

**Федеральное государственное образовательное бюджетное учреждение
высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)**

**Кафедра анализа данных и машинного обучения
Факультета информационных технологий и анализа больших данных**

УТВЕРЖДАЮ

Проректор по учебной и
методической работе

_____ Е.А. Каменева

24.05.2024 г.

В.И. Соловьев, В.Г. Феклин, Н.Ю. Чулина

Прикладные задачи машинного обучения

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки:
09.03.03 - Прикладная информатика,
ОП «Инженерия данных»; ОП «Цифровые платформы управления предприятиями»;
ОП «Прикладные информационные системы в экономике и финансах»

*Рекомендовано Ученым советом
Факультета информационных технологий и анализа больших данных
(протокол № 44 от 21.05.2024 г.)*

*Одобрено советом Кафедры анализа данных и машинного обучения
(протокол № 01 от 06.05.2024 г.)*

Москва 2024

СОДЕРЖАНИЕ

1. Наименование дисциплины.....	3
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине	3
3. Место дисциплины в структуре образовательных программ.....	4
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	4
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий.....	5
5.1. Содержание дисциплины.....	5
5.2. Учебно–тематический план.....	7
5.3. Содержание семинаров, практических занятий	9
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине.....	10
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы	10
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю	11
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине.....	13
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	18
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.....	19
10. Методические указания для обучающихся по освоению дисциплины.	20
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем	21
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.....	22

1. Наименование дисциплины

Дисциплина «Прикладные задачи машинного обучения».

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ОП «Инженерия данных», ОП «Прикладные информационные системы в экономике и финансах»			
ПКН-4	Способность проектировать и создавать интеллектуальные информационные системы, выбирать метод обучения в соответствии с анализом задачи	1. Демонстрирует знание основных понятий машинного обучения и интеллектуального анализа данных, понимание области и границ применимости, основные виды задач.	Знать: основные принципы постановки задач машинного обучения. Уметь: осуществлять постановку прикладных задач машинного обучения для создания ценности.
		2. Демонстрирует знание популярных инструментальных средств машинного обучения, собирает датасет, строит модели, проводит их анализ и диагностику, делает содержательные выводы.	Знать: компьютерную реализацию основных методов и алгоритмов машинного обучения в пакетах прикладных программ Уметь: выбирать компьютерную реализацию методов и алгоритмов машинного обучения для решения бизнес-задач
		3. Презентабельно демонстрирует результаты анализа данных и машинного обучения в форме, доступной непрофессионалу, структурирует отчет по проведенному анализу.	Знать: современные инструменты визуализации результатов анализа данных и машинного обучения Уметь: применять современные инструменты визуализации результатов анализа данных и машинного обучения в форме, доступной непрофессионалу
ОП «Инженерия данных»			
ПКП-1	Способность формировать требования, производить подготовку данных и проводить аналитические работы с использованием	1. Демонстрирует знания в области компьютерных технологий, используемых для анализа больших данных.	Знать: компьютерную реализацию основных методов и алгоритмов машинного обучения в пакетах прикладных программ Уметь: выбирать компьютерную реализацию методов и алгоритмов машинного обучения для выполнения аналитических работ на основе больших данных.

