**Cetli (“Shopping List”) Competition Rules**

# 1. The Organizer

Nextent Informatika Zrt. (registered seat: 1138 Budapest, Madarász Viktor utca 47-49. data management registration number: NAIH-83388/2015, hereinafter referred to as: the “**Organizer**”) hereby launches a competition and a call for participation in relation to its Cetli (“Shopping List”) application campaign (hereinafter jointly referred to as: the “**Competition**”).

# 2. Data challenge description

## Task1: Guess The Product!

### Description

We provide the train.csv file, which contains shopping list items from January to August 2016.

We take the shopping list (“Cetlis”) of July and August, and randomly remove one item from each Cetli. The task is to predict the missing item (the target variable is the ‘Product’ column) of each individual shopping list.

### Files

train.csv - the training dataset

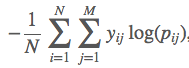
test.csv - shopping list items whose product name needs to be predicted

sample\_submission.csv - a sample file, which looks like a submission file

### Evaluation

Percentage accuracy is used as an evaluation metric. That means we will calculate a P value based on your subscription, and the P value will represent the percentage of correctly predicted product names. The calculation is the number of correctly predicted product names divided by the number of lines. The leaderboard is ranked based on P.

Submissions are evaluated using the multi-class logarithmic loss. For each Shopping List, you must submit a set of predicted probabilities (one for every Product in the dataset). The formula is:



where N is the number of ShoppingLists in the test set, M is the number of product names, log is the natural logarithm, yij is 1 if observation i is of class j and 0 otherwise, and pij is the predicted probability that observation i belongs to class j.

The submitted probabilities for a given visit are not required to sum to one because they are rescaled prior to being scored (each row is divided by the row sum). In order to avoid the extremes of the log function, predicted probabilities are replaced with max(min(p,1−10−15),10−15)

<https://www.kaggle.com/wiki/MultiClassLogLoss>

### Submission Format

You must submit a csv file with the objectIds of the Cetlis, and all candidate product probability for each class (260 columns). The order of the rows matters, it is in alphabetical order. The file must have a header and should look like the following:

**objectId,ablakmosó,alma,...**

hjdsme6Rt2,0.1,0.8,...

PzN7frABpB,0,0.3,..

## Task2: Freestyle task

In this task, you can suggest a freestyle research plan, without any limitation. You should define a goal or a question to answer, and you can submit a research plan as to how to solve it.

This can be a visualization, a route planner, a shopping behaviour analysis, and you can use additional external data, the options are up to you. We suggest that you maintain regular contact with the organizers to share and discuss your findings.

If you wish to participate then please send us a short description, explaining the way you want to use the data, and the insights that you expect from your analysis.

Plans and solutions should be sent to [hello@bevasarlocetli.hu](mailto:hello@bevasarlocetli.hu)

Evaluation:

## Dataset description

The original data is from [Cetli mobile application](http://bevasarlocetli.hu), which is a real life shopping list app. You can manage your shopping list (Cetli in Hungarian) in the application, you can share your Cetlis with your family members, you can add items to your shopping lists, and you can cross out the items once you buy them. You can try the application in real life, so you can get a good Business Understanding (as the CRISP-DM states) of the dataset.

To get an insight of the data (Data Understanding in CRIPS-DM) we shall provide a description of the columns of the train.csv for you.

We shall provide a training.csv file, which contains shopping list items for the first half of the year 2016.

We shall remove all the items suitable to identify any real person. The anonymity and protection of privacy data is our number one priority. So, our real life data are anonymized, blurred and changed, but they are still suitable for getting data insights and making predictions for the challenge.

The dataset contains GPS coordinates of the most popular shopping locations in Budapest. So, you can get an overview of the shopping habits of our capital city. These coordinates are filtered and blurred, so no real user habits can be identified by them.

### Columns

**objectId** - Unique row identifier of the item added to a Cetli

**cetliId.objectId** - Unique identifier of the Cetli

**owner.objectId** - The unique identifier of the user who adds an item to the list. Note: there are families where two unique users can add items to a Cetli.

