 +/- 0.010	0.627	0.655

As we can observe from Table 1 the best performing algorithm is multi-task Lasso regularized logistic regression with AUC 0.644, but other two multi-task algorithms performed similarly with 0.643 and 0.641 for Trace norm and $l_{2,1}$ norm, respectively. As expected, logistic regression using all data and task independent logistic regressions were 3% or 4% worse compared to multi-task algorithms. This is an indicator that sharing knowledge between tasks improve the performance of the algorithm. There are many examples where the similar finding is found, i.e. medicine (Zhou et al., 2012) and protein-protein interactions (Kshirsagar et al., 2013).

Table 2. Micro AUCs of the ski injury prediction models

Method	Average +/- S.D.	Min	Max
AD-LR	0.607 +/- 0.010	0.590	0.618
TI-LR	0.615 +/- 0.013	0.597	0.637
MT-LN-LR	0.642 +/- 0.011	**0.626**	0.658
MT-TN-LR	0.644 +/- 0.014	0.621	0.660
MT-L21-LR	**0.647 +/- 0.011**	0.625	**0.661**

Similar results are obtained using micro AUC (Table 2). The best performances are presented in bold letters. Performance is better on multi-task models compared to the baselines. However, Lasso norm logistic regression was the worst performing algorithm and $l_{2,1}$ norm was the best. Since, the difference is on the third decimal place we can state that performances are similar. Before going to further analysis we present information about ski lifts (Table 3).

Table 3. Information about ski lifts

Ski lift	# Observations	% Injuries
Centar	1735	4.84
Duboka 1	2436	4.68
Duboka 2	2881	6.42
Gobelja relej	1676	2.21
Gvozdac	1880	2.39
Karaman	2624	2.06
Karaman greben	3516	12.32
Kneževe bare	2341	1.84
Mali karaman	3244	8.69
Malo jezero	2884	1.70
Marine vode	2679	1.42
Mašinac	3216	1.74
Pančićev vrh	3110	3.95
Sunčana dolina	1763	1.87

Having abovementioned results in mind we would like to inspect performances on ski lifts and discuss effects of a number of observations and class imbalance on performances. This part of the analysis will be done on MT-LN-LR, MT-TN-LR, and MT-L21-LR methods because these methods performed well in the previous part of the experiment. The results are presented in Table 4 and 5 for macro AUC and micro AUC, respectively.

In Table 4 we can see average macro AUCs by ski lifts. Unfortunately, there aren't clear patterns about performances based on a number of observations and class imbalance.

Table 4. Macro AUC by ski lifts

Ski lift	MT-LN-LR	MT-TN-LR	MT-L21-LR
Centar	0.586	**0,609**	0,584
Duboka 1	0.664	0,664	**0,687**
Duboka 2	**0.663**	0,645	0,637
Gobelja relej	0.615	0,630	**0,655**
Gvozdac	0.651	**0,700**	0,674
Karaman	0.639	**0,649**	0,639
Karaman greben	**0.671**	0,658	0,664
Kneževe bare	0.648	**0,658**	0,634
Mali karaman	**0.714**	0,685	0,710
Malo jezero	**0.588**	0,559	0,544
Marine vode	0.613	0,613	**0,634**
Mašinac	0.587	0,584	**0,593**
Pančićev vrh	0.652	0,647	**0,663**
Sunčana dolina	0.681	0,675	**0,715**

Similar is observed in micro AUCs (Table 5).

