



SHAKHMARDAN YESSENOV
FOUNDATION



YESSENOV
DATA LAB

17 июня -
13 июля

Алматы 2024

Оглавление

- Что такое летняя школа Yessenov Data Lab 2024?
- Программа обучения:
 - ▶ Неделя 1. Python для данных
 - ▶ Неделя 2. Регрессия и аналитика
 - ▶ Неделя 3. Классификация данных
 - ▶ Неделя 4. Глубокое обучение и компьютерное зрение





Что такое Yessenov Data Lab 2024?

Летняя школа Yessenov Data Lab 2024 - это 4-недельный интенсив.
3 недели онлайн обучение,
1 неделя офлайн обучение (в г. Алматы).

Даты проведения: 17 июня – 13 июля 2024

ЗА 4 НЕДЕЛИ ШКОЛЫ ВЫ СМОЖЕТЕ:

- обучиться навыкам анализа больших данных, которые сможете применять на практике;
- пообщаться с лекторами, задать интересующие вопросы;
- познакомиться с единомышленниками, установить новые связи.

Помимо сертификата школы YDL отличники обучения, получат академический транскрипт/сертификат от AlmaU, приравненный к сертификату повышения квалификации, который студентам идет в зачёт обучения.

- Расходы по проезду до Алматы и обратно, питанию и проживанию победителей конкурса Yessenov Data Lab, живущих за пределами г. Алматы, в период с 8 по 13 июля, фонд берет на себя. Питанием обеспечиваются все участники школы в период офлайн-обучения.



Online



YDL YESSENOV
DATA LAB

Программа обучения

Лектор:

Тимур Бакибаев

Ph.D. по информатике, профессор AlmaU



Неделя 1. Python для данных

Цель: научиться основам языка программирования Python для анализа данных

День 1 Искусственный интеллект и работа с ChatGPT

- 10:00 – 10:10 Открытие летней школы, приветственные слова
- 10:10 – 11:30 Что такое Искусственный интеллект
- 11:45 – 13:15 Принципы работы ChatGPT
- 14:30 – 16:00 Лабораторная работа с ChatGPT API
- 16:15 – 18:00 Разбор лабораторных работ

День 2 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 3 Визуализация данных

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

День 4 Основы прогнозирования

- 10:00 – 11:30 Знакомство с задачей регрессии
- 11:45 – 13:15 Важность признаков при построении модели
- 14:30 – 16:00 Лабораторная работа: анализ и предобработка данных
- 16:00 – 17:00 Разбор лабораторных работ
- 17:00 – 18:00 Мотивационная встреча

День 5 StreamLit: 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)

SHAKHMARDAN YESSENOV
FOUNDATION





Программа обучения

Лектор:

Кирилл Якунин

Ph.D. по вычислительной технике и программному обеспечению,
ML Engineer в Metaculus



Неделя 2. Регрессия и аналитика

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

День 1 Статистический анализ

| | |
|---------------|---|
| 10:00 – 11:30 | Разведывательный анализ, гистограммы, выбросы, распределения, корреляционный анализ |
| 11:45 – 13:15 | Тестирование гипотез. One-sample test. Two-sample test |
| 14:30 – 16:00 | Лабораторная работа |
| 16:15 – 18:00 | Лабораторная работа, обсуждение |

День 2 Regression modelling p.1

| | |
|---------------|---|
| 10:00 – 11:30 | Основы линейной регрессии. Метрики. Верификация данных. |
| 11:45 – 13:15 | Предобработка данных. |
| 14:30 – 16:00 | Лабораторная работа |
| 16:15 – 18:00 | Лабораторная работа, обсуждение |

День 3 Regression modelling p.2

| | |
|---------------|--|
| 10:00 – 11:30 | Полиномиальная регрессия. Регуляризация регрессионных моделей. |
| 11:45 – 13:15 | Градиентный спуск. |
| 14:30 – 16:00 | Лабораторная работа |
| 16:15 – 18:00 | Лабораторная работа, обсуждение |

День 4 Обучение без учителя

| | |
|---------------|---|
| 10:00 – 11:30 | Важность признаков. |
| 11:45 – 13:15 | Обучение без учителя. Понижение размерности |
| 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 | Полный проект машинного обучения |
| 11:45 – 13:15 | Q&A |
| 14:30 – 16:00 | Защита проекта. |
| 16:15 – 18:00 | Защита проекта и обсуждение результатов. |



Online



YEDL
YESSENOV
DATA LAB

Программа обучения

Лектор:

Куаныш Абешев

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



Неделя 3. Классификация данных

Цель: построение предиктивных моделей классификации данных

День 1 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 Обучение на примерах. Вероятностный классификатор

- 10:00 – 11:30 Метод К-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 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 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 Метод принципиальных компонент (PCA)
- 11:45 – 13:15 Лабораторная работа.
- 14:30 – 16:00 Работа над проектом.
- 16:15 – 18:00 Выступление и защита проекта.

SHAKHMARDAN YESSENOV
FOUNDATION



Offline

Алматы



Программа обучения

Лектор:

Владислав Ющенко

магистр информатики, старший разработчик машинного обучения в Care Analytics



Неделя 4. Глубокое обучение и компьютерное зрение

Цель: освоение методов глубокого обучения и нейронных сетей

День 1 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 2 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 3 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
- 16:15 – 18:00 Лабораторная работа, обсуждение

День 4 Глубокое обучение на изображениях. Детекция объектов

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

День 5 Введение в MLOps

- 10:00 – 11:30 Введение в MLOps, этапы автоматизации.
- 11:45 – 13:15 Трекинг экспериментов и развертывание моделей. Передовой опыт применения MLOps
- 14:30 – 16:00 Лабораторная работа
- 16:00 – 17:00 Лабораторная работа, обсуждение
- 17:00 – 18:00 Мотивационная встреча

День 6

- 10:00 – 16:00 Закрытие школы. Тим билдинг. Нетворкинг.

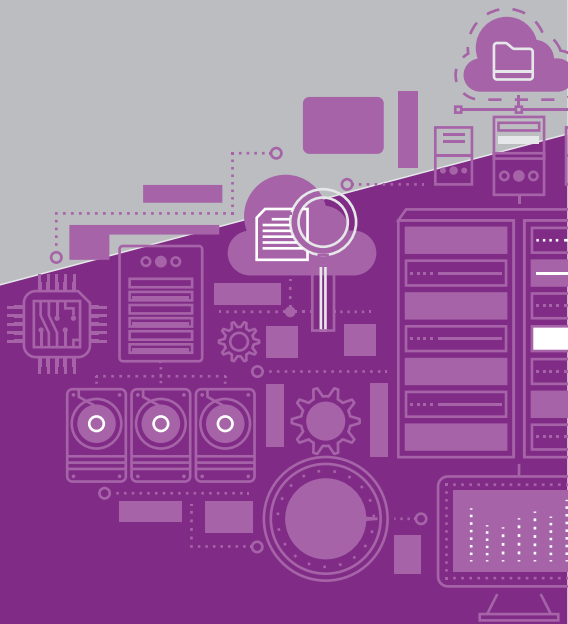




**YESSENOV
DATA LAB**



**SHAKHMARDAN YESSENOV
FOUNDATION**



Мы: yessenovfoundation.org
ВСЕГДА: info@yessenovfoundation.org
НА СВЯЗИ: **+7 771 759 5944**