**createdAt** - The timestamp of when the item is added to the Cetli.

**updatedAt** - The timestamp of when the item is modified.

**product** - Name of the product added to the Cetli.

**quantity** - Quantity of the product added to the Cetli.

**selected** - The user crossed out the item from the Cetli. We assume that he/she has bought the product, or has placed it into the basket.

**deleted -** The user deleted the item from the Cetli.

**crossedOutLocation.latitude** - GPS latitude coordinate of the user where the item was crossed out. It is probably a grocery store. ;)

**crossedOutLocation.longitude** - GPS longitude coordinate of the user where the item was crossed out.

## Participants

The Competition is open only to those natural persons aged 18 or over as competitors or applicants who are not subject to the restrictions set out in item 7.

The Participant agrees to use the data made available by the Organizer during the Competition and constituting the subject of the Competition, and any information obtained (by data analysis or any other method) originating from the masses of data for the sole purpose of participation in the Competition. It is strictly prohibited to use those for any other purpose, to utilize those or to forward those to third parties.

# 3. Timeline

Competition starts: 15th September 2016

Final submission deadline: 31th October 2016

Deadline is at 11:59 PM GMT+1 on the corresponding day unless otherwise noted. The competition organizers reserve the right to update the contest timeline if they deem it necessary.

# 4. Prizes

Task1:

1st place - A VIP Ticket for the Big Data Universe Conference 2017 (<http://www.bdu.hu> ) + 50 USD Amazon Gift Card + Big Data T-Shirt

2nd place - 100 USD Amazon Gift Card + Big Data T-Shirt

3rd place - 50 USD Amazon Gift Card + Big Data MUG

Task2:

1st place - A VIP Ticket for the Big Data Universe Conference 2017 (<http://www.bdu.hu> ) + 50 USD Amazon Gift Card + Big Data T-Shirt

2nd place - 100 USD Amazon Gift Card + Big Data T-Shirt

3rd place - 50 USD Amazon Gift Card + Big Data MUG

# 5. Right of use concerning the Entries

By submission of the Entry prepared by the Applicant, the Applicant, as the author of the Entry, shall grant a right of use to the Organizer without any royalties claimed, and unrestricted as to time and geographical region for the presentation of the submitted Entry in television commercials, at the www.bevasarlocetli.hu site, and the www.nextent.hu website, at the Organizer’s and the Cetli’s Facebook site, in print and online publications reports on the result of the Call for Participation, and in any form during the development of the Cetli application, and also for the Organizer to use it in the course of their own activity, for the purposes of the development of the Cetli application or any other product. The Organizer shall also be entitled to use parts of the Entries submitted by the Applicants.

A condition to participation in the Call for Participation shall be the acceptance of these Competition Rules, and of the license to use attached as Annex No. 1.

The authorization of the above content shall also apply to cases where the intellectual creation of the Applicant does not qualify as a copyright work (e.g. know-how, ideas).

In addition thereto, the Applicant declares that they are entitled to freely dispose of the Entry prepared by them and any parts thereof, those do not violate the copyrights, rights relating to personality and other rights of other persons, and that any other persons qualified as being entitled to the Entry or any parts thereof have expressly consented to the preparation, upload and use as specified above of the Entry. The Applicant acknowledges that they shall be held fully liable for any potential infringements, and the Organizer’s liability is excluded in relation thereto. Any misuses of the Entry shall automatically result in an exclusion from the Call for Participation.

If the Entry submitted by the Applicant has been jointly prepared by several persons, i.e. the Entry is a joint copyright work (or any other intellectual creation), then by its registration the Applicant shall guarantee that the Applicant is entitled to proceed on behalf of the authors, and that the Participant also represents its co-authors. The Organizer hereby excludes its liability regarding the Game, the above license to use and the prize towards the co-authors of the joint copyright works.

Furthermore, the right of use shall be governed by the provisions of Act LXXVI of 1999 on Copyrights and Act V of 2013 on the Civil Code.