	технологий больших данных	2. Владеет навыками управления аналитическими работами на основе больших данных.	Знать: основные принципы постановки задач машинного обучения в аналитических работах на основе больших данных. Уметь: осуществлять постановку прикладных задач машинного обучения для выполнения аналитических работ на основе больших данных.
		3. Выполняет аналитические работы на основе больших данных.	Знать: алгоритмы машинного обучения и их компьютерную реализацию в пакетах прикладных программ Уметь: применять компьютерную реализацию методов и алгоритмов машинного обучения для выполнения аналитических работ на основе больших данных
ОП «Цифровые платформы управления предприятиями»			
ПКП-6	Способность разрабатывать, реализовывать и применять методы интеллектуального анализа данных и машинного обучения для автоматизации решения неструктурированных и слабоструктурированных задач экономических предметных областей	1. Использует знания современных методов интеллектуального анализа данных (в том числе, больших данных) и способы их программной реализации.	Знать: основные принципы постановки задач машинного обучения. Уметь: осуществлять постановку прикладных задач машинного обучения для создания ценности.
		2. Осуществляет поиск, сбор, анализ и интерпретацию данных экономических предметных областей с применением методов искусственного интеллекта и машинного обучения	Знать: компьютерную реализацию основных методов и алгоритмов машинного обучения в пакетах прикладных программ Уметь: выбирать компьютерную реализацию методов и алгоритмов машинного обучения для решения бизнес-задач
		3. Владеет современным инструментарием искусственного интеллекта и его использованием при разработке и развитии существующих финансово-экономических информационных систем.	Знать: алгоритмы машинного обучения и их компьютерную реализацию в пакетах прикладных программ Уметь: применять компьютерную реализацию методов и алгоритмов машинного обучения при разработке и развитии существующих финансово-экономических информационных систем

3. Место дисциплины в структуре образовательных программ

Дисциплина «Прикладные задачи машинного обучения» является дисциплиной Цикла профиля (элективный) по направлению подготовки 09.03.03 - Прикладная информатика, ОП «Инженерия данных», ОП «Цифровые платформы управления предприятиями», ОП «Прикладные информационные системы в экономике и финансах».

4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

очная форма обучения

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 7 (в часах)
Общая трудоёмкость дисциплины	3/108	108
Контактная работа - Аудиторные занятия	50	50
<i>Лекции</i>	<i>16</i>	<i>16</i>
<i>Семинары, практические занятия</i>	<i>34</i>	<i>34</i>
Самостоятельная работа	58	58
Вид текущего контроля		контрольная работа
Вид промежуточной аттестации		зачет

очно-заочная, очно-заочная (ИОО) формы обучения

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 8 (в часах)
Общая трудоёмкость дисциплины	3/108	108
Контактная работа - Аудиторные занятия	34	34
<i>Лекции</i>	<i>16</i>	<i>16</i>
<i>Семинары, практические занятия</i>	<i>18</i>	<i>18</i>
Самостоятельная работа	74	74
Вид текущего контроля		контрольная работа
Вид промежуточной аттестации		зачет

заочная форма обучения (ИОО)

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 8 (в часах)
Общая трудоёмкость дисциплины	3/108	108
<i>Контактная работа - Аудиторные занятия</i>	<i>12</i>	<i>12</i>
<i>Лекции</i>	<i>4</i>	<i>4</i>
<i>Семинары, практические занятия</i>	<i>8</i>	<i>8</i>
Самостоятельная работа	96	96
Вид текущего контроля		контрольная работа
Вид промежуточной аттестации		зачет

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. От данных к ценности: введение в науки о данных

Основные определения: интеллектуальный анализ данных, большие данные, машинное обучение.

Методы и задачи машинного обучения, области применения методов и технологий машинного обучения.

Примеры задач машинного обучения: поиск информации в интернете, распознавание изображений, лиц, эмоций, пола, возраста, ..., распознавание речи, языка, эмоциональной окраски текстов, прогнозирование продаж, прогнозирование оттока клиентов, кредитный скоринг, рекомендательные системы и др.

Цикл обработки данных: поиск данных, сбор данных, очистка данных, трансформация данных, интеллектуальный анализ данных, интерпретация и практическое применение результатов.

Статистические основы обработки данных.

Обзор современных технологий машинного обучения: Microsoft Azure ML, Python, R.

Тема 2. Практическое использование моделей классификации

Принципы разработки и оценки систем машинного обучения. Основные классы моделей машинного обучения: обучение с учителем (классификация и регрессия) и без учителя (кластеризация и поиск аномалий).

Задача классификация с обучением. Задача кредитного скоринга.

