



SHAKHMARDAN YESSENOV  
FOUNDATION



**YESSENOV  
DATA LAB**

13 июля – 14 августа 2020  
Алматы

## Оглавление

- Что такое летняя школа Yessenov data lab 2020?
- Программа обучения
  - ▶ Неделя 1. Язык программирования Python
  - ▶ Неделя 2. Линейные модели классификации и регрессии
  - ▶ Неделя 3. Анализ данных, работа с признаками и деревья
  - ▶ Неделя 4. Кластеризация и нейронные сети
  - ▶ Неделя 5. Компьютерное зрение

## ▶ Что такое Yessenov Data Lab 2020?

Летняя школа Yessenov Data Lab 2020 это 5-недельный интенсив, по обучению казахстанских молодых ученых базовым навыкам анализа больших данных для дальнейшего применения знаний в реальных практических работах, а также развития их коммуникативных навыков

**Даты проведения:** 13 июля – 14 августа 2020

### **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)
4. Визуализирования данных и нахождения в них зависимости
5. Составления прогноза на основании исторических данных
6. Понимания различных алгоритмов обучения
7. Правильного выбора модели обучения



## Программа обучения

### Неделя 1. Язык программирования Python

#### День 1

- 10:00 – 10:15 Открытие летней школы, приветственные слова
- 10:15 – 11:30 Что такое Data Science, Machine Learning, примеры
- 11:45 – 13:15 Введение в Python: переменные, списки
- 14:30 – 16:00 Введение в Python, условия, циклы
- 16:15 – 18:00 Лабораторная работа, обсуждение ошибок

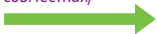
#### День 2

- 10:00 – 11:30 Работа со строками. Latex, Markdown
- 11:45 – 13:15 Знакомство с библиотекой NumPy
- 14:30 – 16:00 Лабораторная работа: задачи по NumPy
- 16:15 – 18:00 Лабораторная работа, обсуждение ошибок

#### День 3

- 10:00 – 11:30 Структуры данных: списки, множества, словари
- 11:45 – 13:15 Алгоритмы на языке Python: сортировки, поиск
- 14:30 – 16:00 Лабораторная работа: структуры данных и NumPy
- 16:15 – 18:00 Лабораторная работа, обсуждение ошибок

#### День 4

- 10:00 – 11:30 Знакомство с библиотекой Pandas, Matplotlib
- 11:45 – 13:15 Группировка данных. Фильтры, сортировка.
- 14:30 – 16:00 Лабораторная работа: работа с Pandas, Matplotlib.
- 16:15 – 18:00 Современное рабочее пространство, совместная/командная работа и коммуникация  Спикер

#### День 5

- 10:00 – 11:30 Закрепление материала по Pandas и Matplotlib
- 11:45 – 13:15 Знакомство с библиотекой Seaborn
- 14:30 – 16:00 Лабораторная работа: Визуализация
- 16:15 – 18:00 Лабораторная работа, обсуждение ошибок



**Куаныш Абешев**

декан Школы инженерного менеджмента, AlmaU



**Алмас Молдаканов**

Менеджер по работе с сектором образования в Казахстане и странах СНГ, Microsoft





## Неделя 2. Линейные модели классификации и регрессии



**Назым Сатбекова**

специалист по машинному обучению, Kolesa Group

### День 1

- 10:00 – 11:30 Линейные модели регрессии (Simple Linear Regression, MultReg, PolyReg)
- 11:45 – 13:15 Метрики качества (MSE, MAE, RMSE)
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, обсуждение ошибок

### День 2

- 10:00 – 11:30 Задача классификации
- 11:45 – 13:15 Метод ближайших соседей (kNN), Логистическая регрессия
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, обсуждение ошибок

### День 3

- 10:00 – 11:30 Вероятностные методы. Наивный Байес
- 11:45 – 13:15 Лабораторная работа
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, обсуждение ошибок

### День 4

- 10:00 – 11:30 Метод опорных векторов (SVM)
- 11:45 – 13:15 Лабораторная работа
- 14:30 – 18:00 Лабораторная работа, обсуждение ошибок

### День 5

- 10:00 – 11:30 Методы статистического анализа
- 11:45 – 13:15 Проектная работа
- 14:30 – 16:00 Проектная работа
- 16:15 – 18:00 Защита результатов, обсуждение ошибок





## Неделя 3. Анализ данных, работа с признаками и деревья

### День 1

10:00 – 11:30	Регрессия и классификация деревьями
11:45 – 13:15	Корреляция и ковариация
14:30 – 16:00	Лабораторная работа
16:15 – 18:00	Лабораторная работа, обсуждение ошибок

