

СИЛЛАБУС
Осенний семестр 2025 – 2026 учебного года
Образовательная программа «6В06102 Информационные системы»

ID и наименование дисциплины	Самостоятельная работа обучающегося (СРО)	Кол-во кредитов			Общее кол-во кредитов	Самостоятельная работа обучающегося под руководством преподавателя (СРОП)			
		Лекции (Л)	Практ. занятия (ПЗ)	Лаб. занятия (ЛЗ)					
101293, Интеллектуальный анализ данных	4	1,5	1,5	6,0	5	4			
АКАДЕМИЧЕСКАЯ ИНФОРМАЦИЯ О ДИСЦИПЛИНЕ									
Формат обучения	Цикл, модуль компонент	Типы лекций	Типы практических занятий	Форма и платформа итогового контроля					
Офлайн	П КпВ	Обзорные лекции – основные понятия интеллектуального анализа данных (Data Mining), этапы и методы	Построение моделей классификации, кластеризации, регрессии на Python/R	Устный офлайн.					
Лектор - (ы)	Байкувеков Мейржан Берикович								
e-mail:	baikuvekov_meirzhan3@live.kaznu.kz								
Телефон:	+7 702 6478122								
Ассистент- (ы)	Бейбітхан Еркегүл								
e-mail:	Beibutkhan_Yerkegul@live.kaznu.kz								
Телефон:	+77016289523								
АКАДЕМИЧЕСКАЯ ПРЕЗЕНТАЦИЯ ДИСЦИПЛИНЫ									
Цель дисциплины	Ожидаемые результаты обучения (РО)*			Индикаторы достижения РО (ИД)					
Цель дисциплины – сформировать у студентов фундаментальные знания и практические навыки в области методов и технологий интеллектуального анализа данных, научить применять современные алгоритмы и инструменты для обработки, анализа, интерпретации и визуализации больших данных, а также разработки аналитических решений для поддержки принятия решений в различных предметных областях.	1. Студент знает основные понятия, этапы и задачи интеллектуального анализа данных и процесса KDD.			1.1 Объясняет ключевые этапы процесса KDD и их взаимосвязь.					
				1.2 Анализирует и классифицирует примеры применения Data Mining в различных областях.					
	2. Умеет собирать, очищать, преобразовывать и визуализировать данные для анализа.			2.1 Выполняет предобработку данных: очистку, обработку пропусков, нормализацию, кодирование признаков.					
				2.2 Строит графики и визуализации, демонстрирующие структуру и распределение данных.					
	3. Применяет методы классификации, регрессии, кластеризации и ассоциативного анализа к реальным задачам.			3.1 Реализует алгоритмы классификации и регрессии на реальных датасетах.					
				3.2 Выполняет кластеризацию и поиск ассоциативных правил, интерпретируя полученные результаты.					
	4. Оценивает качество моделей и интерпретирует результаты анализа данных.			4.1 Сравнивает различные модели и обосновывает выбор оптимальной.					
				4.2 Сравнивает различные модели и обосновывает выбор оптимальной.					
	5. Разрабатывает аналитические проекты и применяет Data Mining для решения прикладных задач.			5.1 Реализует полный цикл анализа данных (от предобработки до визуализации) в рамках мини-проекта.					
				5.2 Представляет результаты анализа в виде отчета и презентации, демонстрируя умение интерпретировать выводы.					
Пререквизиты	Основы программирования (Python / Java / C++). Алгоритмы и структуры данных. Базы данных и системы управления базами данных (СУБД).								

требования к ресурсам	Машинное обучение и глубокое обучение (Machine Learning & Deep Learning). Аналитика больших данных (Big Data Analytics).
Литература:	основная, дополнительная.
	1. Géron A. Прикладное машинное обучение с использованием Scikit-Learn, Keras и TensorFlow. – СПб.: Питер, 2022. – 864 с. 2. Aggarwal C.C. Data Mining: The Textbook. – Springer, 2021. – 734 p. 3. Хан Дж., Камбер М., Пей Д. Интеллектуальный анализ данных: современные методы. – М.: Вильямс, 2021. – 752 с. 4. Тан П., Штайнбах М., Кумар В. Введение в интеллектуальный анализ данных. – М.: Вильямс, 2020. – 600 с. 5. Уиттен И., Франк Э., Холл М. Data Mining: Практическое руководство по анализу данных. – М.: Диалектика, 2022. – 640 с.
	Исследовательская инфраструктура
	1. Язык программирования Python (версии 3.8+). 2. Библиотеки анализа данных: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn. 3. Среды разработки: Jupyter Notebook, Google Colab, VS Code. 4. Системы управления базами данных: MySQL, PostgreSQL. 5. Kaggle – платформа для работы с датасетами и проведения соревнований. 6. UCI Machine Learning Repository – репозиторий открытых наборов данных.
	Профессиональные научные базы данных
	1. IEEE Xplore Сайт: IEEE Xplore 2. ACM Digital Library сайт: ACM Digital Library . 3. SpringerLink Сайт: SpringerLink 4. ScienceDirect Сайт: Science Direct 5. Web of Science Сайт: Web of Science
	Интернет-ресурсы.
	1. https://archive.ics.uci.edu 2. https://scikit-learn.org 3. https://pandas.pydata.org 4. https://www.kaggle.com
	Программное обеспечение.
	1. Python 2. Jupyter Notebook 3. Google Colab