Модель логистической регрессии и ее компьютерная реализация.

Методы оценки качества моделей классификации: доля правильных ответов, точность, полнота, F1, AUC. Цены ошибок первого и второго рода.

Модели деревьев и лесов решений. Усиление (бустинг) деревьев решений. Нейронные сети. Компьютерная реализация методов классификации.

Недообучение и переобучение в моделях классификации.

Основы технологии улучшения моделей машинного обучения. Подбор оптимальных параметров моделей. Важность подготовки данных. Генерация синтетических признаков. Работа с пропущенными данными. Работа с несбалансированными выборками.

Тема 3. Практическое использование моделей регрессии

Задача прогнозирования продаж. Модель множественной линейной регрессии. Методы оценки качества моделей регрессии. Коэффициент детерминации, средняя абсолютная ошибка предсказания, средняя относительная ошибка предсказания.

Методы классификации и их компьютерная реализация.

Недообучение и переобучение в моделях регрессии.

Основы технологии улучшения моделей машинного обучения в задачах регрессии.

Тема 4. Практическое использование моделей кластерного анализа и поиска аномалий

Кластерный анализ и его компьютерная реализация. Задача сегментирования потребителей. Иерархические агломеративные алгоритмы. Метод К-средних.

Методы оценки качества моделей кластерного анализа. Расстояния между кластерами, расстояния между объектами внутри кластеров. Характеристики центров кластеров.

Задача снижения размерности факторного пространства. Метод главных компонент и его компьютерная реализация. Геометрическое и экономическое содержание метода главных компонент. Использование методов снижения размерности для улучшения качества моделей машинного обучения. Использование методов снижения размерности для выявления латентных факторов.

Поиск аномалий. Задача поиска мошеннических транзакций.

Тема 5. Построение рекомендательных систем

Понятие рекомендательной системы, типы оценок. Примеры рекомендательных сервисов.

Принципы создания рекомендательных систем: коллаборативная фильтрация, контентная фильтрация, гибридные подходы.

Алгоритмы, используемые рекомендательными системами: корреляция Пирсона, алгоритмы кластеризации, байесовские сети доверия, цепи Маркова, метод Роккио.

Проблемы рекомендательных систем.

Компьютерная реализация рекомендательных систем.

5.2. Учебно–тематический план

очная форма обучения

№ п/п	Наименование тем (разделов) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Всего	*Контактная работа - Аудиторная работа			Самос- тояте- льная работа	
			Общая, в т.ч.	Лекции	Семинары, практи- ческие занятия		
1.	От данных к ценности: введение в науки о данных	12	6	2	4	6	тест
2.	Практическое ис- пользование моделей классификации	24	12	4	8	12	соревнование
3.	Практическое ис- пользование моделей регрессии	24	12	4	8	12	соревнование

4.	Практическое использование моделей кластерного анализа и поиска аномалий	24	12	4	8	12	соревнование
5.	Построение рекомендательных систем	24	8	2	6	16	тест
	В целом по дисциплине	108	50	16	34	58	Согласно учебному плану: контрольная работа
	Итого в %		46	32	68	54	

Очно-заочная, очно-заочная (ИОО) формы обучения

№ п/п	Наименование тем (разделов) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Всего	*Контактная работа - Аудиторная работа			Самос- тоятельная работа	
			Общая, в т.ч.	Лекции	Семинары, практи- ческие занятия		
1.	От данных к ценности: введение в науки о данных	12	4	2	2	8	тест
2.	Практическое ис- пользование моделей классификации	24	10	4	6	14	соревнование
3.	Практическое ис- пользование моделей регрессии	24	8	4	4	16	соревнование
4.	Практическое использование моделей кластерного анализа и поиска аномалий	24	8	4	4	16	соревнование
5.	Построение рекомендательных систем	24	4	2	2	20	тест
	В целом по дисциплине	108	34	16	18	74	Согласно учебному плану: контрольная работа
	Итого в %		31	47	53	69	

