Ski Resorts Recommendation using Deep Neural Networks

Y. Serdouk, T. Couble, E. Couble, and C. Marcone
Valraiso, Web Agency, Lyon France

RecTour

Sunday, September 26 2021

Introduction

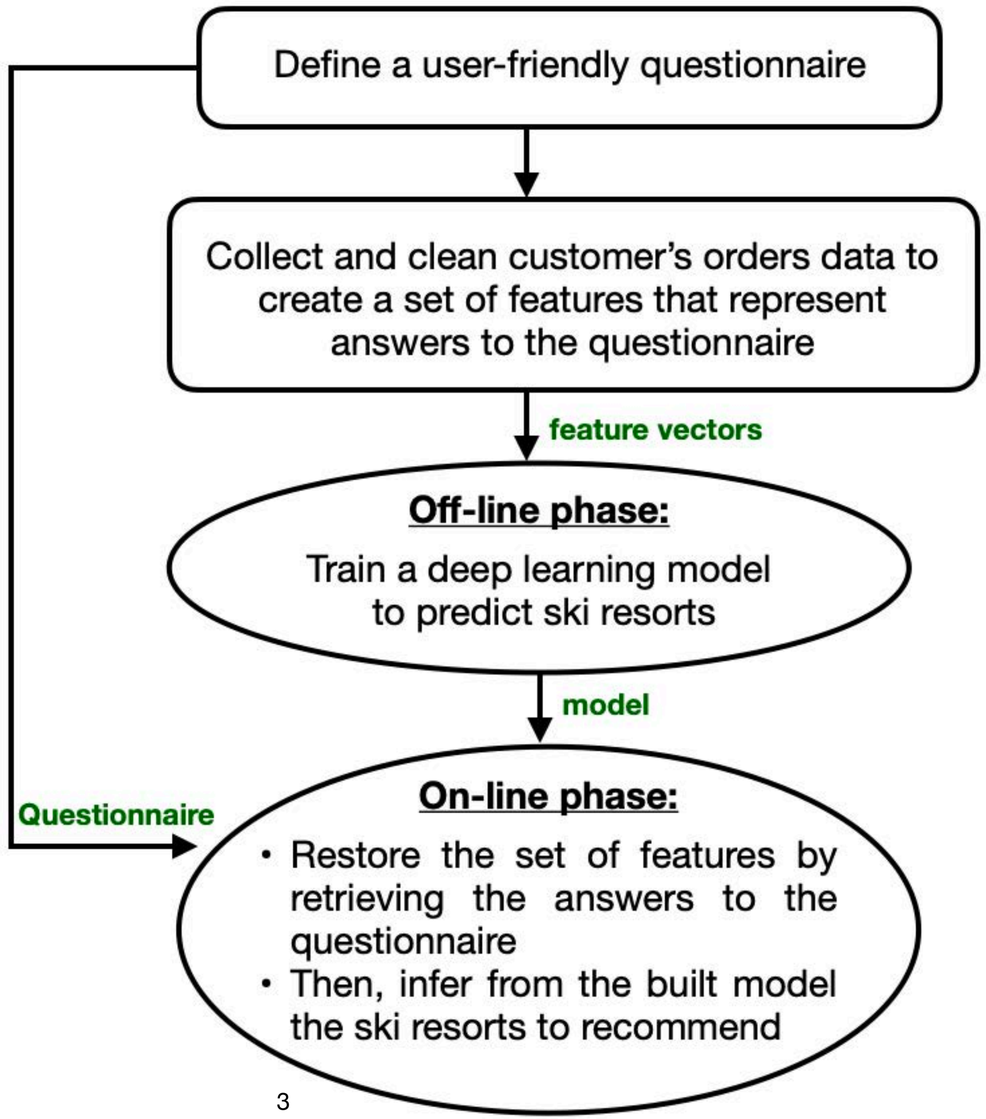
- A deep learning based Recommender System for a web business company
  - Valraiso makes commercial websites for the French Ski School

- **The aim:** Recommend suitable ski resorts for the customers visiting these websites
- **Our proposition:**
  - Collect historical data and introduce it to a deep neural network to build a model that can generate to new incoming data
  - Communicate with the customers through an online interface in order to determine their needs: the budget, the customer's location, date of stay, …
  - Using the collected data through the interface and the previously built model, a recommendation of the most suitable ski resorts is made.

