Petitions analysis

Int20h Hackathon

Authors:
Kaminska O., Komisarenko V., Kulik O.

February 4, 2018

Task

General

Data: dataset of petitions to Kyiv City Government Administration.

Problem: a lot of useful petitions didn't receive enough votes.

Goal: define general characteristics of successful and unsuccessful petitions.

Solution

Develop tool, which can predict success of petition and give some recommendations, how to improve it!

Figure 1: Happy user

Data processing

Steps

1. Remove wrong symbols;
2. Split on successful/unsuccessful petitions;
3. Calculate numeric features;
4. Data processing: delete stop-words / punctuation / URLs etc, extract stemmas, find the most popular;
5. Detect common popular words in both types, delete them.

Format of data

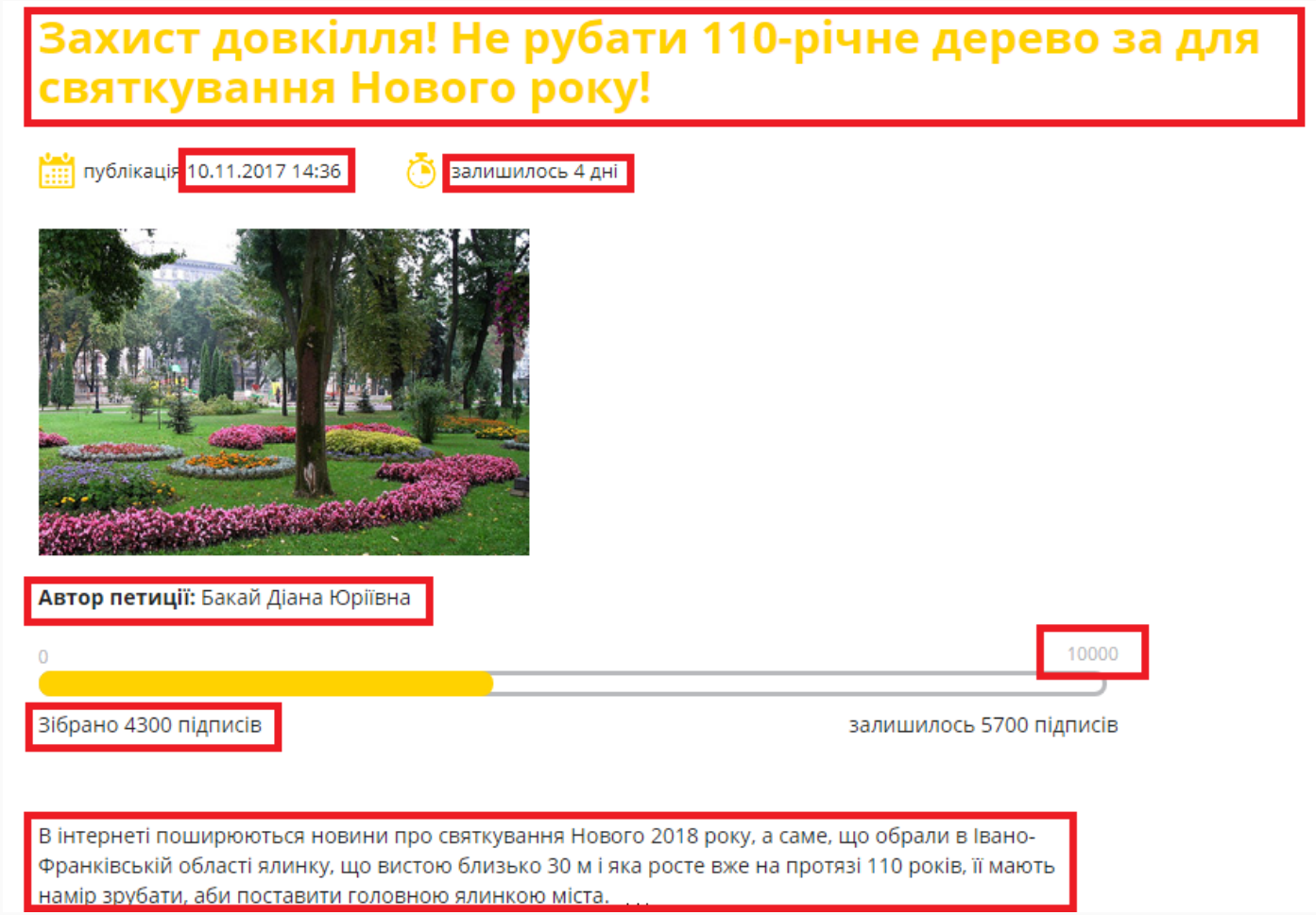


Figure 2: Example of data

Some statistics

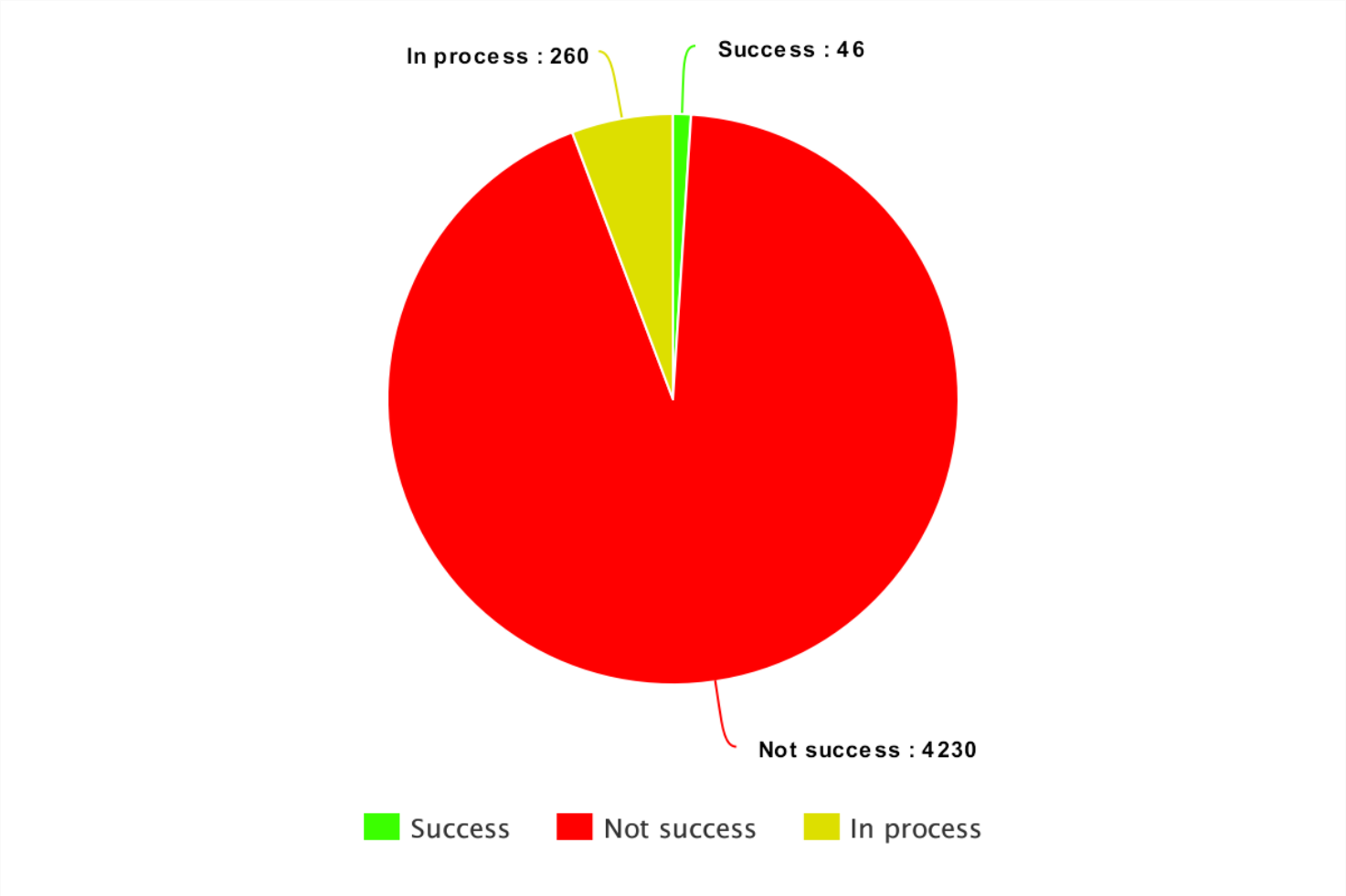


Figure 3: Value of successful/unsuccessful petitions

Some statistics

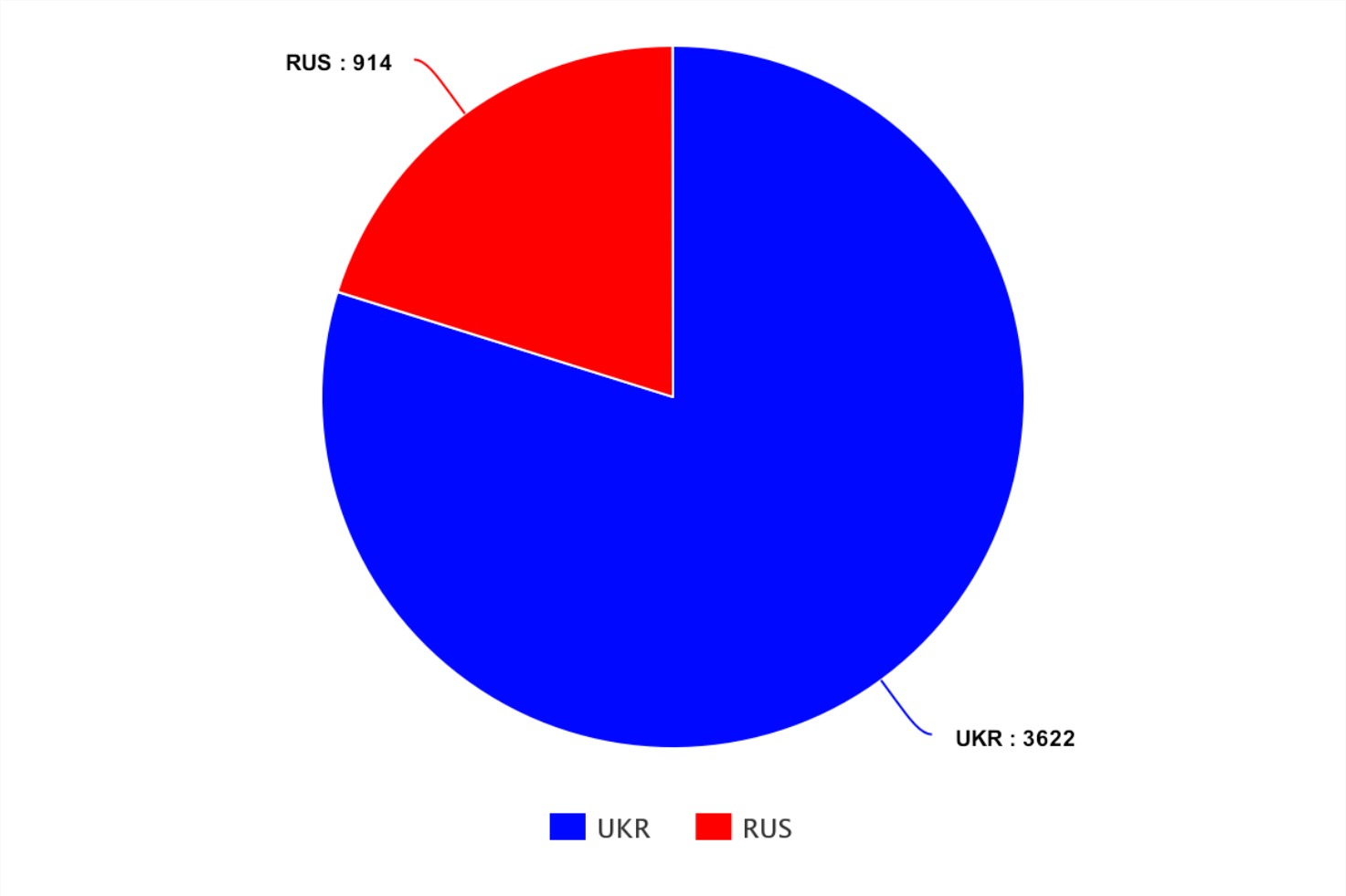