### День 2

10:00 – 11:30	Ансамбли: бэггинг, бустинг, случайный лес
11:45 – 13:15	Лабораторная работа
14:30 – 16:00	Лабораторная работа
16:15 – 18:00	Лабораторная работа, обсуждение ошибок

### День 3

10:00 – 11:30	Информативность, построение и отбор признаков
11:45 – 13:15	Семплирование данных (sampling)
14:30 – 16:00	Лабораторная работа
16:15 – 18:00	Лабораторная работа, обсуждение ошибок

### День 4

10:00 – 11:30	Ансамбли: Дилемма смещения–дисперсии, XGBoost
11:45 – 13:15	Лабораторная работа
14:30 – 16:00	Лабораторная работа, обсуждение ошибок
16:15 – 18:00	Личный бренд в профессиональных социальных медиа помогает открывать новые карьерные возможности. Как создать эффективный профайл в LinkedIn?



### День 5

10:00 – 11:30	Градиентный спуск на деревьях: CatBoost
11:45 – 13:15	Проектная работа
14:30 – 16:00	Проектная работа
16:15 – 18:00	Защита результатов, обсуждение ошибок



**Дархан Аймаганов**  
аналитик продукта, Arbutz.kz



**Алмас Молдаканов**  
Менеджер по работе с сектором образования в Казахстане и странах СНГ, Microsoft





## Неделя 4. Кластеризация и нейронные сети

### День 1

- 10:00 – 11:30 Проверка статистических гипотез
- 11:45 – 13:15 Статистические модели: Анализ выживаемости
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, обсуждение ошибок

### День 2

- 10:00 – 11:30 Обучение без учителя: PCA, кластеризация
- 11:45 – 13:15 Алгоритм K-Means
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, обсуждение ошибок

### День 3

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

### День 4

- 10:00 – 11:30 Введение в библиотеку Keras
- 11:45 – 13:15 Сверточные нейронные сети
- 14:30 – 16:00 Лабораторная работа, обсуждение ошибок
- 16:00 – 18:00 Атлас новых профессий и компетенций Казахстана.  
Роль науки в формировании перспективных  
отраслей и профессий



Спикер

### День 5

- 10:00 – 11:30 Проблема переобучения нейронных сетей, аугментация данных
- 11:45 – 13:15 Проектная работа
- 14:30 – 16:00 Проектная работа
- 16:15 – 18:00 Защита результатов, обсуждение ошибок



**Айдос Сарсембаев**  
ассистент-профессор, МУИТ



**Саясат Нурбек**  
Руководитель BTS Education





## Неделя 5. Компьютерное зрение (Computer Vision)

### День 1

- 10:00 – 11:30 Intro to CV
- 11:45 – 13:15 Intro to CNNs
- 14:30 – 16:00 Лабораторная работа - Intro to Pytorch
- 16:15 – 18:00 Лабораторная работа - Intro to CNNs

### День 2

- 10:00 – 11:30 Regularization (Dropout, Batchnorm, Augmentations)
- 11:45 – 13:15 Classification Architectures
- 14:30 – 16:00 Лабораторная работа - ML Pipeline
- 16:15 – 18:00 Лабораторная работа - Dataloaders, Transfer Learning

### День 3

- 10:00 – 11:30 Optimization (Optimizers)
- 11:45 – 13:15 Detection Architectures
- 14:30 – 16:00 Лабораторная работа - Detection Pipeline
- 16:15 – 18:00 Лабораторная работа - Segmentation Pipeline, Intro to Project

### День 4

- 10:00 – 11:30 Segmentation Architectures
- 11:45 – 13:15 Проектная работа - Pneumothorax Segmentation
- 14:30 – 16:00 Проектная работа - Pneumothorax Segmentation
- 16:15 – 18:00 Проектная работа - Pneumothorax Segmentation

### День 5

- 10:00 – 11:30 RNNs for CV (optional)
- 11:45 – 13:15 Проектная работа - Pneumothorax Segmentation
- 14:30 – 17:00 Командные презентации результатов
- 17:00 – 17:30 Официальное закрытие



**Рустем Бурханов**

Computer Vision Engineer,  
BTS Digital







## ПРОГРАММЫ ФОНДА



Научные  
стажировки



IT



Стипендия  
Есенова



Есеновские  
чтения



Английский  
язык



Магистратура



Книги  
Конференции



Алматы  
марафон



АQT

**МЫ:** [yessenovfoundation.org](http://yessenovfoundation.org)

**ВСЕГДА:** [info@yessenovfoundation.org](mailto:info@yessenovfoundation.org)

**НА СВЯЗИ:** **+7 727 346 92 88**





**YESSENOV  
DATA LAB**



**SHAKHMARDAN YESSENOV  
FOUNDATION**