Академическая политика дисциплины	Академическая политика дисциплины определяется <u>Академической политикой и Политикой академической честности КазНУ имени аль-Фараби</u>. Документы доступны на главной странице ИС Univer. Интеграция науки и образования. Научно-исследовательская работа студентов, магистрантов и докторантов – это углубление учебного процесса. Она организуется непосредственно на кафедрах, в лабораториях, научных и проектных подразделениях университета, в студенческих научно-технических объединениях. Самостоятельная работа обучающихся на всех уровнях образования направлена на развитие исследовательских навыков и компетенций на основе получения нового знания с применением современных научно-исследовательских и информационных технологий. Преподаватель исследовательского университета интегрирует результаты научной деятельности в тематику лекций и семинарских (практических) занятий, лабораторных занятий и в задания СРОП, СРО, которые отражаются в силлабусе и отвечают за актуальность тематик учебных занятий и заданий. Посещаемость. Дедлайн каждого задания указан в календаре (графике) реализации содержания дисциплины. Несоблюдение дедлайнов приводит к потере баллов. Академическая честность. Практические/лабораторные занятия, СРО развивают у обучающегося самостоятельность, критическое мышление, креативность. Недопустимы плагиат, подлог, использование шпаргалок, списывание на всех этапах выполнения заданий. Соблюдение академической честности в период теоретического обучения и на экзаменах помимо основных политик регламентируют «Правила проведения итогового контроля», «Инструкции для проведения итогового контроля осеннего/весеннего семестра текущего учебного года», «Положение о проверке текстовых документов обучающихся на наличие заимствований». Документы доступны на главной странице ИС Univer. Основные принципы инклюзивного образования. Образовательная среда университета задумана как безопасное место, где всегда присутствуют поддержка и равное отношение со стороны преподавателя ко всем обучающимся и обучающимся друг к другу независимо от гендерной, расовой/ этнической принадлежности, религиозных убеждений, социально-экономического статуса, физического здоровья студента и др. Все люди нуждаются в поддержке и дружбе ровесников и
-----------------------------------	---

сокурсников. Для всех студентов достижение прогресса скорее в том, что они могут делать, чем в том, что не могут. Разнообразие усиливает все стороны жизни.

Все обучающиеся, особенно с ограниченными возможностями, могут получать консультативную помощь по телефону/ e-mail либо +7 7026478122 посредством видеосвязи в MS Teams <https://us04web.teams.us/j/73674837609?pwd=AlD37HhIShnBYtJOgU7qqmDQ9oK4na.1>

Интеграция MOOC (massive open online course). В случае интеграции MOOC в дисциплину, всем обучающимся необходимо зарегистрироваться на MOOC. Сроки прохождения модулей MOOC должны неукоснительно соблюдаться в соответствии с графиком изучения дисциплины.

ВНИМАНИЕ! Дедлайн каждого задания указан в календаре (графике) реализации содержания дисциплины, а также в MOOC. Несоблюдение дедлайнов приводит к потере баллов.

ИНФОРМАЦИЯ О ПРЕПОДАВАНИИ, ОБУЧЕНИИ И ОЦЕНИВАНИИ

Календарь (график) реализации содержания дисциплины. Методы преподавания и обучения.