Figure 4: Value of RUS/UKR language petitions

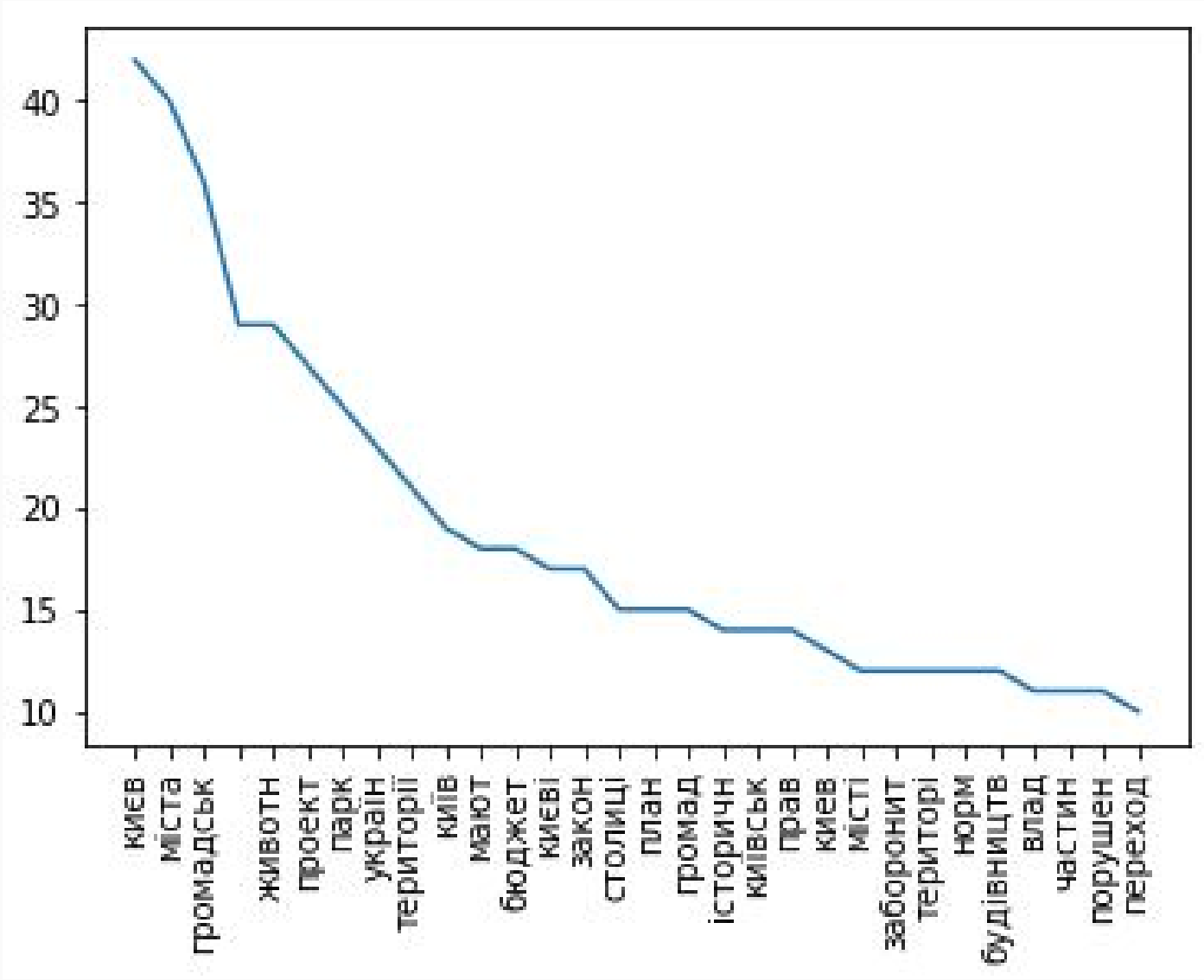
Features

- ✔ Number of words written with CAPS LOCK
- ✔ Average length of words in symbols
- ✔ Average length of sentences in words
- ✘ Language: RU / UKR
- ✘ Number of exclamation marks!!!
- ✘ Sentiment intensity of petition