заочная форма обучения (ИОО)

№ п/п	Наименование тем (разделов) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Всего	*Контактная работа - Аудиторная работа			Самос- тоятельная работа	
			Общая, в т.ч.	Лекции	Семинары, практи- ческие занятия		
1.	От данных к ценности: введение в науки о данных	12	4	4	-	8	тест
2.	Практическое ис- пользование моделей классификации	24	2	-	2	22	соревнование
3.	Практическое ис- пользование моделей регрессии	24	2	-	2	22	соревнование
4.	Практическое использование моделей кластерного анализа и поиска аномалий	24	2	-	2	22	соревнование
5.	Построение рекомендательных систем	24	2	-	2	22	тест
	В целом по дисциплине	108	12	4	8	96	Согласно учебному плану: контрольная работа
	Итого в %		11	33	67	89	

* объем контактной работы в очно-заочной/заочной формах обучения и индивидуальных учебных планах определяется соответствующими учебными планами. Темы, реализуемые в виде контактной работы, определяются преподавателем самостоятельно, исходя из уровня их сложности.

5.3. Содержание семинаров, практических занятий

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9 (указывается раздел и порядковый номер источника)	Формы проведения занятий
От данных к ценности: Введение в науки о данных	Построение компьютерной системы визуализации данных, анализ конкретной ситуации и обсуждение результатов. Построение компьютерной системы визуализации данных, анализ конкретной	Интерактив: построение компьютерной системы машинного обучения, анализ конкретной ситуации и обсуждение

	ситуации и обсуждение результатов – 100% от трудоемкости семинарского занятия. <i>р.8, [1-3]; р.9, [4]-[8]</i>	результатов – 100% от трудоемкости семинарского занятия
Практическое использование моделей классификации	Построение системы прогнозирования продаж. Построение системы кредитного скоринга. Оптимизация системы кредитного скоринга. <i>р.8, [1-3]; р.9, [4]-[9]</i>	Лабораторная работа «Построение системы кредитного скоринга». (построение компьютерной системы машинного обучения, анализ конкретной ситуации и обсуждение результатов).
Практическое использование моделей регрессии	Кластерный анализ и его компьютерная реализация. Иерархические агломеративные алгоритмы. Метод К-средних. Методы оценки качества моделей кластерного анализа. Расстояния между кластерами, расстояния между объектами. Поиск аномалий. Задача поиска мошеннических транзакций. Сегментирование потребителей. <i>р.8, [1]; р.9, [4]-[9]</i>	Лабораторная работа «Построение системы прогнозирования продаж» (построение компьютерной системы машинного обучения, анализ конкретной ситуации и обсуждение результатов).
Практическое использование моделей кластерного анализа и поиска аномалий	Финансовые технологии, основанные на обработке данных и машинном обучении: интеллектуальные кредитные сервисы, интеллектуальные страховые сервисы, интеллектуальные сервисы интернета вещей. Построение рекомендательного сервиса. Анализ эмоциональной окраски записей в социальных сетях. <i>р.8, [1]; р.9, [4]-[9]</i>	Лабораторная работа «Сегментирование потребителей» (построение компьютерной системы машинного обучения, анализ конкретной ситуации и обсуждение результатов).
Построение рекомендательных систем	Построение компьютерной системы визуализации данных, анализ конкретной ситуации и обсуждение результатов. Построение компьютерной системы визуализации данных, анализ конкретной ситуации и обсуждение результатов – 100% от трудоемкости семинарского занятия. <i>р.8, [1-3]; р.9, [4]-[8]</i>	Лабораторная работа «Построение рекомендательного сервиса» (построение компьютерной системы машинного обучения, анализ конкретной ситуации и обсуждение результатов).