Table 5. Micro AUC by ski lifts

Ski lift	MT-LN-LR	MT-TN-LR	MT-L21-LR
Centar	0.586	**0.609**	0.600
Duboka 1	0.664	0.664	**0.674**
Duboka 2	**0.644**	0.633	0.640
Gobelja relej	0.621	**0.665**	0.649
Gvozdac	0.659	0.654	**0.679**
Karaman	**0.670**	0.613	0.646
Karaman greben	0.655	**0.664**	0.659
Kneževe bare	**0.699**	0.684	0.648
Mali karaman	0.719	0.712	**0.728**
Malo jezero	**0.584**	0.558	0.551
Marine vode	0.618	0.624	**0.628**
Mašinac	0.572	0.581	**0.591**
Pančićev vrh	0.659	0.647	**0.660**
Sunčana dolina	0.658	**0.696**	0.685

However, we can observe that lower class imbalance (bigger percentage of ski injuries) do tend to have greater AUCs. Based on the performances we, unfortunately, cannot say which model is the best one.

Finally, we can present logistic regression coefficients for multi-task algorithms. Due to the big number of elements, these images are presented in the Appendix (Figure A1, Figure A2, and Figure A3). We can observe that Trace norm regularization included all available data and, therefore, every classification model has non-zero coefficients for every input attribute. Lasso norm regularization influenced the classification model to select only some of the input attributes. We can observe that the majority of features are zero, especially weather-related ones. A similar pattern is seen for $l_{2,1}$ norm. Since this norm forces classification models to select the same set of features we see that weather related features are omitted, i.e. coefficients are forces to be zero. This is an indicator that weather does not have an effect to ski injuries. More specifically, skiers do not tend to go skiing when the weather is bad. This finding is in accordance to the (Delibašić et al., 2017b). Also, we can observe that coefficient for average time on the track is negative. This is an indicator that the beginning of the skiing session can represent risky skiing behavior due to lack of preparation as noticed by (Hume et al., 2015).

5 Conclusion

This paper proposes a multi-task approach for ski injury prediction. This is motivated by the fact that ski lifts have different properties (width of adjacent slopes, difficulty of adjacent slopes etc.) and ski injury patterns may differ from ski lift to ski lift. However, knowledge about patterns should be exchanged from one ski lift to another ski lift, and even, between ski resorts. Therefore, we utilized three multi-task methods or regularizations. One of them is Lasso norm extension for multi-task setting, which forces coefficients of logistic regression to zero. This way algorithms are less prone to overfitting and therefore have greater generalizability. Another regularization used in this paper was Trace norm which tries to capture the relationship between tasks by assuming that every classification model shares a common low-dimensional subspace. Finally, $l_{2,1}$ norm forces tasks to select the same set of attributes. These methods were compared to baseline methods which are using all data for logistic regression and creating logistic regression for each ski lift independently.

Predictive performance of multi-task was better for 3% or 4% compared to baseline methods using macro AUC and micro AUC as a performance measure. Namely, AUCs of multi-task methods were ~0.64 for both macro and micro version of AUCs.

For a further research, we plan to employ class imbalance techniques in order to reduce class imbalance and to create decision tree multi-task algorithms.

References

Bianchi, G., Brügger, O., & Niemann, S. (2017). Skiing and snowboarding in Switzerland: Trends in injury and fatality rates over time. In *Snow Sports Trauma and Safety* (pp. 29-39). Springer, Cham.

Bohanec, M., & Delibašić, B. (2015, May). Data-mining and expert models for predicting injury risk in ski resorts. In *International Conference on*

Decision Support System Technology (pp. 46-60). Springer, Cham.

Chamarro, A., & Fernández-Castro, J. (2009). The perception of causes of accidents in mountain sports: a study based on the experiences of victims. *Accident Analysis & Prevention*, *41*(1), 197-201.

Delibašić B., Radovanović, S., Jovanović, M., & Suknović, M. (2018b, May). Improving decision making in ski resorts by analysing ski lift transportation. In *Proceedings of the XIII Balkan Conference on Operational Research – BALCOR 2018* (pp. 16). Belgrade, Serbia, 2018.

Delibašić, B., Radovanović, S., Jovanović, M., Bohanec, M., & Suknović, M. (2018a). Integrating knowledge from DEX hierarchies into a logistic regression stacking model for predicting ski injuries. *Journal of Decision Systems*, 1-8.