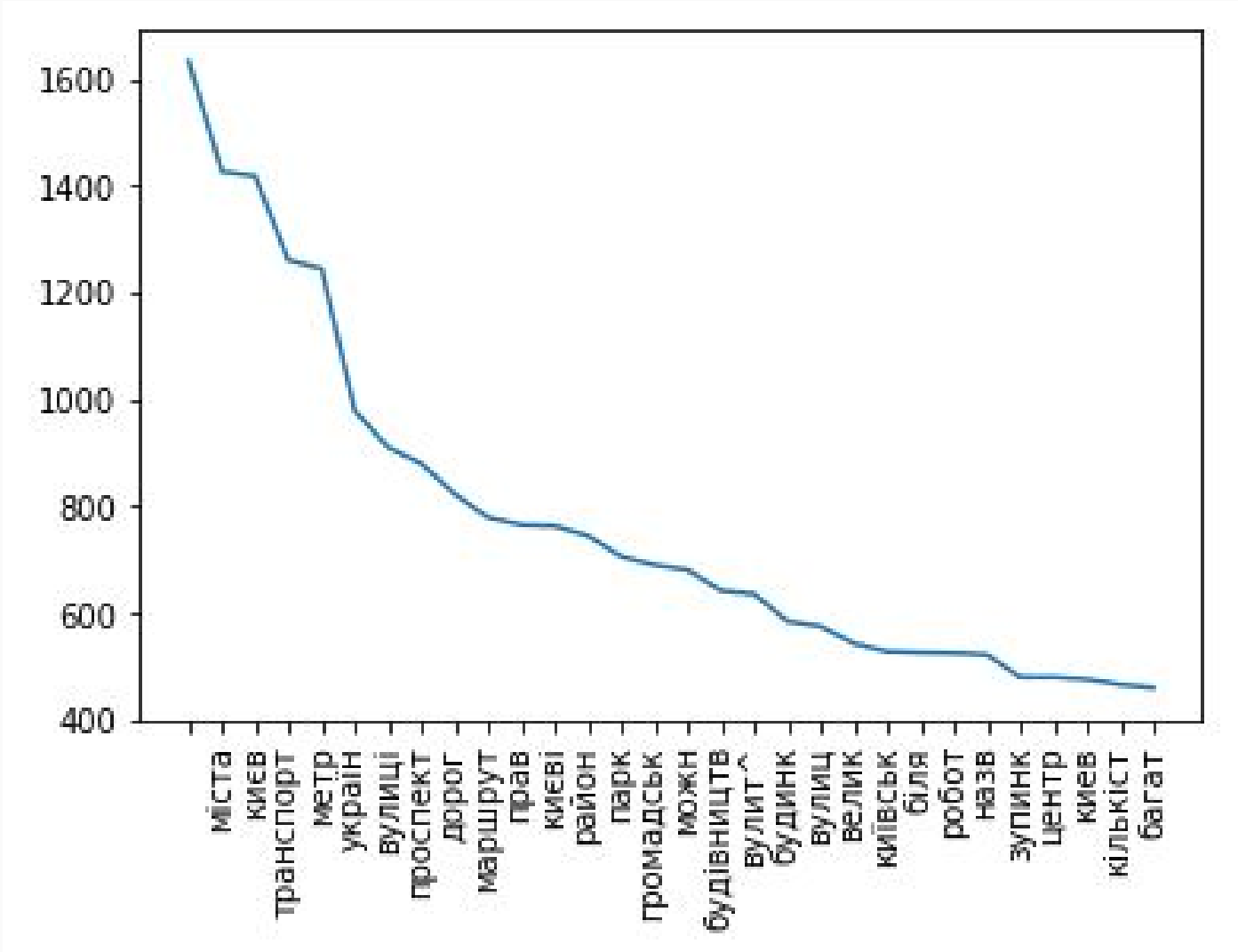
All hypothesizes were checked with t-test

