# Machine Learning 3 Module 3, 2019

# Daria Dzyabura

New Economic School

ddzyabura@nes.ru

## TAs: Konstantin Korotkiy, kkorotkiy@nes.ru

#### **Course description**

This course will be an applied course on business data analytics. We will focus primarily on consumer data (loyalty card, purchase, promotion, product development data, etc). The goal is for you to get experience working with real world, "messy" data sets, and use them to answer relevant business questions. [This should be a short description of the course from one to three paragraphs.

It should include course goals and learning objectives.

It should also describe what skills and knowledge students obtain during the course.]

#### Course requirements, grading, and attendance policies

You are expected to have working knowledge of R or Python and have taken Econometrics 1 and 2, and Machine Learning 1. We will meet 7 times. Your grade will consist of the following:

Mini Assignments (6)	20%
Homeworks (4)	50%
Class attendance and participation	10%
Final project	20%

<u>Mini-Assignments.</u> For every class, we will use a new data set. Before class, you are required to complete a mini assignment, which involves: (1) loading the data set, (2) running some summary statistics (can be a table or a plot), and (3) prepare one or two questions you would like to answer from the data. You should be ready to explain in class what you have done.

<u>Homeworks.</u> In class, we will discuss an analysis plan for the data set. In the seminars (Saturdays), you will carry out the analysis together with the Teaching Assistant. Whatever you don't have time to complete in the seminar, will be your homework. These assignments should all be submitted through my.nes by the assigned deadline (typically Thursday evening). They will involve a combination of analysis, as well as interpretation of results. You will need to submit your code, output, and interpretations, which may be in either English or Russian. While everyone should submit all 6 assignments, we will only count 4 of your highest grades.

<u>Class Attendance and Participation.</u> Classes will consist of lectures as well as in class discussions, focused on data analysis plan (Which models we should estimate? How should we do sampling? What is the correct way to do validation? Which metrics are best? etc.). In seminars, you will work on carrying out the analysis plan, which will combine reproducing results shown in lecture, as well as generating new results. It is important that you attend all lectures and seminars. I also hope that you will participate actively in class discussions. Class attendance and participation constitutes 10% of your final grade. If you miss a class or seminar, it is your responsibility to make up the missed content.

<u>Final Project.</u> In the last week of the class, each of you will be assigned one of the 6 data sets we have analyzed in class for deeper study in the final project. The project should extend the analysis. You may choose the direction of the extension: either estimate a more complex model on the data, test a new hypothesis with the data, or derive a prescriptive recommendation for the firm based on the data. For example, you may propose an optimal design of a product, or a line of products, propose the next experiment the firm should run, design a new promotion, identify customer segments,... the direction is really up to you. You may even carry out a similar analysis on a related data set and compare results. The projects will be graded based on the ambitiousness of the analysis carried out and rigor of the recommendation made.

### **Course contents**

- Product design & Customer preferences (Conjoint analysis, multiattribute utility models, heterogeneous preferences) Tools/approaches - regression, clustering, random coefficients model
- Customer preferences II (Prediction, recommendation systems, high-dimensional data, unstructured data)
  - Tools/approaches feature selection, k-Nearest Neighbors, collaborative filtering
- 3. Demand elasticity (purchase data, pricing promotions, profitability analysis)
- Customer Relationship Management (loyalty card data, customer profitability, customer retention, promotion evaluation) Tools/approaches - field experiments, design of experiments, predictive analytics
- Brand Management (customer brand perceptions, brand monitoring, survey methods, social media mining) Tools/approaches (Principal Components Analysis, Topic Modeling)
- Brand Management II (social media mining continued) Tools/approaches (Image analytics, word embeddings)

## Description of course methodology

The class will use a combination of resources, surrounding the data-based case studies. We will use the content of my lecture slides, various online software modules, and academic papers to guide our analyses.

## **Course materials**

#### **Required textbooks and materials**

We will not use a textbook, and rely on separate sources. Below is a subset of the readings. TBA

#### **Additional materials**

TBS

## Academic integrity policy

Cheating, plagiarism, and any other violations of academic ethics at NES are not tolerated.