

Assignment 5

This assignment will be graded by 100 point scale and could compose 10% of your total grade. You must send the presentation to na.melkonyan@gmail.com.

Deadline: 04 June 2019, 11:59 pm (GMT+4)

Programming

You are allowed to use any packages and tools that suit to you.

Data

The dataset has been obtained from a Kaggle Competition “What’s cooking?” hosted by Yummly. It is a popular website and application which provides recipe recommendations tailored to the individual's preferences, semantic recipe search and a digital recipe box. It contains 39774 instances of cuisine and ingredient list pairings, in which each cuisine has several ingredients.

Task

Your task will be to use your learnt techniques in combination with your research skills to classify cuisine based on ingredients. Use methods you know so far to extract knowledge from Yummly data (KDD process). Thus I expect creative solutions.

Presentation

Each presentation should be no more than 10 minutes in length, so consider not to exceed the given time frame.

Before forming your solutions, you should first tell about the problem, then the data, your approach and at the end show the results.

Extra Points

Extra points will be granted for extra creativity, e.g. Sturges' formula has been used to find the number of bins in the histogram.

Rules

Take attention on the feedbacks of previous assignments.

References

There are a lot of research and papers on this data. You are recommended to go over some of them.