Text analysis

Popular words in successful petitions



Popular words in unsuccessful petitions



UNIQUE popular words - successful

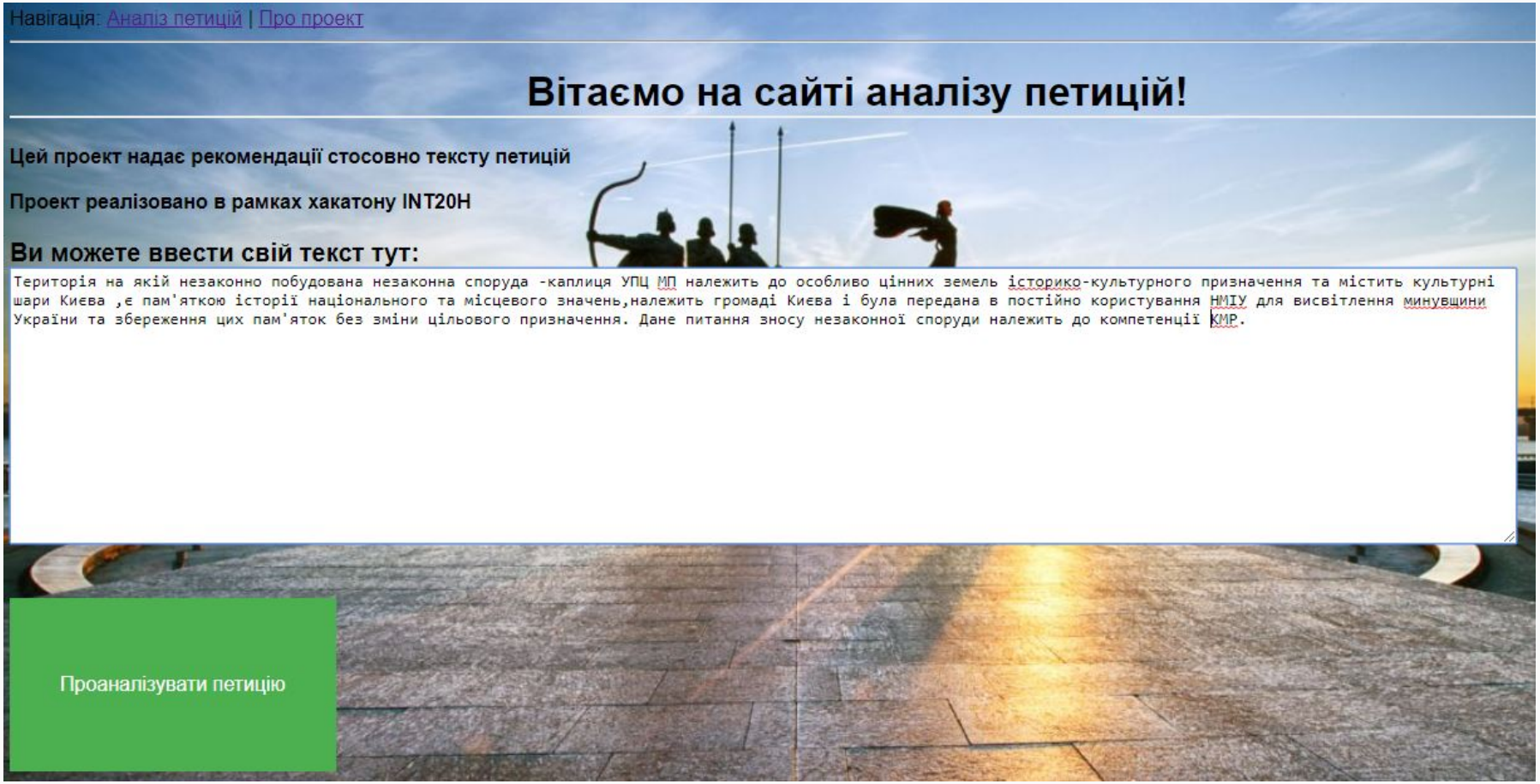
UNIQUE popular words - unsuccessful

Prediction number of votes

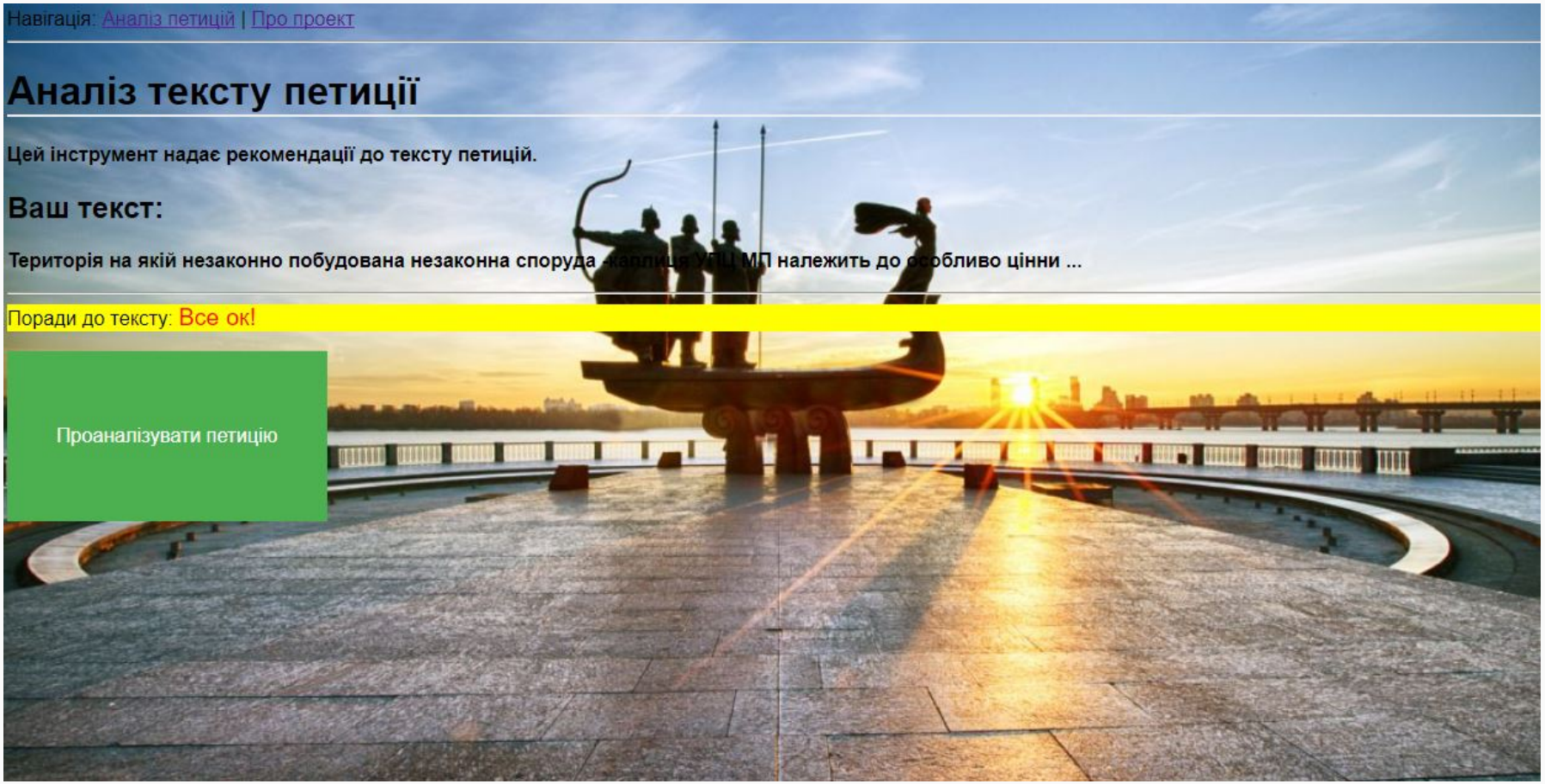
We used **Word2Vec** for Ukrainian language (on the base of Wikipedia) - to convert text of each petition into 300-dimensional vector.
We train **Decision Tree Regression Model** on all petitions.
Also, we use those features to detect the closest successful and unsuccessful petitions.

App description

Start page


Навігація: Аналіз петицій | Про проект
Вітаємо на сайті аналізу петицій!
Цей проект надає рекомендації стосовно тексту петицій
Проект реалізовано в рамках хакатону INT20H
Ви можете ввести свій текст тут:
Територія на якій незаконно побудована незаконна споруда -каплиця УПЦ МП належить до особливо цінних земель історико-культурного призначення та містить культурні шари Києва ,є пам'яткою історії національного та місцевого значень,належить громаді Києва і була передана в постійно користування НМІУ для висвітлення минувщини України та збереження цих пам'яток без зміни цільового призначення. Дане питання зносу незаконної споруди належить до компетенції КМР.
Проаналізувати петицію

Result


Навігація: Аналіз петицій | Про проект
Аналіз тексту петиції
Цей інструмент надає рекомендації до тексту петицій.
Ваш текст:
Територія на якій незаконно побудована незаконна споруда -капліця УПЦ МП належить до особливо цінни ...
Поради до тексту: Все ок!
Проаналізувати петицію

What we also tried

Almost...

