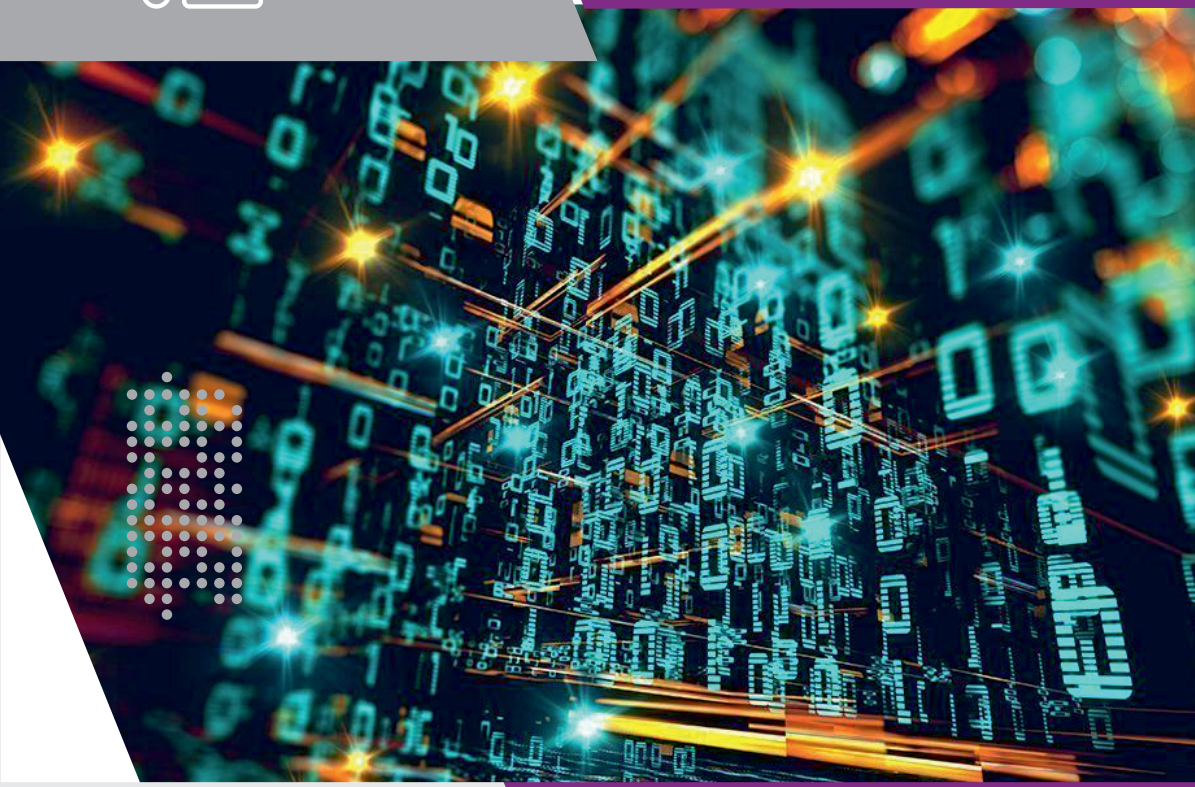
SHAKHMARDAN YESSENOV FOUNDATION



YESSENOV DATA LAB

June 17 – July 13

Almaty 2024

YESSENOV DATA LAB

# Contents


- What is Yessenov Data Lab 2024 Summer School?
- Training Program:
  - Week 1 Python for Data
  - Week 2. Regression and Analysis
  - Week 3. Data Classification
  - Week 4. Deep Learning and Computer Vision

SHAKHMARDAN YESSENOV FOUNDATION

YESSENOV DATA LAB

# What is Yessenov Data Lab 2024 Summer School?

Yessenov Data Lab Summer School 2024 is a 4-week intensive course.
3 weeks of online training,
1 week of offline training (in Almaty).

**Dates:** June 17 – July 13, 2024

## A 4-WEEK SCHOOL GIVES YOU THE OPPORTUNITY TO:

- learn big data analysis skills that you can apply in practice;
- communicate with lecturers, ask your questions;
- meet like-minded people and build new connections.

In addition to the YDL School certificate, excellent students will also receive an academic transcript/certificate from AlmaU, equivalent to a certificate of advanced training, which counts towards the students' training.

- The Foundation covers return tickets, food and accommodation expenses for the winners of the Yessenov Data Lab contest who come to Almaty from other regions from July 8 to July 13.
The School participants are provided with food during the offline training.

SHAKHMARDAN YESSENOV FOUNDATION

Online

YESSENOV DATA LAB

# Training Program

Lecturer:

## Timur Bakibayev

Ph.D. in Computer Science, Professor at AlmaU

## Week 1. Python for Data

**Objective:** to learn the basics of the Python programming language for data analysis

**Day 1 Artificial Intelligence and Work with ChatGPT**

| | |
|---|---|
| 10:00 – 10:10 | Opening of the Summer School. Welcome Message |
| 10:10 – 11:30 | What is Artificial Intelligence |
| 11:45 – 13:15 | ChatGPT Principles |
| 14:30 – 16:00 | Laboratory Work with ChatGPT API |
| 16:15 – 18:00 | Laboratory Works' Analysis |

**Day 2 Working with Data**

| | |
|---|---|
| 10:00 – 11:30 | NumPy and Pandas Libraries' Review |
| 11:45 – 13:15 | Data Grouping. Filters, Sorting |
| 14:30 – 16:00 | Laboratory Work with Pandas |
| 16:15 – 18:00 | Laboratory Works' Analysis |

**Day 3 Data Visualization**

| | |
|---|---|
| 10:00 – 11:30 | Matplotlib Review |
| 11:45 – 13:15 | Seaborn Library Review |
| 14:30 – 16:00 | Laboratory Work: Visualization |
| 16:15 – 18:00 | Laboratory Works' Analysis |

