

# Ski Resorts Recommendation using Deep Neural Networks

YASMINE SERDOUK, TIMOTHÉE COUBLE, ERIC COUBLE, and CÉDRIC MARCONE, Valraiso - e-Commerce Web Agency, France

In this work, we present a recommender system based on deep neural networks for a novel application that predicts suitable ski resorts for a web agency customers. A special focus is paid to the features construction where they are created from historical heterogeneous data according to a user-friendly questionnaire. The evaluation is performed through the accuracy metric and the obtained results show a promising performance with a TOP-10 accuracy that reaches 98.95% and 99.30% for French and non-French customers, respectively.

CCS Concepts: • **Computing methodologies** → *Artificial intelligence*; **Machine learning approaches**.

Additional Key Words and Phrases: neural networks, deep learning, recommender systems, ski resorts

## 1 INTRODUCTION

Traditionally, Recommender Systems (RS) are designed to predict missing ratings and provide accurate top-N recommendation for queries. However, traditional techniques only use user ratings for the different items and can not exploit feature vectors that can better capture the complex relations between the user and items. Thus, deep learning has been changing dramatically recommendation architectures lately [6, 8] with MultiLayer Perceptron (MLP) [2, 7], AutoEncoder (AE) [1, 13], Convolutional Neural Networks (CNN) [9] and other hybrid methods [11, 14].

This work proposes a novel RS deep learning based for a web business company in order to provide their customers with the best online experience. The web company, called Valraiso makes commercial websites for the French Ski School (ESF: École de Ski Françaises) and this work aims at recommending suitable Ski resorts for the customers visiting these websites. In other words, we want to develop an online interface that communicates with clients to determine their needs and collect data, such as the budget, the customer's location, etc. Then, on the basis of these data and the history of other previously collected orders, we build a model through deep learning techniques in order to recommend suitable ski resorts. In the literature, some works related to tourism have been explored like in [4] where authors analyse ski-lift transportations data to shape the decision making process in order to improve safety in ski resorts. Another interesting work treats about RS for the travel and hospitality industries with a skimatcher application [3]. Also, a case study was carried out in [5] about a constraint-based RS that was integrated into a travel advisory system for an Austrian spa resort.

The rest of this paper is organized as follows. Section 2 presents the dataset that we created and the process we went through for collecting and cleaning it. Section 3 is for to the experimental results. Finally, main conclusions and future works are drawn in Section 4.

## 2 DATASET CONSTRUCTION

The used dataset is named **SRESP** for "Ski RESorts Prediction". Initially, the Valraiso company has archived various heterogeneous data since its creation. In order to know which data to select, we first define the questionnaire.

### 2.1 Define the questions

In a real time context, a questionnaire interacts with the customer and asks few questions in order to recover his needs and desires. From the customer's responses, we reconstruct a feature vector that will be introduced into the prediction

model for recommendation. In order to allow a user-friendly interface, we need to have few questions and they need to be concise and allow simple answers divided into categories as follows.

- (1) What kind of skier(s) are you ? (In the case of a group or a family, please specify the average level)  
*Beginner, Intermediate, Confirmed, or Expert*
- (2) What type of ski area suits you best ? *Small, Medium, Large, or Very Large resorts,*
- (3) How high would you like to ski ? *Low, Average, High, or Very high altitude,*
- (4) What week do you plan to go ?
- (5) What is your budget level for this trip ? *Low, Average, or No limit cost*
- (6) Where do you live ?
- (7) How old are your children ? *Under 6 years old, Between 6 and 12 years old, More than 12 years old, or No children*

## 2.2 Data Selection and Creation

From the archived data of Valraiso, we select the most relevant features that can contribute to the choice of a ski resort and which can be used to reconstruct the responses of the defined questionnaire. The dataset is organized, such that each row is a client order for a specific ski resort while the columns represent the features. The created features are:

2.2.1 **Group.** This feature regroup four information about the nature of the group with different numeric values.

- The existence of at least one child under 6 years old: numeric value **3**
- The existence of at least one child aged between 6 and 12: numeric value **5**
- The existence of children in both age categories, which means aged under 12: numeric value **8**
- Or no children within the order, which implies only teenagers or adults in the order: numeric value **0**

2.2.2 **Budget.** As most orders are for at least one week, we include the weekly pass fees and the total cost in euros within this feature. Also, as the *daily\_pass* information is for one person (an adult), it is multiplied by the number of people included in the order.  $Budget = (5 \times daily\_pass \times nb\_people) + totalEuro$

Since the questionnaire allows only few categories of answers, feature values are categorized.

- **Low cost** category  $\in ]0, 300]€$  having a numeric value **0**, **Average cost** category  $\in ]300, 400]€$  having a numeric value **1**, and a **No limit cost** category  $\in ]400, +\infty[€$  having numeric value **2**.

The range of each category of the selected features is chosen such that the data are fairly distributed over each group. Each category is converted into a unique numeric value.