Неделя	Название темы	Кол-во часов	Макс. балл
МОДУЛЬ 1: Основы анализа данных			
1	Л 1. Введение в интеллектуальный анализ данных, история и современное применение	1	
	Семинар 1. Обзор кейсов Data Science, примеры из бизнеса и науки	1	
	ЛЗ 1. Установка Python, Jupyter; знакомство с Pandas, анализ простого датасета	1	
2	Л 2. Этапы анализа данных: CRISP-DM, KDD	1	
	СЗ 2. Разбор этапов обработки данных на примере датасета	1	
	ЛЗ 2. Очистка данных: пропущенные значения, дубликаты, категориальные признаки	1	
3	Л 3. Предварительная обработка данных: нормализация, кодирование категорий	1	

4	СЗ 3. Практическая работа с преобразованием признаков и масштабированием	1	
	ЛЗ 3. Подготовка датасета для машинного обучения	1	
	Л 4. Визуализация данных и EDA	1	
	СЗ 4. Построение графиков, выявление закономерностей	1	
	ЛЗ 4. Визуальный анализ: гистограммы, корреляции, boxplot	1	
5	СРОП 1. Консультации по выполнению СРО 1 на тему «Прогнозирование потребительского поведения»	1	
	Л 5. Основы Python для анализа данных (NumPy, Pandas)	1	0
	СЗ 5. Работа с библиотеками: агрегирование, фильтрация, группировка	1	3
	ЛЗ 5. Анализ датасета: статистики, сводные таблицы	1	7
	СРО 1. Прием СРС 1 в форме оцениваться по отчёту, демонстрации проекта.	1	30
МОДУЛЬ 2: Методы машинного обучения			
6	Л 6. Классификация: логистическая регрессия, kNN, деревья решений	1	0
	СЗ 6. Применение методов классификации на реальном датасете	1	3
	ЛЗ 6. Построение модели классификации, оценка точности, precision, recall	1	7
	СРОП 2. Консультации по выполнению СРО 2 на тему «Анализ и классификация текстов (NLP)»		0
7	Л 7. Регрессия: линейная, полиномиальная, оценка качества моделей	1	0
	СЗ 7. Практическая работа с регрессионными задачами	1	3
	ЛЗ 7. Построение модели прогнозирования числовых значений (например, цены на жильё)	1	7
	СРО 2. Прием СРС 2 в форме оцениваться по отчёту, демонстрации проекта.		30
8	Л 8. Кластеризация: k-means, иерархическая кластеризация, DBSCAN	1	0
	СЗ 8. Анализ кластеров, визуализация результатов	1	3
	ЛЗ 8. Кластеризация пользователей/объектов по признакам	1	7
	Рубежный контроль 1		100
9	Л 9. Снижение размерности: PCA, t-SNE	1	0
	СЗ 9. Визуализация многомерных данных после снижения размерности	1	3
	ЛЗ 9. Применение PCA/t-SNE для визуализации сложного датасета	1	7
	СРОП 3. Консультация по выполнению СРО 3 на тему «Прогнозирование цен на недвижимость»		0
10	Л 10. Введение в Scikit-learn: построение и оценка моделей	1	0
	СЗ 10. Работа с пайплайнами и cross-validation	1	3
	ЛЗ 10. Создание пайплайна: обработка данных, обучение модели, оценка качества	1	7
	СРОП 3. Консультация по выполнению СРО 3 на тему «Прогнозирование цен на недвижимость»		0
МОДУЛЬ 3 Продвинутое методы и проекты			
11	Л 11. Ансамблевые методы: Random Forest, Gradient Boosting, XGBoost	1	0
	СЗ 11. Разбор кейсов с ансамблевыми методами, сравнение моделей	1	3
	ЛЗ 11. Построение ансамблевой модели, анализ важности признаков	1	7
	СРО 3. Прием СРС 3 в форме оцениваться по отчёту, демонстрации проекта.		15
12	Л12. Работа с текстовыми данными: NLP, TF-IDF, Bag-of-Words, классификация	1	0
	СЗ 12. Классификация текстов: токенизация, очистка, обучение моделей	1	3
	ЛЗ 12. Классификация текстов (отзывы, новости) с TF-IDF	1	7
	СРОП 4. Консультация по выполнению СРО 4 на тему «Кластеризация и сегментация пользователей»		0
13	Л 13. Введение в нейронные сети и глубокое обучение (TensorFlow/PyTorch)	1	0
	СЗ 13. Построение простой нейронной сети для классификации изображений/текста	1	3
	ЛЗ 13. Реализация нейронной сети и оценка метрик качества	1	7
	СРОП 4. Консультация по выполнению СРО 4 на тему «Кластеризация и сегментация пользователей»		0
14	Л 14. Оценка качества моделей: метрики, cross-validation, регуляризация	1	0
	СЗ 14. Разбор ошибок моделей, подбор гиперпараметров	1	3
	ЛЗ 14. Cross-validation, регуляризация, ROC/AUC анализ	1	7
	СРОП 4. Консультация по выполнению СРО 4 на тему «Кластеризация и сегментация пользователей»		0
15	Л 15. Этика и безопасность в интеллектуальном анализе данных, подготовка проекта	1	0
	СЗ 15. Презентация мини-проектов студентов, обсуждение результатов	1	3
	ЛЗ 15. Мини		

