

СИЛЛАБУС
Весенний семестр 2023-2024 учебного года
Образовательная программа «7М06302 – Аудит информационной безопасности»

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)		
98861 Безопасность больших данных	4	1,7	0	3,3	5	9
АКАДЕМИЧЕСКАЯ ИНФОРМАЦИЯ О ДИСЦИПЛИНЕ						
Формат обучения	Цикл, компонент	Типы лекций	Типы практических занятий	Форма и платформа итогового контроля		
Офлайн	БД, КВ	Проблемно-ориентированный	Изучение концепций обработки естественных языков с помощью моделей машинного обучения	Устный офлайн		
Лектор - (ы)	Карюкин Владислав Игоревич					
e-mail:	vladislav.karyukin@gmail.com vladislav.karyukin@kaznu.kz					
Телефон:	+77019405992					
Ассистент- (ы)	Карюкин Владислав Игоревич					
e-mail:	vladislav.karyukin@gmail.com vladislav.karyukin@kaznu.kz					
Телефон:	+77019405992					
АКАДЕМИЧЕСКАЯ ПРЕЗЕНТАЦИЯ ДИСЦИПЛИНЫ						
Цель дисциплины	Ожидаемые результаты обучения (РО)*		Индикаторы достижения РО (ИД)			
Анализ понятия больших данных и их безопасности, использование методов анализа для решения в области безопасности больших данных, обнаружения сетевых аномалий.	1. (когнитивный) Теоретические понятия больших данных		1.1 Понимает базовые и расширенные понятия больших данных			
			1.2 Анализирует особенности методов обработки больших данных			
			1.3 Применяет методы разработки приложений, использующих большие данные			
	2. (функциональный) Работа с библиотеками обработки больших данных		2.1 Использует знания установки библиотек для работы с большими данными			
			2.2 Применяет данные библиотеки для работы с большими данными			
			2.3 Формирует навыки работы с библиотеками при создании приложений			
	3.(функциональный) Разрабатывать программы, анализирующие и обрабатывающие большие данные		3.1 Разрабатывает методы формирования баз данных			
			3.2 Создает полнофункциональную программу обработки больших данных			
			3.3 Разрабатывает методы анализа и визуализации больших данных			
	4. (системный) Создавать модули защиты данных		4.1 Создает конфигурацию обеспечения безопасности больших данных			
			4.2 Проводит анализ уязвимостей больших данных			
			4.3 Применяет методы машинного обучения для обеспечения безопасности больших данных			

	5. (системный) Создавать веб-приложения, использующие методы защиты больших данных	5.1 Создает веб-приложение, использующее большие данные 5.2 Конфигурирует модули обеспечения безопасности веб-приложения 5.3 Проводит оценку анализа уровня защищенности веб-приложения
Пререквизиты	Аудит информационной безопасности, Методы семантического анализа для обеспечения информационной безопасности	
Постреквизиты	Безопасность веб-приложений	
Учебные ресурсы	<p>Литература: Основная:</p> <ul style="list-style-type: none"> – Natural Language Processing with Python and spaCy: A Practical Introduction, Yuli Vasiliev, 2021. – Machine Learning and Deep Learning in Natural Language Processing, Anitha S. Pillai, Roberto Tedesco, 2023. – Natural Language Processing: A Machine Learning Perspective Yue Zhang, Zhiyang Teng, 2021. – Natural Language Processing Projects: Build Next-Generation NLP Applications Using AI Techniques, Akshay Kulkarni, Adarsha Shivananda, Anoosh Kulkarni, 2021. – Security and Privacy for Big Data, Cloud Computing and Applications. Wei Ren, Lizhe Wang, Kim-Kwang Raymond Choo, Fatos Xhafa, 2019. – Big Data Security. Shibakali Gupta, Indradip Banerjee and Siddhartha Bhattacharyya, 2019 <p>Дополнительная:</p> <ul style="list-style-type: none"> – Learning Scientific Programming with Python, Christian Hill, 2021 – Deep Learning for Natural Language Processing: Creating Neural Networks with Python. Palash Goyal, Sumit Pandey, Karan Jain, 2018 <p>Профессиональные научные базы данных:</p> <ul style="list-style-type: none"> – Лабораторная аудитория 514 – Лабораторная аудитория 323 <p>Интернет–ресурсы:</p> <ul style="list-style-type: none"> – Python Exercises, Practice, Solution – https://www.w3resource.com/python-exercises/ – Natural Language Toolkit – https://www.nltk.org/ – Tensorflow – https://www.tensorflow.org/?hl=ru – Machine learning mastery – https://machinelearningmastery.com/start-here/ <p>Программное обеспечение: Python IDE, Anaconda Navigator Python, NLTK, Microsoft Office Word, WinRAR, Power Point, Adobe Reader, Paint.</p>	
Академическая политика дисциплины	<p>Академическая политика дисциплины определяется <u>Академической политикой и Политикой академической честности КазНУ имени аль-Фараби</u>. Документы доступны на главной странице ИС Univer.</p> <p>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 syllabusе и отвечают за актуальность тематик учебных занятий и заданий.</p> <p>Посещаемость. Дедлайн каждого задания указан в календаре (графике) реализации содержания дисциплины. Несоблюдение дедлайнов приводит к потере баллов.</p> <p>Академическая честность. Практические/лабораторные занятия, СРО развивают у обучающегося самостоятельность, критическое мышление, креативность. Недопустимы плагиат, подлог, использование шпаргалок, списывание на всех этапах выполнения заданий.</p> <p>Соблюдение академической честности в период теоретического обучения и на экзаменах помимо основных политик регламентируют <u>«Правила проведения итогового контроля»</u>, <u>«Инструкции для проведения итогового контроля осеннего/весеннего семестра текущего учебного года»</u>,</p>	