2.2.3 **Delta\_week.** The delta between the stay week and the validation week (i.e. the week the order was placed) is important and can influence greatly the choice of the ski resort.  $Delta\_week = stay\_week - validation\_week$

2.2.4 **stay\_week.** In addition to the delta order, the stay week has an important influence on the choice of the resort

2.2.5 **altitude\_max.** Some customers like to ski at very high altitude while others (that have children for instance) would choose resorts with a low or an average altitude. Four categories representing each of:

- **Low altitudes**  $\in ]0, 2100]$  meters having a numeric value **0**, **Average altitudes**  $\in ]2100, 2550]$  meters having a numeric value **1**, **High altitudes**  $\in ]2550, 2800]$  meters having a numeric value **2**, **Very high altitudes**  $\in ]2800, +\infty[$  meters having a numeric value **3**

Table 1. Composition of the dataset.

dataset	# classes	# samples at training	# samples at test	total number of samples
<b>SRESP_fr</b>	147	916451	392766	1309217
<b>SRESP_nonFr</b>	147	160365	68729	229094

2.2.6 **slopes\_difficulty**. This feature regroup the number of green ( $nb_g$ ), blue ( $nb_b$ ), red ( $nb_r$ ) and black ( $nb\_black$ ) ski slopes to create a new feature that gives a notion of the resort’s difficulty and though, infer the average level of skiers in the order. Thus, it is calculated so as to slightly weight  $nb_g$  and  $nb_b$  and to heavily weight  $nb_r$  and  $nb\_black$  as follows.  $slopes\_d = (0.1 \times nb_g) + (0.2 \times nb_b) + (0.6 \times nb_r) + nb\_black$

Then, four categories are created.

- **Beginner level** having a  $slopes\_d$  between  $]0, 6.2]$  with a numeric value **0**, **Intermediate level** having a  $slopes\_d$  between  $]6.2, 14.7]$  with a numeric value **1**, **Confirmed level** having a  $slopes\_d$  between  $]14.7, 22.4]$  with a numeric value **2**, **Expert level** having a  $slopes\_d$  between  $]22.4, +\infty[$  with a numeric value **3**,

2.2.7 **km\_slopes**. This feature represents the size of the ski area of the resort in kilometers.

- **Small ski resorts** having  $km\_slopes$  between  $]0, 70]$  kilometers with a numeric value **0**, **Medium ski resorts** having  $km\_slopes$  between  $]70, 146]$  kilometers with a numeric value **1**, **Large ski resorts** having  $km\_slopes$  between  $]146, 160]$  kilometers with a numeric value **2**, **Very large ski resorts** having  $km\_slopes$  between  $]160, 250]$  kilometers with a numeric value **3**

2.2.8 **country\_id**. The country from which the order has been made is converted into a numeric value. The dataset contains 174 different countries, the numeric values vary from 0 to 173.

2.2.9 **dep**. This feature is only dedicated to customers who ordered from France. Indeed, as more than 85% of the orders in the dataset are made from French customers and according to the market experts, the French and non French customer’s behaviours are different, we decided to build two models. One for the French customers and the other for the remainder countries. Therefore,  $dep$  depicts for the department of French customers.

### 2.3 SRESP construction

Note that two different inputs are presented into the machine learning. The first is for French and the second for non-French customers.

$[Group, stay\_week, Delta\_week, Budget, altitude\_max, slopes\_diff, km\_slopes, dep],$

$[Group, stay\_week, Delta\_week, Budget, altitude\_max, slopes\_diff, km\_slopes, country\_id],$

On the other hand, Table 1 details the composition of the dataset. It can be downloaded from [12]. We split **SRESP** into two subsets: training and test. As our dataset has unbalanced classes, the train-test split is made according to the number of samples on each class, i.e. 70% of the samples of each class are used for training while the rest are for the evaluation. Finally, as the feature values have widely different ranges, a min-max scaling is done in the range of  $[0-1]$  for a faster convergence [10].

Table 2. Accuracy (%) results using DNN for both French and non-French data.

TOP Accuracy	TOP-1	TOP-2	TOP-3	TOP-4	TOP-5	TOP-6	TOP-7	TOP-8	TOP-9	TOP-10
French model	52.15	72.95	83.24	89.11	92.77	95.11	96.66	97.75	98.45	98.95
non-French model	56.50	77.54	88.22	93.14	96.04	97.76	98.67	98.95	99.14	99.30

### 3 EXPERIMENTAL RESULTS

In this work, the parameters are selected through a 5-fold cross validation (5 different train-test splits). On the other hand, note that reducing the feature inputs, reduces the information variability and complexity to be introduced into the model. Also, categorizing the feature values, further reduces the variability within the values of our inputs. Therefore, the challenge is to have a strong architecture that can deal with such inputs. Finally, our problem can be summed up into a supervised classification problem that aims to predict resorts for new customers. The evaluation is carried out by computing the accuracy metric that assesses overall, how often our model is correct. It is presented for both French and non-French models.

