

Министерство науки и высшего образования РФ
Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Б1.О.32.10 БАЗОВЫЕ ДИСЦИПЛИНЫ
СПЕЦИАЛИЗАЦИИ

Интеллектуальный анализ и технологии Big Data

наименование дисциплины (модуля) в соответствии с учебным планом

Направление подготовки / специальность

38.05.01 Экономическая безопасность

Направленность (профиль)

38.05.01.31 Экономическая безопасность и финансовый мониторинг
национальных, региональных и корпоративных социально-
экономических систем

Форма обучения

заочная

Год набора

2022

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Программу составили _____

канд.техн.наук, Доцент, Капустина С.В.

должность, инициалы, фамилия

1 Цели и задачи изучения дисциплины

1.1 Цель преподавания дисциплины

Ознакомление студентов с процессами, алгоритмами и инструментами, относящимися к основным принципам машинного обучения, технологиям Big Data.

Формирование устойчивых навыков использования программного обеспечения в области обработки данных и бизнес-аналитики, необходимых для принятия управленческих решений.

1.2 Задачи изучения дисциплины

- приобретение знаний в области технологий обработки больших данных, принципов машинного обучения, классов моделей (линейные, логические, нейросетевые), метриками качества и подходами к предварительной обработке данных;

- развитие цифровых навыков в процессе интеллектуального анализа больших данных с использованием программных средств для решения практических задач.

1.3 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Код и наименование индикатора достижения компетенции	Запланированные результаты обучения по дисциплине
ОПК-6: Способен использовать современные информационные технологии и программные средства при решении профессиональных задач.	
ОПК-6.2: Использует современные информационные технологии для подготовки исходных данных; подготавливает исходные данные исходя из постановки задачи и взаимосвязи исходных данных и конечного результата; формирует необходимые исходные данные в соответствии с особенностями профессиональных задач и инструментальными средствами программных приложений, используемыми для их решения.	методы разведочного анализа данных (EDA) методы одномерного и двумерного анализ данных (диаграммы рассеивания, корреляционный анализ) методы поиска аномальных и пропущенных данных, кодирование признаков выполнять подготовку данных для анализа визуализировать данные владеть средствами глубокого изучения данных с использованием описательной статистики и визуальных средств языком программирования Python, Google Colab, GitHub, Azure ML Studio
ПК-2: Способен на основе обоснованного выбора методик и действующей нормативно-правовой базы проводить количественную и качественную оценку угроз экономической и финансовой безопасности государства, региона, корпоративного сектора на основе критериев допустимых уровней по целевым задачам для принятия управленческих решений и формирования соответствующих систем обеспечения	

<p>ПК-2.2: Применяет на практике методы, техники, технологии, программные средства и информационные базы идентификации и анализа угроз экономической и финансовой безопасности государства, региона, корпоративного сектора на основе критериев допустимых уровней по целевым задачам</p>	<p>Методы решения задач машинного обучения с учителем методы сокращения размерности количественных данных Применять инструментарий ППП Statistica, Loginom для разрабатывать программы на Python инструментами цифровых технологий для автоматизации информационно-аналитической деятельности</p>
<p>ПК-3: Способен исследовать социально-экономические процессы, анализировать, структурировать, классифицировать нормативно-правовую документацию, экономическую и финансовую информацию, в том числе о подозрительных операциях и сделках</p>	
<p>ПК-3.2: Разрабатывает методики проведения анализа нормативно-правовой документации, экономической и финансовой информации, в том числе в целях ПОД/ФТ</p>	<p>методики проведения анализа нормативно-правовой документации федеральные законы и приказы Росфинмониторинга о противодействии нелегальных сделок выполнять оценку риска совершения клиентом операций с использованием информационных технологий</p>
<p>ПК-3.5: Формулирует математическую постановку задач математического программирования на основе анализа социально-экономических процессов и находить решение задачи.</p>	<p>постановки задач математического программирования и их решение строить модели классификации подозрительных сделок на Python компьютерными технологиями анализа социально-экономических процессов</p>

1.4 Особенности реализации дисциплины

Язык реализации дисциплины: Русский.

Дисциплина (модуль) реализуется без применения ЭО и ДОТ.

2. Объем дисциплины (модуля)

Вид учебной работы	Всего, зачетных единиц (акад.час)	Семестр					
		1	2	3	4	5	6

3 Содержание дисциплины (модуля)

3.1 Разделы дисциплины и виды занятий (тематический план занятий)

№ п/п		Модули, темы (разделы) дисциплины		Контактная работа, ак. час.							
				Занятия лекционного типа		Занятия семинарского типа				Самостоятельная работа, ак. час.	
						Семинары и/или Практические занятия		Лабораторные работы и/или Практикумы			
				Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС
1. Большие данные. Современные подходы к обработке и хранению											
		1. Особенности Big Data, инструменты, технологии, этапы обработки данных, моделирование. Технологии обработки больших данных Hadoop, MapReduce, NoSQL		1							
		2. Предварительный анализ больших наборов данных (структурирование, исключение аномальных данных, нормализация, кодирование, визуализация)				2					
		3. Программирование на Python в среде разработки Jupiter Notebook, Google Colab							24		
2. Общая постановка задачи обучения по прецедентам											
		1. Обучение с учителем: регрессия, классификация, прогнозирование, ранжирование. Метрики качества.		1							
		2. Метрические и логические алгоритмы классификации				2					
		3. Нейронные сети для решения задачи классификации				2					

4. Изучение scikitlearn, tensorflow							24	
3. Обучение без учителя и с подкреплением.								
1. Иерархические и неиерархические алгоритмы кластеризации. Метрики качества кластеризации.	2							
2. Алгоритмы кластеризации. Глубокое обучение нейронных сетей.			2					
3. Определение подозрительной активности во время покупок							44	
Всего	4		8				92	

4 Учебно-методическое обеспечение дисциплины

4.1 Печатные и электронные издания:

1. Коэльо Л. П. Построение систем машинного обучения на языке Python (Москва: ДМК Пресс).
2. Рашка С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения(Москва: ДМК Пресс).
3. Бонцанини М. Анализ социальных медиа на Python. Извлекайте и анализируйте данные из всех уголков социальной паутины на Python: научное издание(Москва: ДМК Пресс).
4. Содем Я. Э. Программирование компьютерного зрения на языке Python (Москва: ДМК Пресс).

4.2 Лицензионное и свободно распространяемое программное обеспечение, в том числе отечественного производства (программное обеспечение, на которое университет имеет лицензию, а также свободно распространяемое программное обеспечение):

1. Python, Pandas, Numpy, Tensorflow
2. Google Colab

4.3 Интернет-ресурсы, включая профессиональные базы данных и информационные справочные системы:

1. UCI Machine Learning Repository
2. Kaggle
3. GitHub

5 Фонд оценочных средств

Оценочные средства находятся в приложении к рабочим программам дисциплин.

6 Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине (модулю)

Для осуществления образовательного процесса необходимо:

Интерактивная доска,
проектор,
компьютерный класс