«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 обучающихся 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 vladislav.karyukin@gmail.com / +77019405992 либо посредством видеосвязи в MS Teams

ИНФОРМАЦИЯ О ПРЕПОДАВАНИИ, ОБУЧЕНИИ И ОЦЕНИВАНИИ

Балльно-рейтинговая буквенная система оценки учета учебных достижений				Методы оценивания		
Оценка	Цифровой эквивалент баллов	Баллы, % содержание	Оценка по традиционной системе			
A	4,0	95-100	Отлично	Критериальное оценивание – процесс соотнесения реально достигнутых результатов обучения с ожидаемыми результатами обучения на основе четко выработанных критериев. Основано на формативном и суммативном оценивании. Формативное оценивание – вид оценивания, который проводится в ходе повседневной учебной деятельности. Является текущим показателем успеваемости. Обеспечивает оперативную взаимосвязь между обучающимся и преподавателем. Позволяет определить возможности обучающегося, выявить трудности, помочь в достижении наилучших результатов, своевременно корректировать преподавателю образовательный процесс. Оценивается выполнение заданий, активность работы в аудитории во время лекций, семинаров, практических занятий (дискуссии, викторины, дебаты, круглые столы, лабораторные работы и т. д.). Оцениваются приобретенные знания и компетенции. Суммативное оценивание – вид оценивания, который проводится по завершению изучения раздела в соответствии с программой дисциплины. Проводится 3-4 раза за семестр при выполнении СРО. Это оценивание освоения ожидаемых результатов обучения в соответствии с дескрипторами. Позволяет определять и фиксировать уровень освоения дисциплины за определенный период. Оцениваются результаты обучения.		
A-	3,67	90-94				
B+	3,33	85-89				Хорошо
B	3,0	80-84	Активность на лекциях	0		
B-	2,67	75-79	Работа на практических занятиях	25		
C+	2,33	70-74	Самостоятельная работа	25		
C	2,0	65-69	Удовлетворительно	Проектная и творческая деятельность	10	
C-	1,67	60-64		Итоговый контроль (экзамен)	40	
D+	1,33	55-59		Неудовлетворительно	ИТОГО	
D	1,0	50-54				
FX	0,5	25-49				
F	0	0-24				

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

Неделя	Название темы	Кол-во часов	Макс. балл
МОДУЛЬ 1 Основы операции работы с текстовыми данными			
1	Л 1. Введение в область больших данных	1	
	ЛЗ 1. Основные операции работы с большими данными	2	5
2	Л 2. Технологии хранения и обработки больших данных	1	0
	ЛЗ 2. Создание базы данных для хранения информации	2	5
	СРОП 1. Консультации по выполнению СРО1 на тему «Реализация проекта анализа и обработки больших данных»		
3	Л 3. Выполнение операций обработки текстовых данных	1	
	ЛЗ 3. Разработка программы обработки текстовых данных	2	7
	СРОП 2. Прием СРО 1		20
4	Л 4. Выполнение операции стемминга и лемматизации текстовых данных	1	
	ЛЗ 4. Создание программы стемминга и лемматизации	2	7
	СРОП 3. Проведение коллоквиума по темам 1-3 недель		5
5	Л 5. Выполнение операции моделирования и прогнозирования событий	1	
	ЛЗ 5. Создание программы прогнозирования событий	2	7
	СРОП 4. Консультация по выполнению СРО 2 на тему «Анализ методов защиты больших данных»		

