



**YESSENOV
DATA LAB**

12 маусым - 08 шілде
2023 ж
Алматы



▶ Мазмұны

- Yessenov Data Lab 2023 жазғы мектебі дегеніміз не?
- Оқыту бағдарламасы
 - ▶ 1 апта. Python бағдарламалау тілі
 - ▶ 2 апта. Деректерді талдау. Регрессиялық мәліметтерді талдау
 - ▶ 3 апта. Деректер классификациясы
 - ▶ 4 апта. Кластерлеу және терең оқыту



▶ Yessenov Data Lab 2023 жазғы мектебі дегеніміз не?

Yessenov Data Lab 2023 жазғы мектебі - бұл білімді нақты практикалық жұмыста қолдану үшін қазақстандық жас ғалымдар мен мамандарды үлкен деректерді талдаудың негізгі дағдыларына үйрету бойынша 4 апталық қарқынды оқыту.

Өткізу күні: "2023" сөзін жойыңыз 12 маусым – 08 шілде 2023 ж

МЕКТЕП ТҮЛЕКТЕРІ ТӨМЕНДЕГІДЕЙ ДАҒДЫЛАРҒА ИЕ БОЛАДЫ:

- 1. Деректерді талдау аясында Python тілінде бағдарламалау
- 2. Статистикалық талдау әдістері
- 3. Деректерді алдын ала өңдеу, оларды кейінгі талдау үшін дайындау (preprocessing) және тәуелділіктерін табу. Болжам жасау
- 4. Деректерді визуализациялау және оларға тәуелділік себебін табу
- 5. Тарихи деректер негізінде болжам жасау
- 6. Машиналық оқытудың әртүрлі алгоритмдерін түсіну
- 7. Оқыту моделін дұрыс таңдау



▶ Оқыту бағдарламасы

Тимур Бакибаев

Ph.D., AlmaU профессоры
DSA Engineering



▶ 1 апта. Python бағдарламалау тілі

Мақсаты: Деректермен жұмыс істеу үшін Python бағдарламалау тілінің озық тәжірибелерін үйрену

1-күн **Объектіге бағытталған бағдарламалау**

- 10:00 – 10:10 Data Science, Machine Learning
- 10:10 – 11:30 Деректер туралы ғылым, Машиналық оқыту дегеніміз не?
- 11:45 – 13:15 Python: ОББ, типтік аннотациялар және тесттер
- 14:30 – 16:00 ОББ бойынша зертханалық жұмыс
- 16:15 – 18:00 Зертханалық жұмыстарды талдау

2-күн **Графтар теориясы**

- 10:00 – 11:30 Python тілінде графтарды жүзеге асыру
- 11:45 – 13:15 Кеңдігінен іздеу - Дейкстра алгоритмі
- 14:30 – 16:00 Зертханалық жұмыс: кеңдігінен іздеу
- 16:15 – 18:00 Зертханалық жұмыстарды талдау

3-күн **Деректермен жұмыс**

- 10:00 – 11:30 NumPy, Pandas кітапханаларына шолу
- 11:45 – 13:15 Деректерді топтастыру. Сүзгілер, сұрыптау.
- 14:30 – 16:00 Зертханалық жұмыс: Pandas-пен жұмыс
- 16:15 – 18:00 Зертханалық жұмыстарды талдау

4-күн **Деректерді визуализациялау**

- 10:00 – 11:30 Matplotlib кітапханасына шолу
- 11:45 – 13:15 Seaborn кітапханасына шолу
- 14:30 – 16:00 Зертханалық жұмыс: бейнелеу
- 16:00 – 17:00 Зертханалық жұмыстарды талдау
- 17:00 – 18:00 Мотивациялық кездесу

5-күн **StreamLit: деректерді жинауға арналған Интернет**

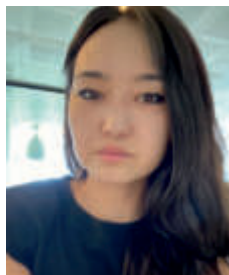
- 10:00 – 11:30 StreamLit-пен танысу Шолуы
- 11:45 – 13:15 Зертханалық жұмыс: StreamLit
- 14:30 – 16:00 Зертханалық жұмыс: StreamLit
- 16:00 – 18:00 Серверде StreamLit-ті орналастыру



▶ Оқыту бағдарламасы

Әйгерім Садықова

Информатика және бағдарламалық жасақтама саласындағы бакалавр дәрежесі, PwC компаниясының бұрынғы аға деректер талдаушысы



▶ 2 апта. Деректерді талдау. Регрессиялық модельдеу

Мақсаты: статистикалық талдау теориясы мен дағдыларын меңгеру; болжамды регрессиялық модельдерді құру

1 күн Деректерді барлап талдау

- 10:00 – 11:30 Барлап талдау, гистограммалар, шығарындылар
- 11:45 – 13:15 Үлестіру, корреляциялық талдау
- 14:30 – 16:00 Зертханалық жұмыс
- 16:15 – 18:00 Зертханалық жұмыс, талқылау

2 күн Гипотезаны тексеру

- 10:00 – 11:30 Гипотезаларды тексеру. Бір үлгі сынағы
- 11:45 – 13:15 Гипотезаларды тексеру. Екі үлгі сынағы
- 14:30 – 16:00 Зертханалық жұмыс
- 16:15 – 18:00 Зертханалық жұмыс, талқылау

3 күн Регрессиялық модельдеу 1-бет

- 10:00 – 11:30 Сызықтық регрессияның түрлері
- 11:45 – 13:15 Регрессия есептерінің көрсеткіштері
- 14:30 – 16:00 Зертханалық жұмыс
- 16:15 – 18:00 Зертханалық жұмыс, талқылау