6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
От данных к ценности: Введение в науку о данных	Наиболее динамично развивающиеся области искусственного интеллекта: обработка и анализ и генерация текстов и изображений. Этические проблемы искусственного интеллекта.	Работа с учебной литературой. Разбор вопросов по теме занятия.
Практическое использование моделей классификации	Технологии улучшения моделей машинного обучения. Подбор оптимальных параметров моделей. Важность подготовки данных. Генерация синтетических признаков. Работа с пропущенными данными. Работа с несбалансированными выборками	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение и оформление лабораторных работ.
Практическое использование моделей регрессии	Использование методов снижения размерности для улучшения качества моделей машинного обучения. Использование методов снижения размерности для выявления латентных факторов.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение и оформление лабораторных работ.
Практическое использование моделей кластерного анализа и поиска аномалий	Современные практические исследования Финансового университета в области обработки данных и машинного обучения: индикаторы смены состояний финансовых рынков, анализ сложных сетей в управлении городами, энергетике, на транспорте, предиктивный ремонт оборудования, системы скоринга.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение и оформление лабораторных работ.
Построение рекомендательных систем	Наиболее динамично развивающиеся области искусственного интеллекта: обработка и анализ и генерация текстов и изображений. Этические проблемы искусственного интеллекта.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение и оформление лабораторной работы.

6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

Примерные задания контрольной работы

1. Как не используют выборки из генеральной совокупности аналитики больших данных:

- а) как метод формирования комплексного суждения о генеральной совокупности случайной величины;
- б) как метод тестирования полученных моделей;
- в) как метод верификации исходных данных;
- г) как метод обучения моделей;
- д) как данные для обучения моделей.

2. Укажите лишний этап построения модели машинного обучения:

- а) сбор и верификация исходных данных;
- б) выбор факторов;
- с) построение модели;
- д) получение оценок;
- е) согласование полученных результатов с заинтересованными лицами;
- ф) проверка статистической значимости модели.

3. Глубокое обучение включает в себя:

- а) регрессионные модели;
- б) совокупность различных нейросетевых моделей;
- в) методы классификации;
- г) градиентный бустинг;
- д) обучение с подкреплением.

4. Какой метод верификации исходных данных не применяется для верификации данных о стоимости активов:

- а) семантические анализаторы;
- б) матрицы граничных значений;
- с) конверторы отраслевых классификаторов;
- д) наборы решающих правил;
- е) проверка данных с использованием колл-центра;
- ф) тестовые и валидационные выборки.

5. Какие нейронные сети лучше подходят для задач поиска аналога исследуемого объекта?

- а) сети Кохонена;
- б) сети встречного распространения;
- с) сети на радиальных базисных функциях;
- д) любые MLP нейросети;
- е) все перечисленное.

6. Какая проблема решается путем логарифмического шкалирования исходных данных?

- а) мультиколлинеарности;
- б) робастности;
- с) гетерескедастичности;
- д) гомоскедастичности.

7. Какие требования к факторам предъявляют классические статистические модели:

- а) значимость;
- б) независимость;
- с) внятная экономическая интерпретация;
- д) все вышеперечисленное.

8. При возникновении новых, не описанных ранее ситуаций, какая технология машинного обучения реагирует на них, получая данные из внешней среды?

- а) обучение с подкреплением;
- б) обучение с противником;
- с) вероятностное прогнозирование;
- д) распознавание образов.

Критерии бальной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях Кафедры анализа данных и машинного обучения Факультета информационных технологий и анализа больших данных.

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2. «Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».

Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний

Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания

ОП «Инженерия данных», ОП «Прикладные информационные системы в экономике и финансах»			
ПКН-4 Способность проектировать и создавать интеллектуальные информационные системы, выбирать метод обучения в соответствии с анализом задачи	1. Демонстрирует знание основных понятий машинного обучения и интеллектуального анализа данных, понимание области и границ применимости, основные виды задач.	Знать: основные принципы постановки задач машинного обучения. Уметь: осуществлять постановку прикладных задач машинного обучения для создания ценности.	Перечислите основные классы моделей машинного обучения, их характеристики и отличия. Осуществите сбор статистической информации по выбранной самостоятельно компании для проведения сегментирования потребителей. Представьте эти данные в виде файла формата csv. Проведите первичную обработку данных. Проведите сегментирование потребителей.
	2. Демонстрирует знание популярных инструментальных средств машинного обучения, собирает датасет, строит модели, проводит их анализ и диагностику, делает содержательные выводы.	Знать: компьютерную реализацию основных методов и алгоритмов машинного обучения в пакетах прикладных программ Уметь: выбирать компьютерную реализацию методов и алгоритмов машинного обучения для решения бизнес-задач	Перечислите основные принципы построения и практической реализации моделей классификации, регрессии, кластерного анализа, поиска аномалий, текстовой аналитики и компьютерного зрения. Осуществите сбор статистической информации по выбранной самостоятельно страховой компании для построения модели оценки величины страховой выплаты. Представьте эти данные в виде файла формата csv. Проведите первичную обработку данных. Проведите оценку страховых выплат.
	3. Презентабельно демонстрирует результаты анализа данных и машинного обучения в форме, доступной непрофессионалу, структурирует отчет по проведенному анализу.	Знать: современные инструменты визуализации результатов анализа данных и машинного обучения Уметь: применять современные инструменты визуализации результатов анализа данных и машинного обучения в форме, доступной непрофессионалу	Перечислите основные этапы развития искусственного интеллекта По одним и тем же исходным данным о заемщиках постройте модели прогнозирования вероятности банкротства на основе логистической регрессии, усиленных деревьев LightGBM и усиленных деревьев CatBoost. Сравните основные метрики качества построенных моделей и обоснуйте выбор наилучшей модели.

ОП «Инженерия данных»			
ПКП-1 Способность формировать требования, производить подготовку данных и проводить аналитические работы с использованием технологий больших данных	1. Демонстрирует знания в области компьютерных технологий, используемых для анализа больших данных	<p>Знать: компьютерную реализацию основных методов и алгоритмов машинного обучения в пакетах прикладных программ</p> <p>Уметь: выбирать компьютерную реализацию методов и алгоритмов машинного обучения для выполнения аналитических работ на основе больших данных.</p>	<p>Приведите примеры эффективного использования сервисов обработки естественного языка, текстовой аналитики, компьютерного зрения, генерации текстов по изображениям для развития цифровых продуктов.</p> <p>По данным о прошлых заемщиках вычислите доход банка в сегменте потребительского кредитования. Постройте систему машинного обучения для прогнозирования вероятности банкротства заемщика и вычислите ожидаемый доход банка в результате внедрения разработанной системы машинного обучения. Проведите обсуждение сравнения результатов до и после внедрения системы искусственного интеллекта. Перечислите основные принципы и методы сбора статистических данных. Как инструменты могут быть использованы для визуализации данных.</p> <p>Осуществите сбор статистической информации по выбранной самостоятельно компании для построения системы прогнозирования продаж. Представьте эти данные в виде файла формата csv. Проведите первичную обработку данных. Постройте компьютерную модель прогнозирования продаж. Перечислите и дайте определения основных метрик качества, которые используются при оценке качества моделей классификации.</p>