Delibašić, B., Radovanović, S., Jovanović, M., Obradović, Z., & Suknović, M. (2017a). Ski injury predictive analytics from massive ski lift transportation data. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology* (2017), 1754337117728600.

Delibašić, B., Radovanović, S., Jovanović, M., Vukićević, M., & Suknović, M. (2017b). An Investigation of Human Trajectories in Ski Resorts. In *International Conference on ICT Innovations* (pp. 130-139). Springer, Cham.

Grave, E., Obozinski, G. R., & Bach, F. R. (2011). Trace lasso: a trace norm regularization for correlated designs. In *Advances in Neural Information Processing Systems* (pp. 2187-2195).

Hume, P. A., Lorimer, A. V., Griffiths, P. C., Carlson, I., & Lamont, M. (2015). Recreational snow-sports injury risk factors and countermeasures: a meta-analysis review and Haddon matrix evaluation. *Sports Medicine*, *45*(8), 1175-1190.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 112). New York: Springer.

Kshirsagar, M., Carbonell, J., & Klein-Seetharaman, J. (2013). Multitask learning for host–pathogen protein interactions. *Bioinformatics*, *29*(13), i217-i226.

Nie, F., Huang, H., Cai, X., & Ding, C. H. (2010). Efficient and robust feature selection via joint $\ell 2$, 1-norms minimization. In *Advances in Neural Information Processing Systems* (pp. 1813-1821).

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345-1359.

Radovanović, S., Delibašić, B., & Suknović, M. (2018). A multilabel prediction model for predicting part of the body and type of ski injury. In *Proceedings of the 4th International Conference on Decision Support System Technology – ICDSST 2018 & PROMETHEE DAYS 2018* (pp. 41). Heraklion, Greece.

Ruedl, G., Benedetto, K. P., Fink, C., Bauer, R., & Burtscher, M. (2016b). Factors associated with self-reported failure of binding to release among recreational skiers: an epidemiological study. *Current Issues in Sport Science* (CISS).

Ruedl, G., Helle, K., Tecklenburg, K., Schranz, A., Fink, C., & Burtscher, M. (2016a). Factors associated with self-reported failure of binding release among ACL injured male and female recreational skiers: a catalyst to change ISO binding standards?. *British Journal of Sports Medicine*, *50*(1), 37-40.

Ruedl, G., Kopp, M., Sommersacher, R., Woldrich, T., & Burtscher, M. (2013). Factors associated with injuries occurred on slope intersections and in snow parks compared to on-slope injuries. *Accident Analysis & Prevention*, *50*, 1221-1225.

Ruedl, G., Philippe, M., Sommersacher, R., Dünnwald, T., Kopp, M., & Burtscher, M. (2014). Current incidence of accidents on Austrian ski slopes. *Sportverletzung Sportschaden: Organ der Gesellschaft fur Orthopadisch-Traumatologische Sportmedizin*, *28*(4), 183-187.

The National Ski Areas Association. National skier/snowboarder visits, 1979–2016 (2018). Retrieved from http://www.nsaa.org/media/275017/1516_visits.pdf

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(3), 273-282.

Zhou, J., Chen, J., & Ye, J. (2011). MALSAR: Multi-task learning via structural regularization. *Arizona State University*, 2011.

Zhou, J., Liu, J., Narayan, V. A., & Ye, J. (2012, August). Modeling disease progression via fused sparse group lasso. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1095-1103). ACM.