4 күн Регрессиялық модельдеу 2-бет

- 10:00 – 11:30 Сипаттамаларды таңдау. Регуляризация
- 11:45 – 13:15 Регрессиялық модельдерді оңтайландыру
- 14:30 – 16:00 Зертханалық жұмыс
- 16:00 – 17:00 Зертханалық жұмыс, талқылау
- 17:00 – 18:00 Мотивациялық кездесу

5 күн Регрессиялық модельдеу жобасы

- 10:00 – 11:30 Жобаны жоспарлау
- 11:45 – 13:15 Жобаны іске асыру
- 14:30 – 16:00 Зертханалық жұмыс
- 16:15 – 18:00 Зертханалық жұмыс, талқылау



▶ Оқыту бағдарламасы

Қуаныш Әбешев

Ph.D., Сандық технологиялар
мектебінің деканы



▶ 3 апта. Деректерді жіктеу

Мақсаты: Болжамдық мәліметтерді классификациялау модельдерін құру

1-күн Өнімділікті бағалау - классификация

- 10:00 – 11:30 Классификация метрикалары
- 11:45 – 13:15 Логистикалық регрессия (екілік және көп класты)
- 14:30 – 16:00 Зертханалық жұмыстар
- 16:15 – 18:00 Зертханалық жұмыстарды талдау. Жобаны қорғау үшін деректер жиынын тарату

2-күн Мысалдар негізінде оқыту. Біқтималдық классификациясы

- 10:00 – 11:30 К-ең жақын көршілер (KNN) әдісі
- 11:45 – 13:15 Naive Bayes классификаторы
- 14:30 – 16:00 Зертханалық жұмыстар
- 16:15 – 18:00 Зертханалық жұмыстарды талдау

3-күн Ядроға негізделген модельдер

- 10:00 – 11:30 Анықтамалық векторлық машиналар әдісі (SVM) - классификация
- 11:45 – 13:15 Анықтамалық векторлық машина әдісі (SVM) - регрессия
- 14:30 – 16:00 Зертханалық жұмыстар
- 16:15 – 18:00 Зертханалық жұмыстарды талдау

4-күн Ағаш негізіндегі модельдер

- 10:00 – 11:30 Шешім ағаштары
- 11:45 – 13:15 Ансамбльдік әдістер және кездейсоқ ормандар
- 14:30 – 16:00 Зертханалық жұмыстар
- 16:00 – 17:00 Зертханалық жұмыстарды талдау
- 17:00 – 18:00 Мотивациялық кездесу

5-күн Классификация жобасы

- 10:00 – 11:30 Жобаны жоспарлау
- 11:45 – 13:15 Жобаны жүзеге асыру
- 14:30 – 16:00 Зертханалық жұмыстар
- 16:15 – 18:00 Зертханалық жұмыстарды талдау



▶ Оқыту бағдарламасы

Айдос Сәрсембаев

Ph.D., ИТУ ассистент профессоры,
Машиналық оқыту бойынша аға
инженер, re:Point компаниясы



▶ 4 апта. Кластерлеу және терең оқыту

Мақсаты: Бақылаусыз оқытуда теорияны меңгеру және қолданбалы тапсырмаларды орындау; терең оқыту әдістері мен нейрондық желілерді меңгеру

1-күн Бақылаусыз оқыту

10:00 – 11:30 Бақылаусыз оқыту; кластерлеу, K-Means, PCA алгоритмі
11:45 – 13:15 DBScan, Hierarchical clustering. Сапа көрсеткіштері
14:30 – 16:00 Зертханалық жұмыс
16:15 – 18:00 Зертханалық жұмыс, талқылау

2-күн Терең оқыту (MLP, ANN)

10:00 – 11:30 Нейрондық желілерге кіріспе. Көп қабатты перцептрон
11:45 – 13:15 Қатені кері тарату алгоритмі (Backpropagation)
14:30 – 16:00 Зертханалық жұмыс
16:15 – 18:00 Зертханалық жұмыс, талқылау

3-күн PyTorch көмегімен терең оқыту

10:00 – 11:30 PyTorch кітапханасына кіріспе
11:45 – 13:15 Конволюциялық нейрондық желілер (CNN). Кескіндерді жіктеу. Көрсеткіштер
14:30 – 16:00 Зертханалық жұмыс
16:15 – 18:00 Зертханалық жұмыс, талқылау

4-күн Кескін деректерін терең зерттеу. Сегменттеу

10:00 – 11:30 Кескінді сегменттеу модельдері. Деректерді дайындау
11:45 – 13:15 Кескінді сегменттеу модельдері. Модель құру, жаттығу, сапа көрсеткіштері
14:30 – 16:00 Зертханалық жұмыс
16:00 – 18:00 Зертханалық жұмыс, талқылау

5-күн Кескін деректерін терең зерттеу. Нысанды анықтау

10:00 – 11:30 Кескіндердегі объектілерді анықтау модельдері. Деректерді дайындау
11:45 – 13:15 Кескіндердегі объектілерді анықтау модельдері. Модель құру, жаттығу, сапа көрсеткіштері
14:30 – 16:00 Зертханалық жұмыс
16:00 – 17:00 Зертханалық жұмыс, талқылау
17:00 – 18:00 Yessenov Data Lab жабылуы



БІЗ: yessenovfoundation.org
ӘРҚАШАН: info@yessenovfoundation.org
БАЙЛАНЫСТАМЫЗ: +7 771 759 5944