Introduction

Main steps carried out

- Off-line phase
- On-line phase



Questionnaire

- Must be user-friendly
- Not many questions
- Questions need to be simple and concise
- Simple answers need to be provided

Questionnaire

01

Quel type de skieur(s) êtes-vous ?

Dans le cas d'un groupe ou d'une famille précisez le niveau moyen

- A Débutant(s)
- B Intermédiaire(s)
- C Confirmé(s)
- D Expert(s)

Questionnaire

02

Quel type de domaine skiable vous correspond le mieux ?

- **A** **Petite station** Moins de 55km de pistes
- **B** **Moyenne station** Jusqu'à 100km de pistes
- **C** **Grande station** Jusqu'à 150km de pistes
- **D** **Très grande station** Jusqu'à 250km de pistes

Questionnaire

03

Jusqu'à quelle altitude aimeriez-vous skier ?

A	**Très haute altitude**	Plus de 2700m
B	**Haute altitude**	Jusqu'à 2700m
C	**Moyenne altitude**	Jusqu'à 2300m
D	**Basse altitude**	Moins de 1800m

Questionnaire

04

Quelle semaine envisageriez-vous de partir ?

Vous avez choisi la semaine du 22/01/2022

→ SUIVANT

Questionnaire

05

Quel est votre niveau de budget pour ce séjour ?

A	Faible	€
B	Moyen	€€
C	Elevé	€€€
D	Pas de limite	€€€€

Questionnaire

06

Où résidez-vous ?

Pays

France

Département

Rhône

→ SUIVANT

Questionnaire

07

Quel est l'âge de vos enfants ?

Choisissez une ou plusieurs réponses

A Moins de 6 ans

B Entre 6 et 12 ans

C Plus de 12 ans

D Je n'ai pas d'enfants

→ AFFICHER VOS STATIONS

Questionnaire Constraints

- Reduces the feature inputs, which reduces more the information variability and complexity to be introduced into the model
- Categorizes the feature values, which 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.

Dataset Construction

- 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.

- **Feature 1: Group.**

  This feature inform 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

Dataset Construction

- **<u>Feature 2</u>: 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 is for one person (an adult), it is multiplied by the number of people included in the order

$$Budget = (5 \times daily\_pass \times (nb\_people)) + total\,Euro$$

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

  - Low cost category ∈]0,300]€ having a numeric value 0
  - Average cost category∈]300, 400]€ having a numeric value 1
  - A no limit cost category ∈]400, +∞[€ having numeric value 2

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

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.

Dataset Construction

- **<u>Feature 3</u>: 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

$$Delta\_week = stay\_week - validation\_week$$

- **<u>Feature 4</u>: Stay Week.**

In addition to the delta order, the stay week has an important influence on the choice of the resort

Dataset Construction

- **<u>Feature 5</u>: 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

Dataset Construction

- **<u>Feature 6</u>: 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.

$$slopes\_d = (0.1 * nb\_g) + (0.2 * nb\_b) + (0.6 * nb\_r) + nb\_black$$

Then, four categories are created.

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

Dataset Construction

- **Feature 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

Dataset Construction

- **Feature 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.

- **Feature 8: Department.**

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 behaviors are different, we decided to build two models. One for the French customers and the other for the remainder countries. Therefore, this feature depicts for the department of French customers.