# 6. Data protection

The Competitor and the Applicant participating in the Competition (hereinafter jointly referred to as: “**Participant**”) hereby declare that the data provided them are true, that by their registration they accept these Competition Rules and the Terms of Use, and they declare that they fully comply with the terms set forth therein, and their participation does not conflict with the restrictions described therein.

Data supply regarding personal data is voluntary. The Participants expressly agree to the management of their personal data by the Organizer, and they accept and acknowledge all provisions contained in these Competition Rules.

By its registration and the acceptance of these Competition Rules, the Participant provides its express consent to the Organizer to manage within the framework of the relevant legislation all personal data voluntarily provided by the Participant, in particular, but not limited to their full name (first and last name) and e-mail address, and in the case of winners their full name, address (city, postal code, street, number), their mother’s birth name, their place and date of birth.

The Participant agrees to the use of its data by the Organizer for the purpose of its recruitment activities aimed at meeting its own demand for personnel.

In the event of winning the Competition, by the acceptance of these Competition Rules, the Participant acknowledges and expressly agrees without any further conditions to the publication by the Organizer of the Participant’s voluntarily provided personal data the Participant’s name (first name, last name), its address (city and postal code only) and its prize at the www.bevasarlocetli.hu website, at the Cetli’s Facebook site, in print and online publications on the result of the award game, and on their own flyers and in their commercials, anywhere, at any time, in any form without any additional conditions or compensation, solely in relation to the Competition.

The personal data of the Participants (first name, last name, e-mail address, place and date of birth, mother’s birth name, address (city, postal code, street, number) shall be managed by the Organizer, in accordance with the effective data protection laws and legislative provisions.

The data management related to participation in this Competition shall be based upon the consent by the Participant. The purpose of data management is solely related to the marketing activities pertaining to the Competition and the promotion of the Cetli application, and the Organizer’s own recruitment activities. The Participant shall be entitled to request the deletion of its data beyond 45 calendar days following the conclusion of the Competition.

Any complaints by the Participant regarding the management of data may be referred to the National Authority for Data Protection and Freedom of Information.

Additional rules concerning the management and processing of the Participants’ personal data are contained in Act CXII of 2011 on the Right of Informational Self-Determination and on Freedom of Information.

# 7. Miscellaneous statements by the Organizer

In the event of any additional questions arising in relation to the Competition, please contact us at the [hello@bevasarlocetli.hu](mailto:hello@bevasarlocetli.hu) e-mail address.

The Organizer’s associates, and their contractual partners in an agency, contractor’s or other legal relationship with them aimed at the performance of work, and the associates of companies involved in the organization and management of the Competition, and the close relatives of such persons as specified under Section 8:1 (1)1 of Act V of 2013 (the Civil Code) are excluded from the Competition.

The Organizer may not be held liable for any non-performance, default or delayed performance, if those are attributable to a Force Majeure event. During the period of a Force Majeure event the provisions of these Rules shall be suspended to such extent as the performance of the above is rendered impossible by the Force Majeure event. Force Majeure events are unavoidable events outside the Oranizer’s control, such as: natural disasters, fire, floods, administrative provisions, states of emergency, rebellions, civil wars, wars, strikes or discontinuance of work of similar nature.

The Organizer reserves the right to amend the Competition Rules and Procedure.

Beyond 45 calendar days following the conclusion of the Competition, the Organizer shall not accept any complaints whatsoever related to the Competition.

# Annex No. 1

**Statement on Right of Use**

By this statement, I hereby provide my free, unconditional and irrevocable consent to the use by Nextent Informatika Zrt. (registered seat: 1138 Budapest, Madarász Viktor utca 47-49.) of the Entry prepared by me and sent to the Competition Organizer, as a copyright work and/or other intellectual creation (know-how, idea, etc.), without any restrictions as to space and time, in any form, without restraint, for marketing and research purposes.

By this statement I declare that I have no claims and I do not claim any royalties in relation to the above from Nextent Informatika Zrt. or any other legal or natural persons.