



**YESSENOV
DATA LAB**

12 June – 08 July 2023
Almaty



▶ Table of Contents

- What Is Yessenov Data Lab Summer School 2023?
- Training Program
 - ▶ Week 1. Python Programming Language
 - ▶ Week 2. Data Analysis. Regression Data Analysis
 - ▶ Week 3. Data Classification
 - ▶ Week 4. Clustering. Deep Learning



▶ What is Yessenov Data Lab 2023?

Yessenov Data Lab summer school 2023 is a 4-week intensive course to train Kazakhstani young scientists and specialists in the basic skills of big data analysis for further application of knowledge in real practical work.

Dates of the school: 12 June – 08 July 2023

GRADUATES WILL GAIN THE FOLLOWING SKILLS:

- 1. Python Programming as Part of Data Analysis
- 2. Statistical Analysis Methods
- 3. Data Pre-processing and Preparation for Subsequent Analysis
- 4. Data Visualization and Pattern Recognition
- 5. Forecasting Based on Historical Data
- 6. Understanding Various Machine Learning Algorithms
- 7. The Right Choice of Training Model



▶ Training Program

Timur Bakibayev

Ph.D., Professor of ALMAU, Co-founder,
DSA Engineering



▶ Week 1. Python Programming Language

Objective: Python for Data Best Practices

Day 1 Object-Oriented Programming

- 10:00 – 10:10 Opening of Summer School, Welcome Words
- 10:10 – 11:30 What is Data Science, Machine Learning
- 11:45 – 13:15 Python: OOP, Type Annotations and Tests
- 14:30 – 16:00 Laboratory Work on OOP
- 16:15 – 18:00 Laboratory Work, Mistake Discussion

Day 2 Graph Theory

- 10:00 – 11:30 Graph Representation in Python
- 11:45 – 13:15 Breadth-First Search - Dijkstra's Algorithm
- 14:30 – 16:00 Laboratory Work: Breadth-First Search
- 16:15 – 18:00 Laboratory Work Analysis

Day 3 Data is Everything

- 10:00 – 11:30 Numpy, Pandas library review
- 11:45 – 13:15 Data Grouping, Filters, Sorting.
- 14:30 – 16:00 Laboratory Work: Work With Pandas
- 16:15 – 18:00 Laboratory Work Analysis

Day 4 Data is Beautiful

- 10:00 – 11:30 Matplotlib Library Review
- 11:45 – 13:15 SeaBorn Library Review
- 14:30 – 16:00 Laboratory Work: Visualisation
- 16:00 – 17:00 Laboratory Work Analysis
- 17:00 – 18:00 Motivational Meeting

Day 5 StreamLit: Web for Data

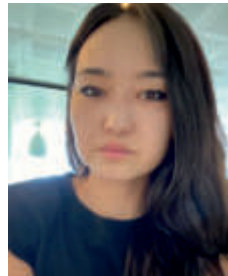
- 10:00 – 11:30 StreamLit Review
- 11:45 – 13:15 Laboratory Work: StreamLit
- 14:30 – 16:00 Laboratory Work: StreamLit
- 16:00 – 18:00 StreamLit Deployment



▶ Training Program

Aigerim Sadykova

Bs in Computer Science and
Software Engineering,
Ex-Senior Data Analyst in PwC



▶ Week 2. Data Analysis. Regression Modelling

Objective: Mastering the Theory and Skills of Statistical Analysis;
Building Predictive Regression Models

Day 1 Exploratory Dana Analysis

- 10:00 – 11:30 Intelligence Analysis, Histograms, Outliers
- 11:45 – 13:15 Distributions, Correlation Analysis
- 14:30 – 16:00 Laboratory Work
- 16:15 – 18:00 Laboratory Work, Discussion

Day 2 Hypothesis Testing

- 10:00 – 11:30 Hypothesis Testing. One-sample Test
- 11:45 – 13:15 Hypothesis Testing. Two-sample Test
- 14:30 – 16:00 Laboratory Work
- 16:15 – 18:00 Laboratory Work, Discussion

Day 3 Regression Modelling p.1

- 10:00 – 11:30 Linear Regression Types
- 11:45 – 13:15 Metrics for Regression Problems
- 14:30 – 16:00 Laboratory Work
- 16:15 – 18:00 Laboratory Work, Discussion