РУБРИКАТОР СУММАТИВНОГО ОЦЕНИВАНИЯ КРИТЕРИИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

СРС 1. Прогнозирование потребительского поведения (30% от 100% РК1)

Критерий	«Отлично» 25-30 %	«Хорошо» 20-24%	«Удовлетворительно» 15-19%	«Неудовлетворительно» 9-14%
Подготовка и анализ данных	Данные тщательно очищены, проведён анализ признаков и визуализация.	Подготовка данных в целом выполнена, но анализ частично неполный.	Подготовка данных частичная, есть ошибки.	Подготовка данных не выполнена.
Построение и обучение модели	Использованы современные методы (ансамблевые, деревья, логистическая регрессия и др.), гиперпараметры оптимизированы.	Модель построена и обучена, но оптимизация частичная.	Модель обучена базово, без настройки.	Модель отсутствует или не функционирует.
Интерпретация и визуализация результатов	Результаты интерпретированы, визуализированы, сделаны выводы о потребительском поведении.	Интерпретация присутствует, но поверхностна.	Интерпретация слабая или неполная.	Интерпретация отсутствует.

СРС 2. Анализ и классификация текстов (NLP) (30% от 100% РК1)

Критерий	«Отлично» 25-30 %	«Хорошо» 20-24%	«Удовлетворительно» 15-19%	«Неудовлетворительно» 9-14%
1. Предобработка и векторизация текста	Текст очищен, нормализован, использованы подходящие методы векторизации (TF-IDF, Word2Vec и др.).	Предобработка выполнена, но с мелкими недочётами.	Предобработка частично реализована.	Предобработка отсутствует.
2. Построение и обучение модели	Применены эффективные алгоритмы, модель обучена и оценена корректно.	Модель построена, но метрики можно улучшить.	Базовая модель без оптимизации.	Модель отсутствует или не работает.
3. Оценка качества и интерпретация	Метрики классификации рассчитаны, результаты интерпретированы и визуализированы.	Метрики рассчитаны, но интерпретация частичная.	Метрики частичные или с ошибками.	Оценка отсутствует.

СРС 3. Прогнозирование цен на недвижимость (15% от 100% РК2)

Критерий	«Отлично» 10-15 %	«Хорошо» 8-10%	«Удовлетворительно» 5-8%	«Неудовлетворительно» 2-5%
Подготовка и анализ данных	Проведена качественная очистка и анализ данных, выявлены значимые признаки.	Подготовка данных выполнена, но без глубокой аналитики.	Подготовка частичная, есть ошибки.	Подготовка данных отсутствует.
Построение модели и оценка	Применены современные регрессионные методы, оптимизированы гиперпараметры, рассчитаны метрики.	Модель обучена и метрики рассчитаны, но оптимизация слабая.	Базовая модель без настройки.	Модель отсутствует или некорректна.
Интерпретация и визуализация результатов	Результаты подробно интерпретированы и визуализированы.	Интерпретация присутствует, но не полная.	Интерпретация частичная.	Интерпретация отсутствует.

СРС 4. Кластеризация и сегментация пользователей (15% от 100% РК2)

Критерий	«Отлично» 10-15 %	«Хорошо» 8-10%	«Удовлетворительно» 5-8%	«Неудовлетворительно» 2-5%
1. Подготовка данных и выбор признаков	Данные тщательно подготовлены, признаки выбраны и обоснованы.	Подготовка данных выполнена, но выбор признаков можно улучшить.	Подготовка частичная, признаки не оптимальны.	Подготовка данных отсутствует.
2. Реализация кластеризации и анализ результатов	Кластеризация выполнена корректно (K-means, DBSCAN и др.), кластеры интерпретированы.	Кластеризация проведена, но интерпретация частично поверхностна.	Кластеризация реализована базово или с ошибками.	Кластеризация не выполнена.
3. Интерпретация и практические рекомендации	Даны четкие стратегии для каждой группы, выводы логичны и аргументированы.	Стратегии присутствуют, но не полностью проработаны.	Стратегии частичные или слабые.	Стратегии отсутствуют.

Декан

Т.С. Иманкулов

Председатель Академического комитета

по качеству преподавания и обучения Ж.А. Бურიбаев

Заведующий кафедрой

А.Н.Шормакова

Лектор

М.Б. Байкувеков