



**YESSENOV  
DATA LAB**

12 июня – 08 июля 2023  
Алматы



## ▶ Оглавление

- **Что такое летняя школа Yessenov Data Lab 2023?**
- **Программа обучения**
  - ▶ **Неделя 1. Язык программирования Python**
  - ▶ **Неделя 2. Анализ данных. Регрессионный анализ данных**
  - ▶ **Неделя 3. Классификация данных**
  - ▶ **Неделя 4. Кластеризация. Глубокое обучение**



## ▶ Что такое Yessenov Data Lab 2023?

Летняя школа Yessenov Data Lab 2023 это 4-недельный интенсив, по обучению казахстанских молодых ученых и специалистов базовым навыкам анализа больших данных для дальнейшего применения знаний в реальных практических работах.

**Даты проведения:** 12 июня – 08 июля 2023

### **ВЫПУСКНИКИ ШКОЛЫ ПРИОБРЕТУТ НАВЫКИ:**

- 1. Программирования на языке Python в рамках анализа данных
- 2. Методов статистического анализа
- 3. Предварительной обработки данных, подготовки их для последующего анализа (preprocessing) и нахождения в них зависимостей. Составления прогнозов
- 4. Визуализирования данных и нахождения в них зависимости
- 5. Составления прогноза на основании исторических данных
- 6. Понимания различных алгоритмов машинного обучения
- 7. Правильного выбора модели обучения



## ▶ Программа обучения

### Тимур Бакибаев

Ph.D, профессор AlmaU,  
DSA Engineering



## ▶ Неделя 1. Python Programming Language/ Язык программирования Python

**Цель:** освоить лучшие практики языка программирования Python для работы с данными

### День 1 Object-oriented programming/объектно-ориентированное программирование

- 10:00 – 10:10 Data Science, Machine Learning
- 10:10 – 11:30 Что такое Data Science, Machine Learning
- 11:45 – 13:15 Python: ООП, Аннотация типов и тесты
- 14:30 – 16:00 Лабораторная работа по ООП
- 16:15 – 18:00 Разбор лабораторных работ

### День 2 Graph Theory/Теория графов

- 10:00 – 11:30 Представление графов на Python
- 11:45 – 13:15 Поиск в ширину - алгоритм Дейкстры
- 14:30 – 16:00 Лабораторная работа: поиск в ширину
- 16:15 – 18:00 Разбор лабораторных работ

### День 3 Data is Everything/Работа с данными

- 10:00 – 11:30 Знакомство с библиотеками Numpy, Pandas
- 11:45 – 13:15 Группировка данных. Фильтры, сортировка
- 14:30 – 16:00 Лабораторная работа: работа с Pandas
- 16:15 – 18:00 Разбор лабораторных работ

### День 4 Data is Beautiful/Визуализация данных

- 10:00 – 11:30 Знакомство с библиотекой Matplotlib
- 11:45 – 13:15 Знакомство с Seaborn
- 14:30 – 16:00 Лабораторная работа: Визуализация
- 16:00 – 17:00 Разбор лабораторных работ
- 17:00 – 18:00 Мотивационная встреча

### День 5 StreamLit: Web for Data/Веб для данных

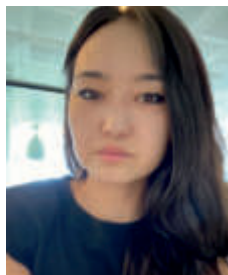
- 10:00 – 11:30 Знакомство с StreamLit
- 11:45 – 13:15 Лабораторная работа: StreamLit
- 14:30 – 16:00 Лабораторная работа: StreamLit
- 16:00 – 18:00 Разворачиваем StreamLit на сервере (Deployment)



## ▶ Программа обучения

### Айгерим Садыкова

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



## ▶ Неделя 2. Data analysis. Regression modelling / Анализ данных. Регрессионное прогнозирование

**Цель:** освоение теории и навыков статистического анализа; построение предиктивных моделей регрессии

### День 1 Exploratory Data Analysis / Разведывательный анализ данных

- 10:00 – 11:30 Разведывательный анализ, гистограммы, выбросы
- 11:45 – 13:15 Распределения, корреляционный анализ
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, обсуждение

### День 2 Hypothesis testing / Тестирование гипотез

- 10:00 – 11:30 Тестирование гипотез. One-sample test
- 11:45 – 13:15 Тестирование гипотез. Two-sample test
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, обсуждение

### День 3 Regression modelling p.1 / Регрессионное прогнозирование ч.1

- 10:00 – 11:30 Типа линейных регрессий
- 11:45 – 13:15 Метрики для регрессионных задач
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, обсуждение