After many experiments, the selected architecture has four hidden layers with 550 nodes each. A ReLU function is placed after each layer, except for the last layer where a softmax function is employed. Then, a batch normalization is employed to normalize each layer's inputs. Lastly, a dropout layer is added to prevent over-fitting by ignoring 25% of neurons during training. The training is performed through 1000 epochs for the French data and only 100 epochs for the non-French data. The Adam Optimizer is used with a batch size of 6360 and an initial learning rate of 0.003 that is decreased using the exponential decay:  $lr = lr_0 * \beta^{(step/\gamma)}$ , where  $lr_0$  is the initial learning rate,  $step$  is the current step, and  $\beta, \gamma$  are two parameters set to  $\beta = 0.96$  and  $\gamma = 90000$ .

The DNN results are presented in Table 2 for French and non-French models. Furthermore, since we want to recommend more than one ski resort, we present the accuracy results from TOP-1 to TOP-10. Note that the TOP-n accuracy corresponds to the accuracy where the true class matches with any one of the n most probable classes predicted by the model. Obviously, the accuracy increases through the TOP results. The results show a promising TOP-10 performance with 98.95% and 99.30%, respectively. Another thing to note is that the non-French accuracies are always higher than those with French data. This can be explained by the fact that in the non-French input vector we have the *country\_id* feature instead of the *department* feature in the French input. Indeed, since we have 171 countries and only 99 different French departments, we think that the countries variability helps more in discriminating between the different classes (i.e. resorts). Also, one can note that the TOP-1 accuracy remains rather low, which can be explained by the constraint of the questionnaire which allows little variability in the data.

### 4 CONCLUSION

In this paper, the proposed system learns various feature vectors that describe important information specific to both customers and target resorts. The obtained results showed promising results using the proposed architecture of DNN with a TOP-10 accuracy of 98.95% and 99.30% for French and non-French data, respectively. As further improvements, we plan to make the customer online experience better by recommending through a chat bot a full stay in the resort.

### REFERENCES

- [1] Bing Bai, Yushun Fan, Wei Tan, and Jia Zhang. 2020. DLTSR: A Deep Learning Framework for Recommendations of Long-Tail Web Services. *IEEE Transactions on Services Computing* 13, 1 (Feb. 2020), 73–85. <https://doi.org/10.1109/TSC.2017.2681666>

- [2] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. Association for Computing Machinery, Boston, USA, 191–198.
- [3] Joaquin Delgado and Richard Davidson. 2002. Knowledge Bases and User Profiling in Travel and Hospitality Recommender Systems. In *Proceedings of the 9th International Conference on Information Technologies in Tourism (ENTER 2002)*. Springer, Innsbruck-Austria, 1–16.
- [4] Boris Delibasic, S. Radovanović, M. Jovanovic, and Milija Suknovic. 2020. Improving Decision-Making in Ski Resorts by Analysing Ski Lift Transportation—A Review. In *Advances in Operational Research in the Balkans*. Springer, 265–273.
- [5] Jannach Dietmar, Zanker Markus, and Fuchs Matthias. 2009. Constraint-Based Recommendation in Tourism: A Multiperspective Case Study. *Information Technology and Tourism* 11 (2009), 139–155.
- [6] Travis Ebesu and Yi Fang. 2017. Neural Citation Network for Context-Aware Citation Recommendation. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*. Association for Computing Machinery, Tokyo, Japan, 1093–1096.
- [7] Jianfeng Gao, Patrick Pantel, Michael Gamon, Xiaodong He, and Li Deng. 2014. Modeling Interestingness with Deep Neural Networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP '14)*. Association for Computational Linguistics, Doha, Qatar, 2–13.
- [8] Xiangnan He and Tat-Seng Chua. 2017. Neural Factorization Machines for Sparse Predictive Analytics. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*. Association for Computing Machinery, Tokyo, Japan, 355–364.
- [9] Donghyun Kim, Chanyoung Park, Jinoh Oh, Sungyoung Lee, and Hwanjo Yu. 2016. Convolutional Matrix Factorization for Document Context-Aware Recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. Association for Computing Machinery, Boston, USA, 233–240.
- [10] GOPAL KRISHNA PATRO and Kishore Kumar Sahu. 2015. Normalization: A Preprocessing Stage. *International Advanced Research Journal in Science, Engineering and Technology* 2, 3 (March 2015), 20–22.
- [11] Maria Nadia Postorino and Giuseppe M. L. Sarnè. 2010. A Neural Network Hybrid Recommender System. In *The 20th Italian Workshop on Neural Nets*, 180–187.
- [12] skiResortData 2021. *Ski Resort dataset*. Retrieved September 14, 2021 from <https://www.kaggle.com/serdoukyasmine/data-frenchskiresorts>
- [13] Florian Strub, Romaric Gaudel, and Jérémie Mary. 2016. Hybrid recommender system based on autoencoders. In *Proceedings of the 1st Workshop on Deep Learning for Recommender System (DLRS '16)*. Association for Computing Machinery, Boston, USA, 11–16.
- [14] Muhammet Çakır, Şule Gündüz Ögüdücü, and Resul Tugay. 2019. A Deep Hybrid Model for Recommendation Systems. In *International Conference of the Italian Association for Artificial Intelligence*. Springer, 321–335.