**Day 4 Basic Forecasting**

| | |
|---|---|
| 10:00 – 11:30 | Regression Problem Review |
| 11:45 – 13:15 | Feature Importance in Model Building |
| 14:30 – 16:00 | Laboratory Work: Analysis and Data Preprocessing |
| 16:00 – 17:00 | Laboratory Works' Analysis |
| 17:00 – 18:00 | Motivational Meeting |

**Day 5 StreamLit: The Web of 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 |

SHAKHMARDAN YESSENOV FOUNDATION

Online

YESSENOV DATA LAB

# Training Program

Lecturer:

## Kirill Yakunin

Ph.D. in Computing Systems and Software,
ML Engineer at Metaculus

## Week 2. Regression and Analysis

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

**Day 1 Statistical Analysis**

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

**Day 2 Regression Modelling P.1**

10:00 – 11:30 Linear Regression Basics. Metrics. Data Verification.
11:45 – 13:15 Data Preprocessing
14:30 – 16:00 Laboratory Work
16:15 – 18:00 Laboratory Work. Discussion

**Day 3 Regression Modelling P.2**

10:00 – 11:30 Polynomial Regression. Regularization of Regression Models
11:45 – 13:15 Gradient Descent
14:30 – 16:00 Laboratory Work
16:15 – 18:00 Laboratory Work. Discussion

**Day 4 Unsupervised Learning**

10:00 – 11:30 Feature Importance
11:45 – 13:15 Unsupervised Learning. Dimensionality Reduction
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 Complete Machine Learning Project
11:45 – 13:15 Q&A
14:30 – 16:00 Project Defence
16:15 – 18:00 Project Defence and Discussion of the Results

SHAKHMARDAN YESSENOV FOUNDATION

Online

YESSENOV DATA LAB

# Training Program

Lecturer:

## Kuanysh Abeshev

Ph.D. in Mathematics,
Dean of the School of Digital Technologies,
AlmaU

## Week 3. Data Classification

**Objective:** building predictive data classification models

**Day 1 Performance Evaluation - 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
16:15 – 18:00 Laboratory Works' Analysis. Distribution of Datasets for the Project Defence

**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 Works' Analysis

**Day 3 Kernel-Based Models**

10:00 – 11:30 Support Vector Machines (SVM) – Classification
11:45 – 13:15 Support Vector Machines (SVM) - Regression
14:30 – 16:00 Laboratory Work
16:15 – 18:00 Laboratory Works' 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 Works' Analysis
17:00 – 18:00 Motivational Meeting

**Day 5 Classification Project**

10:00 – 11:30 Principal Component Analysis (PCA)
11:45 – 13:15 Laboratory Work
14:30 – 16:00 Project Implementation
16:15 – 18:00 Presentation and Defence of the Project

SHAKHMARDAN YESSENOV FOUNDATION

# Offline
## Almaty

YESSENOV DATA LAB

## Training Program

Lecturer:

**Vlad Yushchenko**

M.Sc. Computer Science,
Sr. Machine Learning Engineer
at Cape Analytics

## Week 4. Deep Learning and Computer Vision

**Objective:** mastering deep learning methods and neural networks

**Day 1 Deep Learning (MLP, ANN)**

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

**Day 2 Deep Learning with PyTorch**

10:00 – 11:30 Introduction to PyTorch Library
11:45 – 13:15 Convolutional Neural Networks (CNN). Image Classification. Metrics
14:30 – 16:00 Laboratory Work
16:15 – 18:00 Laboratory Work. Discussion

**Day 3 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
16:15 – 18:00 Laboratory Work. Discussion

**Day 4 Deep Learning on Image Data. Object Detection**

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

**Day 5 Introduction to MLOps**

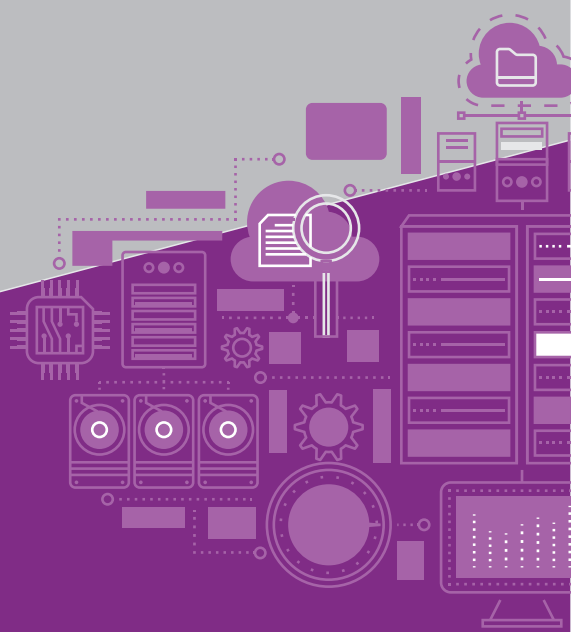
10:00 – 11:30 Introduction to MLOps, Levels of Automation
11:45 – 13:15 Experiment Tracking and Model Deployment. MLOps Best Practices
14:30 – 16:00 Laboratory Work
16:00 – 17:00 Laboratory Work. Discussion
17:00 – 18:00 Motivational Meeting

**Day 6**

10:00 – 16:00 Closing of the Summer School. Team Building. Networking

SHAKHMARDAN YESSENOV FOUNDATION

YESSENOV
DATA LAB



SHAKHMARDAN YESSENOV
FOUNDATION

STAY IN TOUCH WITH US

yessenovfoundation.org
info@yessenovfoundation.org
+7 771 759 5944