МОДУЛЬ 2 Обработка больших данных моделями машинного обучения			
6	Л 6. Подготовка данных интернет-угроз для классификации моделями машинного обучения	1	
	ЛЗ 6. Создание программы обработки датасетов таких интернет-угроз, как DDoS, Man in the middle, SQL injection, Phishing, Malware	2	7
7	Л 7. Классификация интернет-угроз моделями машинного обучения	1	
	ЛЗ 7. Создание программы классификации интернет-угроз моделями Наивного Байеса, Логистической регрессии, Древа решений, Случайного леса и т.д.	2	12
	СРОП 5. Прием СРО 2		25
Рубежный контроль 1			100
8	Л 8. Классификация интернет-угроз нейронными сетями	1	
	ЛЗ 8. Создание программы классификации интернет-угроз моделями Deep neural network, Convolutional neural network и Long short term memory neural network	2	5
	СРОП 6. Консультация по выполнению СРО 3 на тему «Разработка программы классификации интернет-угроз с помощью BERT»		
9	Л 9. Анализ и обработка данных с помощью запросов ChatGPT	1	
	ЛЗ 9. Создание программы обработки данных с API ChatGPT	2	10
10	Л 10. Обеспечение методов защиты больших данных распределенных вычислений	1	
	ЛЗ 10. Создание системы распределенных вычислений	2	
	СРОП 7. Прием СРО 3		25
МОДУЛЬ 3 Работа с парсингом текстовых данных			
11	Л 11. Платформа бизнес-аналитики Power BI	1	
	ЛЗ 11. Установка и настройка платформы Power BI	2	5
	СРОП 8. Консультация по выполнению СРО 4 на тему «Создание приложения, использующего методы защиты больших данных»		
12	Л12. Визуализация отчетов больших данных в Power BI	1	
	ЛЗ 12. Создание подключения к различным источникам больших данных в Power BI	2	5
13	Л 13. Реализация облачных вычислений	1	
	ЛЗ 13. Настройка безопасности облачных хранилищ данных	2	5
	СРОП 9. Прием СРО 4		25
14	Л 14. Продвинутое системы анализа больших данных	1	
	ЛЗ 14. Конфигурация защиты систем анализа больших данных	2	10
15	Л 15. Основные этапы создания веб-приложения с моделями машинного обучения	1	
	ЛЗ 15. Разработка веб-приложения с использованием защиты данных методами машинного обучения	2	10
Рубежный контроль 2			100
Итоговый контроль (экзамен)			100
ИТОГО за дисциплину			100

РУБРИКАТОР СУММАТИВНОГО ОЦЕНИВАНИЯ
КРИТЕРИИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

СРО 1. Реализация проекта анализа и обработки больших данных (20% от 100% РК1)

Критерий	«Отлично» 21-25%	«Хорошо» 11-20%	«Удовлетворительно» 6-10%	«Неудовлетворительно» 0-5%
Знание и понимание основных концепций анализа и обработки больших данных	Понимание степени соответствия, актуальности и достоверности найденных данных. Знание и понимание всех основных элементов и операций анализа и обработки больших данных	Понимание степени соответствия, актуальности и достоверности найденных данных. Знание больше части операций анализа и обработки больших данных	Ограниченное понимание степени соответствия, актуальности и достоверности элементов и операций анализа и обработки больших данных	Поверхностное понимание/ отсутствие понимания степени соответствия, актуальности и достоверности найденных данных. Отсутствие знания элементов и операций анализа и обработки больших данных
Навыки написания программного кода анализа и обработки больших данных	Четкое и ясное представление программного кода, отсутствие в коде синтаксических ошибок	В программном коде имеются небольшие логические ошибки	Большое количество логических и синтаксических ошибок в программном коде, что делают его практически неработоспособным	Отсутствие программного кода или наличие нескольких строк кода
Написание отчета	Письмо демонстрирует ясность, лаконичность и правильность.	Письмо демонстрирует ясность, лаконичность и корректность. В основном отсутствуют ошибки.	В письме есть некоторые ключевые ошибки, и ясность нуждается в улучшении.	Написанное неясно, трудно следовать за содержанием. Много ошибок в тексте