Dataset Construction

Finally, 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]

Dataset Construction

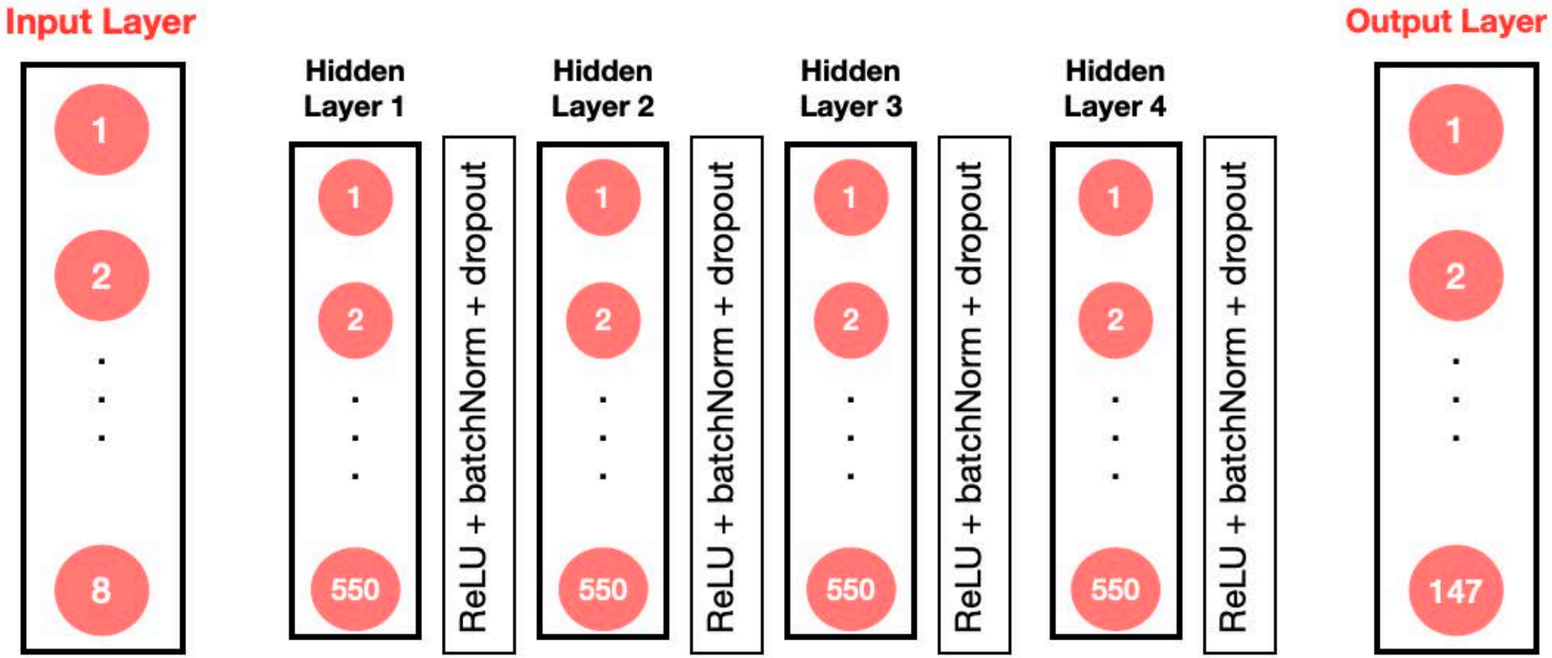
Dataset	#Classes	#Samples at training	#Samples at test	Total number of samples
SRESP_Fr	147	916451	392766	1309217
SRESP_nonFr	147	160365	68729	229094

Experimental Results

- A supervised classification that aims to predict resorts for new customers
- Parameters are selected through a 5-fold cross validation
- The evaluation is carried out by computing the accuracy metric that assesses overall, how often our model is correct

Experimental Results

Architecture of the proposed DNN



Experimental Results

DNN results

- Since we want to recommend more than one ski resort, we present the accuracy results from TOP-1 to TOP-10.
- 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

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

Conclusion

- In this work, the proposed system learns various feature vectors that describe important information specific to both customers and target resorts.
- The obtained results showed promising performance 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.