



YESSENOV
DATA LAB

10 июня – 2 августа 2019
Алматы

Оглавление

- Что такое летняя школа Yessenov data lab?
- Программа обучения
 - ▶ Неделя 1. Язык программирования Python
 - ▶ Неделя 2. Линейные модели классификации и регрессии
 - ▶ Неделя 3. Деревья и Boosting
 - ▶ Неделя 4. Обучение без учителя и нейронные сети
 - ▶ Неделя 5. Приложения ML
 - ▶ Неделя 6. Big Data - I
 - ▶ Неделя 7. Big Data - II
 - ▶ Неделя 6. Проектные задачи

▶ Что такое Yessenov Data Lab?

Летняя школа Yessenov data lab это 8-недельный интенсив, который позволяет за 8 недель ознакомиться с профессией аналитика данных, решать реальные задачи бизнеса и в дальнейшем самостоятельно совершенствовать полученные знания.



Даты проведения: 10 июня – 2 августа 2019

Место проведения:
Almaty Management University

ВЫПУСКНИКИ ШКОЛЫ ПРИОБРЕТУТ НАВЫКИ:

1. Программирования на языке Python в рамках анализа данных
2. Предварительной обработки данных, подготовки их для последующего анализа (preprocessing)
3. Визуализирования данных и нахождения в них зависимости
4. Составления прогноза на основании исторических данных
5. Понимания различных алгоритмов обучения
6. Правильного выбора модели обучения
7. Фундаментальное понимание работы нейронных сетей



Программа обучения

Неделя 1. Язык программирования Python

День 1

- 09:00 – 10:00 Регистрация участников
- 10:00 – 11:30 Что такое Data Mining, Big Data, примеры
- 12:00 – 13:15 Введение в Python: переменные, списки
- 14:30 – 16:00 Введение в Python, условия, циклы
- 16:15 – 18:00 Лабораторная работа: основы Python

День 2

- 10:00 – 11:30 Работа со строками
- 12:00 – 13:15 Знакомство с библиотекой NumPy
- 14:30 – 16:00 Лабораторная работа: задачи по NumPy
- 16:15 – 18:00 Лабораторная работа: задачи по NumPy

День 3

- 10:00 – 11:30 Структуры данных: списки, множества, словари
- 12:00 – 13:15 Алгоритмы на языке Python: сортировки, поиск
- 14:30 – 16:00 Лабораторная работа: структуры данных и NumPy
- 16:15 – 18:00 Лабораторная работа: реализация алгоритмов

День 4

- 10:00 – 11:30 Знакомство с библиотекой Pandas, Matplotlib
- 12:00 – 13:15 Группировка данных. Фильтры, сортировка.
- 14:30 – 16:00 Лабораторная работа: работа с Pandas.
- 16:15 – 18:00 Лабораторная работа: работа с Matplotlib.

День 5

- 10:00 – 11:30 Закрепление материала по Pandas и Matplotlib
- 12:00 – 13:15 Знакомство с задачей регрессии, SciKitLearn.
- 14:30 – 16:00 Лабораторная работа: Линейная регрессия
- 16:15 – 18:00 Лабораторная работа: Линейная регрессия

- 11:30 – 11:45 Кофе-брейк
- 13:15 – 14:30 Обед
- 16:00 – 16:15 Кофе-брейк



Куаныш Абешев
декан Школы
инженерного
менеджмента, AlmaU



Тимур Бакибаев
Профессор, AlmaU



Программа обучения

Неделя 2. Линейные модели классификации и регрессии

День 1

- 10:00 – 11:30 Задача оптимизации, метод градиентного спуска
- 12:00 – 13:15 Лабораторная работа
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа

День 2

- 10:00 – 11:30 Метрики качества.
- 12:00 – 13:15 Лабораторная работа
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа

День 3

- 10:00 – 11:30 Кросс-валидация
- 12:00 – 13:15 Лабораторная работа
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа

День 4

- 10:00 – 11:30 Линейные модели классификации и регрессии.
- 12:00 – 13:15 Лабораторная работа
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа

День 5

- 10:00 – 11:30 Переобучение, генерализация.
- 12:00 – 13:15 Лабораторная работа
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа

- 11:30 – 11:45 Кофе-брейк
- 13:15 – 14:30 Обед
- 16:00 – 16:15 Кофе-брейк



Дмитрий Русанов
Data Scientist



▶ Программа обучения

Неделя 3. Деревья и Boosting



День 1

- 10:00 – 11:30 SVM, задачи оптимизации
- 12:00 – 13:15 Практика, мастер-класс
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, Q&A, обсуждение стандартных ошибок

День 2

- 10:00 – 11:30 Деревья
- 12:00 – 13:15 Практика, мастер-класс
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, Q&A, обсуждение стандартных ошибок

День 3

- 10:00 – 11:30 Ансамбли решающих деревьев (bagging)
- 12:00 – 13:15 Практика, мастер-класс
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, Q&A, обсуждение стандартных ошибок

День 4

- 10:00 – 11:30 Информативность признаков, SHAP (SHapley Additive exPlanations)
- 12:00 – 13:15 Практика, мастер-класс
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, Q&A, обсуждение стандартных ошибок

День 5

- 10:00 – 11:30 Ансамбли решающих деревьев (bias-variance trade-off, XGBoost)
- 12:00 – 13:15 Практика, мастер-класс
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, Q&A, обсуждение стандартных ошибок

День 6

- 10:00 – 11:30 Подробный обзор CatBoost
- 12:00 – 13:15 Практика, мастер-класс
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, Q&A, обсуждение стандартных ошибок

- 11:30 – 11:45 Кофе-брейк
- 13:15 – 14:30 Обед
- 16:00 – 16:15 Кофе-брейк



Программа обучения

Неделя 4. Обучение без учителя и нейронные сети

День 1

- 10:00 – 11:30 Обучение без учителя (Unsupervised Learning)
- 12:00 – 13:15 Практика, мастер-класс
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, Q&A, обсуждение стандартных ошибок

День 2

- 10:00 – 11:30 Unsupervised Learning, ч.2
- 12:00 – 13:15 Практика, мастер-класс
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, Q&A, обсуждение стандартных ошибок

День 3

- 10:00 – 11:30 Введение в нейронные сети
- 12:00 – 13:15 Практика, мастер-класс
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, Q&A, обсуждение стандартных ошибок

День 4

- 10:00 – 11:30 Сверточные нейронные сети
- 12:00 – 13:15 Практика, мастер-класс
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, Q&A, обсуждение стандартных ошибок

День 5

- 10:00 – 11:30 Рекуррентные нейронные сети
- 12:00 – 13:15 Практика, мастер-класс
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, Q&A, обсуждение стандартных ошибок

День 6

- 10:00 – 11:30 Приложения ML: RecSys
- 12:00 – 13:15 Практика, мастер-класс
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа, Q&A, обсуждение стандартных ошибок

11:30 – 11:45 Кофе-брейк

13:15 – 14:30 Обед

16:00 – 16:15 Кофе-брейк



▶ Программа обучения

Неделя 5. Приложения ML



Марина Горлова
Analyst



День 1

- 10:00 – 11:30 Рекомендательные системы. User/Item based.
Коллаборативная фильтрация
- 12:00 – 13:15 Рекомендательные системы. Гибридный recommender
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа

День 2

- 10:00 – 11:30 Анализ изображений. Как решали задачу до нейронных сетей
- 12:00 – 13:15 Анализ изображений. Приложения нейронных сетей
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа

День 3

- 10:00 – 11:30 Анализ текстов. Эмбединги
- 12:00 – 13:15 Анализ текстов. Приложения
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа

День 4

- 10:00 – 11:30 Временные ряды. Статистические подходы
- 12:00 – 13:15 Временные ряды. Приложения нейронных сетей
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа

День 5

- 10:00 – 11:30 Анализ графов
- 12:00 – 13:15 Лабораторная работа
- 14:30 – 16:00 Use cases
- 16:15 – 18:00 Use cases