Day 4 Regression Modelling p.2

- 10:00 – 11:30 Feature Selection. Regularization
- 11:45 – 13:15 Regression Model Optimization
- 14:30 – 16:00 Laboratory Work
- 16:00 – 17:00 Laboratory Work, Discussion
- 17:00 – 18:00 Motivational Meeting

Day 5 Project on Regression Modelling

- 10:00 – 11:30 Project Planning
- 11:45 – 13:15 Project Implementation
- 14:30 – 16:00 Laboratory Work
- 16:15 – 18:00 Laboratory Work, Discussion



▶ Training Program

Kuanysh Abeshev

Ph.D, Dean of School of
Digital Technologies, AlmaU



▶ Week 3. Data Classification

Objective: Building Predictive Data Classification Models

Day 1 Evaluation Performance - Classification

- 10:00 – 11:30 Metrics for Classification Problems
- 11:45 – 13:15 Logistic Regression (Binary and Multi-class)
- 14:30 – 16:00 Laboratory Work Analysis
- 16:15 – 18:00 Laboratory Work Analysis. Distribution of Datasets for the Project Defense

Day 2 Instance-based Learning. Probabilistic Classification

- 10:00 – 11:30 K-nearest Neighbors (knn) Method
- 11:45 – 13:15 Naive Bayes Classifier
- 14:30 – 16:00 Laboratory Work
- 16:15 – 18:00 Laboratory Work Analysis

Day 3 Kernel Based Models

- 10:00 – 11:30 Support Vector Machines (svm) Method - Classification
- 11:45 – 13:15 Support Vector Machines (svm) Method - Regression
- 14:30 – 16:00 Laboratory Work
- 16:15 – 18:00 Laboratory Work Analysis

Day 4 Tree Based Models

- 10:00 – 11:30 Decision Tree
- 11:45 – 13:15 Ensemble Methods and Random Forests
- 14:30 – 16:00 Laboratory Work
- 16:00 – 17:00 Laboratory Work Analysis
- 17:00 – 18:00 Motivational Meeting

Day 5 Classification Project

- 10:00 – 11:30 Project Planning
- 11:45 – 13:15 Project Implementation
- 14:30 – 16:00 Laboratory Work
- 16:15 – 18:00 Laboratory Work Analysis



▶ Training Program

Aidos Sarsembayev

Ph.D., Assistant Professor IITU,
Senior Machine Learning Engineer
re:Point Company



▶ Week 4. Clustering and Deep Learning

Objective: Mastering the Theory and Performing Applied Tasks in Unsupervised Learning; Mastering Deep Learning Methods and Neural Networks

Day 1 Unsupervised Learning

10:00 – 11:30 Unsupervised Learning; Clustering, K-means, Pca Algorithm
11:45 – 13:15 DbSCAN, Hierarchical Clustering. Quality Metrics
14:30 – 16:00 Laboratory Work
16:15 – 18:00 Laboratory Work, Discussion

Day 2 Deep learning (MLP, ANN)

10:00 – 11:30 Introduction to Neural Networks. Multilayer Perspectron
11:45 – 13:15 Backpropagation Algorithm
14:30 – 16:00 Laboratory Work
16:15 – 18:00 Laboratory Work, Discussion

Day 3 Deep Learning With PyTorch

10:00 – 11:30 Introduction to Pytorch Library
11:45 – 13:15 Convolutional Neutral Networks (cnn). Image Classification. Metrics
14:30 – 16:00 Laboratory Work
16:15 – 18:00 Laboratory Work, Discussion

Day 4 Deep Learning on Image Data. Segmentation

10:00 – 11:30 Image Segmentation Models. Data Preparation
11:45 – 13:15 Image Segmentation Models. Model Building, Training, Quality Metrics
14:30 – 16:00 Laboratory Work
16:00 – 18:00 Laboratory Work, Discussion

Day 5 Deep Learning on Image Data. Object Detecion

10:00 – 11:30 Object Detection Models on Images. Data Preparation
11:45 – 13:15 Object Detection Models on Images. Model Building, Training, Quality Metrics
14:30 – 16:00 Laboratory Work
16:00 – 17:00 Laboratory Work, Discussion
17:00 – 18:00 Closing of Yessenov Data Lab



STAY: yessenovfoundation.org
IN: info@yessenovfoundation.org
TOUCH: +7 771 759 5944