- Clusterization by topics - deficit of topics in original data;
- Neuron Network to generate petition text - not enough power of PC;
- Some features on web-app - should make on simple GUI program.

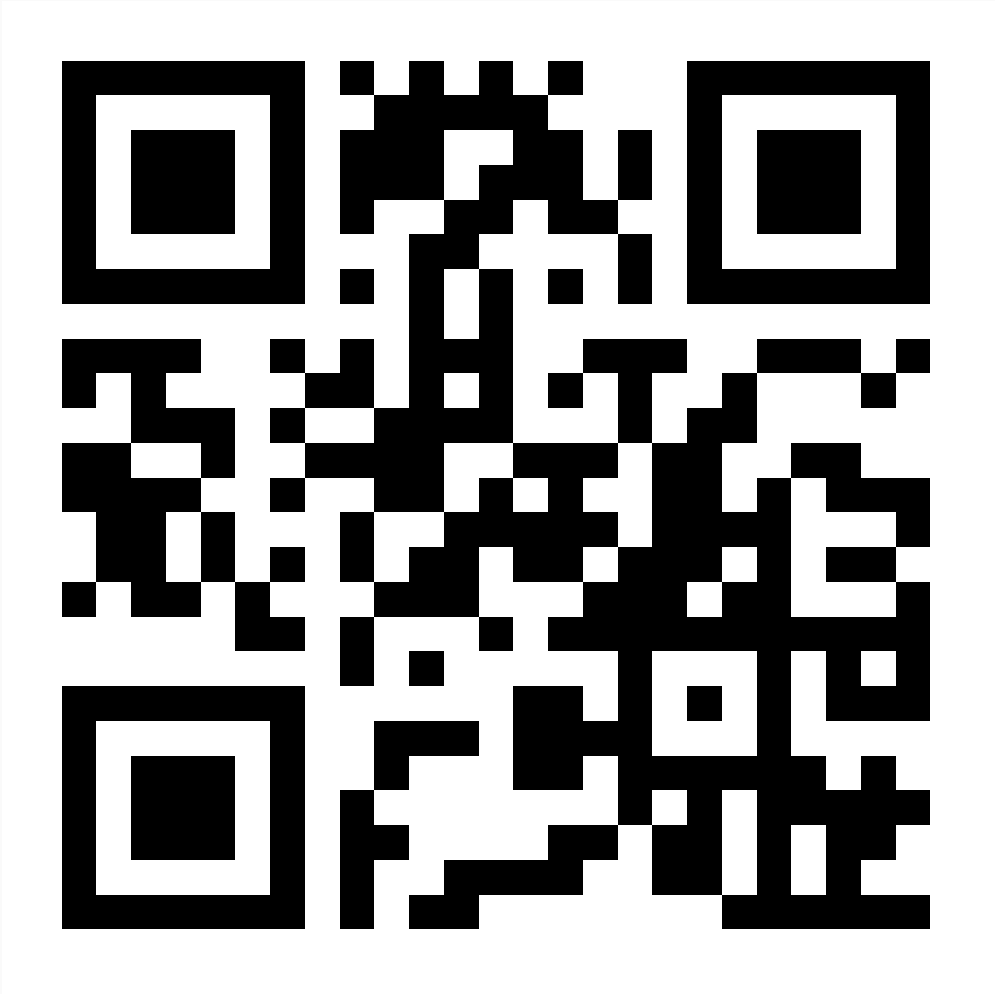
Summary

Total

During Hackaton we built a working prototype, which:

- Analyze our petition text;
- Can give advice how to improve it;
- Predict number of votes, detect emotion of text and closest successful and unsuccessful petitions.

DEMO



https://goo.gl/h87c5L

Thank for your attention!

Questions?