- 11:30 – 11:45 Кофе-брейк
- 13:15 – 14:30 Обед
- 16:00 – 16:15 Кофе-брейк

▶ Программа обучения

Неделя 6. Big Data - I

День 1

- 10:00 – 11:30 Бизнес проблемы
- 12:00 – 13:15 NoSQL
- 14:30 – 16:00 Hadoop: Yarn, MapReduce, Hive, Hdfs, Spark
- 16:15 – 18:00 Продолжение

День 2

- 10:00 – 11:30 MapReduce Internals + Hive
- 12:00 – 13:15 HiveQL Basics (UI or CLI?)
- 14:30 – 16:00 Лабораторная работа
- 16:15 – 18:00 Лабораторная работа

День 3

- 10:00 – 11:30 Spark Core
- 12:00 – 13:15 Практические задания по pySpark + RDD
- 14:30 – 16:00 Практические задания по pySpark + RDD
- 16:15 – 18:00 Практические задания по pySpark + RDD

День 4

- 10:00 – 11:30 Spark SQL
- 12:00 – 13:15 Практические задания по pySpark + DataFrame
- 14:30 – 16:00 Практические задания по pySpark + DataFrame
- 16:15 – 18:00 Практические задания по pySpark + DataFrame

День 5

- 10:00 – 11:30 Разница в использовании ресурсов Compute & Storage
- 12:00 – 13:15 Аналоги Инструментов
- 14:30 – 16:00 Case studies
- 16:15 – 18:00 Case studies

- 11:30 – 11:45 Кофе-брейк
- 13:15 – 14:30 Обед
- 16:00 – 16:15 Кофе-брейк



Валерий Жук
Senior Cloud Computing Engineer



▶ Программа обучения

Неделя 7. Big Data - II

День 1

- 10:00 – 11:30 Бизнес-проблемы
- 12:00 – 13:15 Проблема применения Data Science на больших данных
- 14:30 – 16:00 Обзор инструментов
- 16:15 – 18:00 Лабораторная работа

День 2

- 10:00 – 11:30 Spark MLlib
- 12:00 – 13:15 Spark MLlib (продолжение)
- 14:30 – 16:00 Практические задания по pySpark + MLlib
- 16:15 – 18:00 Практические задания по pySpark + MLlib

День 3

- 10:00 – 11:30 Анализ данных на графах
- 12:00 – 13:15 Обзор GraphX
- 14:30 – 16:00 Практические задания по pySpark + GraphX
- 16:15 – 18:00 Практические задания по pySpark + GraphX

День 4

- 10:00 – 11:30 Data Science on Cloud
- 12:00 – 13:15 Общие вопросы: откуда брать данные?
- 14:30 – 16:00 Общие вопросы: внедрение решения
- 16:15 – 18:00 Общие вопросы: работа аналитиком данных

День 5

- 10:00 – 11:30 Case studies
- 12:00 – 13:15 Case studies
- 14:30 – 16:00 Case studies
- 16:15 – 18:00 Case studies

- 11:30 – 11:45 Кофе-брейк
- 13:15 – 14:30 Обед
- 16:00 – 16:15 Кофе-брейк



Михаил Липкович
Senior Data Scientist


assaia
The Apron AI

► Программа обучения

Неделя 8. Проектные задачи

Проектная неделя предполагает
6 часов работы в день

ПРОГРАММЫ ФОНДА

Знания	Наука	Ресурсы
 Есеновские чтения	 Научные стажировки	 Развитие ИТ компетенций
 Английский язык	 Стипендия Есенова	 Команда SOS
 Популяризация науки	 Обучение в магистратуре	 Повышение личной эффективности
 Казахстанская федерация шахмат	 Алматы марафон	 Алматинская федерация триатлона

МЫ: yessenovfoundation.org

ВСЕГДА: info@yessenovfoundation.org

НА СВЯЗИ: +7 727 346 92 88





**YESSENOV
DATA LAB**



**SHAKHMARDAN YESSENOV
FOUNDATION**



В партнерстве с: **Яндекс**