	2. Владеет навыками управления аналитическими работами на основе больших данных	<p>Знать: основные принципы постановки задач машинного обучения в аналитических работах на основе больших данных.</p> <p>Уметь: осуществлять постановку прикладных задач машинного обучения для выполнения аналитических работ на основе больших данных.</p>	<p>Приведите примеры основных пакетов, реализующих современные алгоритмы градиентного бустинга.</p> <p>Приведите сравнение эффективности использования библиотек CatBoost и LightGBM в различных задачах классификации и регрессии.</p>
	3. Выполняет аналитические работы на основе больших данных	<p>Знать: алгоритмы машинного обучения и их компьютерную реализацию в пакетах прикладных программ</p> <p>Уметь: применять компьютерную реализацию методов и алгоритмов машинного обучения для выполнения аналитических работ на основе больших данных</p>	<p>Приведите примеры эффективного использования сервисов обработки естественного языка, текстовой аналитики, компьютерного зрения, генерации текстов по изображениям для развития цифровых продуктов.</p> <p>По данным о прошлых заемщиках вычислите доход банка в сегменте потребительского кредитования. Постройте систему машинного обучения для прогнозирования вероятности банкротства заемщика и вычислите ожидаемый доход банка в результате внедрения разработанной системы машинного обучения. Проведите обсуждение сравнения результатов до и после внедрения системы искусственного интеллекта.</p>
ОП «Цифровые платформы управления предприятиями»			
ПКП-6 Способность разрабатывать, реализовывать и применять методы интеллектуального анализа данных и машинного обучения для автоматизации	1. Использует знания современных методов интеллектуального анализа данных (в том числе, больших данных) и способы их программной реализации.	<p>Знать: основные принципы постановки задач машинного обучения.</p> <p>Уметь: осуществлять постановку прикладных задач машинного обучения для создания ценности.</p>	<p>Перечислите основные классы моделей машинного обучения, их характеристики и отличия. Осуществите сбор статистической информации по выбранной самостоятельно компании для проведения сегментирования потребителей. Представьте эти данные в виде файла формата csv. Проведите первичную обработку данных. Проведите сегментирование</p>

решения неструктурированных и слабоструктурированных задач экономических предметных областей	2. Осуществляет поиск, сбор, анализ и интерпретацию данных экономических предметных областей с применением методов искусственного интеллекта и машинного обучения	<p>Знать: компьютерную реализацию основных методов и алгоритмов машинного обучения в пакетах прикладных программ</p> <p>Уметь: выбирать компьютерную реализацию методов и алгоритмов машинного обучения для решения бизнес-задач</p>	<p>потребителей.</p> <p>Перечислите основные принципы построения и практической реализации моделей классификации, регрессии, кластерного анализа, поиска аномалий, текстовой аналитики и компьютерного зрения.</p> <p>Осуществите сбор статистической информации по выбранной самостоятельно страховой компании для построения модели оценки величины страховой выплаты. Представьте эти данные в виде файла формата csv. Проведите первичную обработку данных. Проведите оценку страховых выплат.</p>
	3. Владеет современным инструментарием искусственного интеллекта и его использованием при разработке и развитии существующих финансово-экономических информационных систем.	<p>Знать: алгоритмы машинного обучения и их компьютерную реализацию в пакетах прикладных программ</p> <p>Уметь: применять компьютерную реализацию методов и алгоритмов машинного обучения при разработке и развитии существующих финансово-экономических информационных систем</p>	<p>Перечислите основные этапы развития искусственного интеллекта</p> <p>По одним и тем же исходным данным о заемщиках постройте модели прогнозирования вероятности банкротства на основе логистической регрессии, усиленных деревьев LightGBM и усиленных деревьев CatBoost. Сравните основные метрики качества построенных моделей и обоснуйте выбор наилучшей модели.</p>

Примерные вопросы для подготовки к зачету

1. Машинное обучение: основные понятия и методы.
2. Области применения методов и технологий машинного обучения. Примеры задач.
3. Цикл обработки данных.
4. Статистические основы обработки данных.
5. Современные технологии машинного обучения.

6. Принципы разработки и оценки систем машинного обучения.
7. Основные классы моделей машинного обучения.
8. Модель логистической регрессии.
9. Методы оценки качества моделей классификации.
10. Модель множественной линейной регрессии.
11. Кластерный анализ. Методы оценки качества моделей кластерного анализа.
12. Методы снижения размерности и их использование.
13. Метод главных компонент. Его геометрическое и экономическое содержание.
14. Поиск аномалий.
15. Принципы создания рекомендательных систем.
16. Перспективы развития систем машинного обучения.
17. Финансовые технологии, основанные на обработке данных и машинном обучении.
18. Модели деревьев решений.
19. Модели деревьев решений с бустингом.
20. Бэггинг. Случайный лес.
21. Стэкинг.
22. Метод градиентного спуска в машинном обучении.
23. Прогнозирование временных рядов методами AR, MA.
24. Прогнозирование временных рядов методами ARIMA.
25. Прогнозирование временных рядов методами ARIMAX, SARIMAX.