СРО2. Анализ методов защиты больших данных (25% от 100% РК1)

Критерий	«Отлично» 21-25%	«Хорошо» 11-20%	«Удовлетворительно» 6-10%	«Неудовлетворительно» 0-5%
Работа с методами защиты больших данных	Понимание степени соответствия, актуальности и достоверности работы с данными в приложении. Знание и понимание всех основных методов защиты больших данных	Понимание степени соответствия, актуальности и достоверности найденных данных. Знание больше части методов защиты больших данных	Ограниченное понимание степени соответствия, актуальности и достоверности операций по созданию методов защиты больших данных	Поверхностное понимание/ отсутствие понимания степени соответствия, актуальности и достоверности работы с базами данных. Отсутствие знания операций создания методов защиты больших данных
Навыки написания программного кода	Четкое и ясное представление программного кода, отсутствие в коде синтаксических ошибок	В программном коде имеются небольшие логические ошибки	Большое количество логических и синтаксических ошибок в программном коде, что делают его практически неработоспособным	Отсутствие программного кода или наличие нескольких строк кода
Написание отчета	Письмо демонстрирует ясность, лаконичность и правильность.	Письмо демонстрирует ясность, лаконичность и корректность. В основном отсутствуют ошибки.	В письме есть некоторые ключевые ошибки, и ясность нуждается в улучшении.	Написанное неясно, трудно следовать за содержанием. Много ошибок в тексте

СРО 3. Разработка программы классификации интернет-угроз с помощью BERT (25% от 100% РК2)

Критерий	«Отлично» 21-25%	«Хорошо» 11-20%	«Удовлетворительно» 6-10%	«Неудовлетворительно» 0-5%
Работа с моделями машинного обучения классификации интернет-угроз с помощью BERT	Понимание степени соответствия, актуальности и достоверности работы с данными в приложении. Знание и понимание всех основных операций классификации интернет-угроз с помощью BERT	Понимание степени соответствия, актуальности и достоверности найденных данных. Знание больше части операций классификации интернет-угроз с помощью BERT	Ограниченное понимание степени соответствия, актуальности и достоверности операций классификации интернет-угроз с помощью BERT	Поверхностное понимание/ отсутствие понимания степени соответствия, актуальности и достоверности работы с базами данных. Отсутствие знания операций классификации интернет-угроз с помощью BERT
Навыки написания программного кода	Четкое и ясное представление программного кода, отсутствие в коде синтаксических ошибок	В программном коде имеются небольшие логические ошибки	Большое количество логических и синтаксических ошибок в программном коде, что делают его практически неработоспособным	Отсутствие программного кода или наличие нескольких строк кода
Написание отчета	Письмо демонстрирует ясность, лаконичность и правильность.	Письмо демонстрирует ясность, лаконичность и корректность. В основном отсутствуют ошибки.	В письме есть некоторые ключевые ошибки, и ясность нуждается в улучшении.	Написанное неясно, трудно следовать за содержанием. Много ошибок в тексте

СРО 4. Создание приложения, использующего методы защиты больших данных (25% от 100% РК2)

Критерий	«Отлично» 21-25%	«Хорошо» 11-20%	«Удовлетворительно» 6-10%	«Неудовлетворительно» 0-5%
Знание и понимание создания приложения, использующего методы защиты больших данных	Понимание степени соответствия, актуальности и достоверности создания приложения, использующего методы защиты больших данных	Понимание степени соответствия, актуальности и достоверности создания приложения, использующего методы защиты больших данных	Ограниченное понимание создания приложения, использующего методы защиты больших данных	Поверхностное понимание/ отсутствие понимания основных операций создания приложения, использующего методы защиты больших данных
Навыки написания программного кода	Четкое и ясное представление программного кода, отсутствие в коде синтаксических ошибок	В программном коде имеются небольшие логические ошибки	Большое количество логических и синтаксических ошибок в программном коде, что делают его практически неработоспособным	Отсутствие программного кода или наличие нескольких строк кода
Написание отчета	Письмо демонстрирует ясность, лаконичность и правильность.	Письмо демонстрирует ясность, лаконичность и корректность. В основном отсутствуют ошибки.	В письме есть некоторые ключевые ошибки, и ясность нуждается в улучшении.	Написанное неясно, трудно следовать за содержанием. Много ошибок в тексте

И.о. декана _____ Тұрар О.Н.

Заведующий кафедрой _____ Мусиралиева Ш.Ж.

Лектор _____ Карюкин В.И.