Appendix

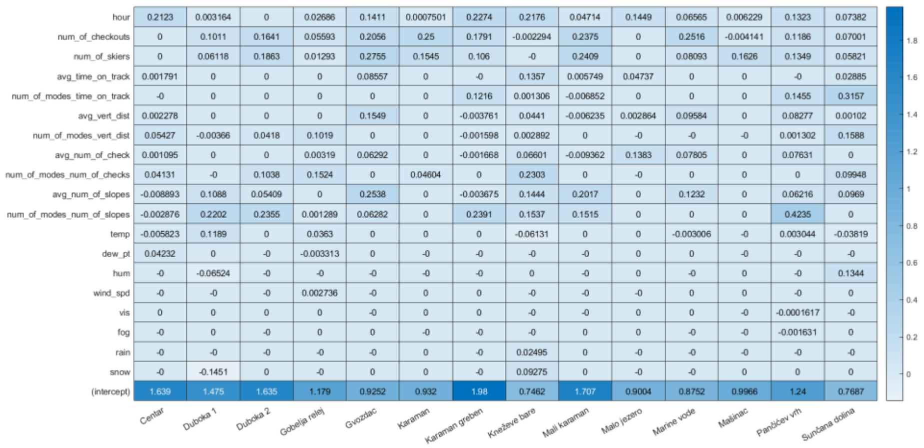

Figure A1. Coefficients of MT-LN-LR

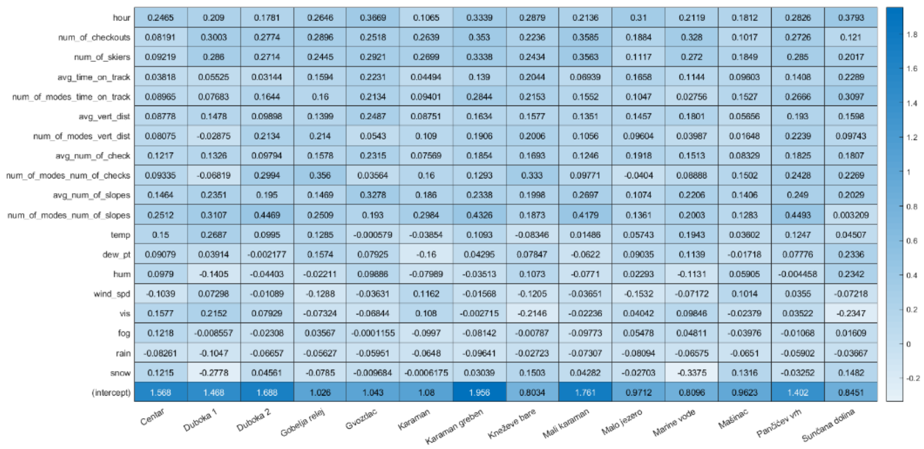

Figure A2. Coefficients of MT-TN-LR

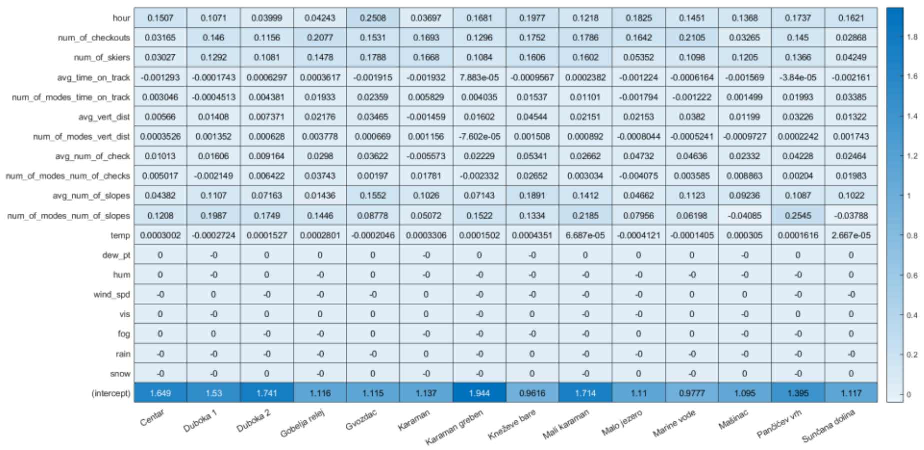

Figure A3. Coefficients of MT-L21-LR