### День 4 Regression modelling p.2 / Регрессионное прогнозирование ч.2

- 10:00 – 11:30 Выбор характеристик. Регуляризация
- 11:45 – 13:15 Оптимизация регрессионных моделей
- 14:30 – 16:00 Лабораторная работа
- 16:00 – 17:00 Лабораторная работа, обсуждение
- 17:00 – 18:00 Мотивационная встреча

### День 5 Проект по регрессионному прогнозированию

- 10:00 – 11:30 Планирование проекта
- 11:45 – 13:15 Реализация проекта
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, обсуждение



## ▶ Программа обучения

### Куаныш Абешев

Ph.D., декан "Школы Цифровых Технологий", AlmaU



## ▶ Неделя 3. Data Classification/Классификация данных

**Цель:** построение предиктивных моделей классификации данных

### День 1 Evaluating Performance - Classification / Оценка производительности - Классификация

- 10:00 – 11:30 Метрики для задач классификации.
- 11:45 – 13:15 Логистическая регрессия (бинарная, многоклассовая)
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Разбор лабораторных работ., Распределение датасетов для защиты проекта

### День 2 Instance-based learning. Probabilistic classification. / Обучение на примерах. Вероятностный классификатор

- 10:00 – 11:30 Метод K-ближайших соседей (KNN)
- 11:45 – 13:15 Наивный байесовский классификатор (Naive Bayes)
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Разбор лабораторных работ

### День 3 Kernel based models / Модели на основе ядра

- 10:00 – 11:30 Метод опорных векторов (SVM) - классификация
- 11:45 – 13:15 Метод опорных векторов (SVM) - регрессия
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Разбор лабораторных работ

### День 4 Tree based models / Модели на основе дерева

- 10:00 – 11:30 Деревья принятия решений (Decision Tree)
- 11:45 – 13:15 Ансамблевые методы и случайные леса
- 14:30 – 16:00 Лабораторная работа
- 16:00 – 17:00 Разбор лабораторных работ
- 17:00 – 18:00 Мотивационная встреча

### День 5 Проект по задаче классификации

- 10:00 – 11:30 Планирование проекта
- 11:45 – 13:15 Реализация проекта
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, обсуждение



## ▶ Программа обучения

### Айдос Сарсембаев

Ph.D., ассистент профессор ИТУ,  
senior Machine Learning Engineer  
компания re:Point



## ▶ Неделя 4. Clustering and deep learning / Кластеризация и глубокое обучение

**Цель:** освоение теории и выполнение прикладных задач в unsupervised learning;  
освоение методов глубокого обучения и нейронных сетей

### День 1 Unsupervised learning / Обучение без учителя

- 10:00 – 11:30 Обучение без учителя: кластеризация, алгоритм K-Means, PCA
- 11:45 – 13:15 DBScan, Hierarchical clustering. Метрики качества
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, обсуждение

### День 2 Deep learning (MLP, ANN) / Глубокое обучение (MLP, ANN)

- 10:00 – 11:30 Введение в нейронные сети. Многослойный перцептрон
- 11:45 – 13:15 Алгоритм обратного распространения ошибки (Backpropagation)
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, обсуждение

### День 3 Deep Learning with PyTorch / Глубокое обучение в библиотеке PyTorch

- 10:00 – 11:30 Введение в библиотеку PyTorch
- 11:45 – 13:15 Сверточные нейронные сети (CNN). Классификация изображений. Метрики
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, обсуждение

### День 4 Deep Learning on image data. Segmentation / Глубокое обучение на изображениях. Сегментация изображений

- 10:00 – 11:30 Модели сегментации изображений. Подготовка данных
- 11:45 – 13:15 Модели сегментации изображений. Построение модели, тренировка, метрики качества
- 14:30 – 16:00 Лабораторная работа
- 16:00 – 18:00 Лабораторная работа, обсуждение

### День 5 Deep Learning on image data. Object Detecion / Глубокое обучение на изображениях. Детекция объектов

- 10:00 – 11:30 Модели детекции объектов на изображениях. Подготовка данных
- 11:45 – 13:15 Модели детекции объектов на изображениях. Построение модели, тренировка, метрики качества
- 14:30 – 16:00 Лабораторная работа
- 16:00 – 17:00 Лабораторная работа, обсуждение
- 17:00 – 18:00 Закрытие Yessenov Data Lab



Мы: [yessenovfoundation.org](https://yessenovfoundation.org)  
ВСЕГДА: [info@yessenovfoundation.org](mailto:info@yessenovfoundation.org)  
НА СВЯЗИ: +7 771 759 5944