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

основная литература:

1. Соловьев, В. И. Анализ данных в экономике: теория вероятностей, прикладная статистика, обработка и визуализация данных в Microsoft Excel: учебник для направления бакалавриата "Экономика и управление" / В. И. Соловьев; Финуниверситет. – Москва : Кнорус, 2019. - 498 с. - Текст : непосредственный. - То же. – 2023. - ЭБС BOOK.ru. - URL: <https://book.ru/book/946789> (дата обращения: 15.05.2024). — Текст : электронный.

2. Кондрашов, Ю. Н. Анализ данных и машинное обучение на платформе MS SQL Server: учебное пособие / Ю. Н. Кондрашов. — Москва: Русайнс, 2020. — 303 с. — Текст : непосредственный.— То же 2023.— ЭБС BOOK.ru. — URL: <https://book.ru/book/947076> (дата обращения: 15.05.2024). — Текст : электронный.

дополнительная литература:

3. Калинина, В. Н. Анализ данных: Компьютерный практикум: учебное пособие / В. Н. Калинина, В. И. Соловьев. — Москва: КНОРУС, 2017. — 166 с. — Текст: непосредственный. — То же 2022. — ЭБС BOOK.ru. — URL: <https://www.book.ru/book/942681> (дата обращения: 15.05.2024). — Текст : электронный.

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Информационно-образовательный портал Финансового университета при Правительстве Российской Федерации <http://portal.ufrf.ru/>.
2. Сайт департамента анализа данных, принятия решений и финансовых технологий. http://fa.ru/dep/data_analysis/
3. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/> (<http://library.fa.ru/files/elibfa.pdf>)
4. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
5. Электронно-библиотечная система «Университетская библиотека ОНЛАЙН» <http://biblioclub.ru/>
6. Электронно-библиотечная система Znanium <http://www.znanium.com>
7. Электронно-библиотечная система издательства «ЮРАЙТ» <https://urait.ru/>
8. Электронно-библиотечная система издательства Проспект <http://ebs.prospekt.org/books>
9. Электронно-библиотечная система издательства Лань <https://e.lanbook.com/>
10. Деловая онлайн-библиотека Alpina Digital <http://lib.alpinadigital.ru/>
11. Электронная библиотека Издательского дома «Гребенников» <https://grebennikon.ru/>
12. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>

13. Национальная электронная библиотека <http://нэб.рф/>
14. Финансовая справочная система «Финансовый директор» <http://www.1fd.ru/>
15. Документация по Azure ML
<https://docs.microsoft.com/ru-ru/azure/machine-learning/studio/>
16. *Applied Machine Learning* / Microsoft. – <https://www.edx.org/course/applied-machine-learning-microsoft-dat203-3x-1>
17. *Data Science Essentials* / Microsoft. – <https://www.edx.org/course/data-science-essentials-microsoft-dat203-1x-3>
18. *Principles of Machine Learning* / Microsoft. – <https://www.edx.org/course/principles-machine-learning-microsoft-dat203-2x-3>
19. Профессиональный ресурс по машинному обучению. – <https://machinelearning.ru>
20. Специализация «Машинное обучение и анализ данных» / МФТИ и Яндекс. – <https://www.coursera.org/specializations/machine-learning-data-analysis>

10. Методические указания для обучающихся по освоению дисциплины
Рекомендации по подготовке к лекционным занятиям
(теоретический курс)

Самостоятельная работа студентов проходит аудиторно и внеаудиторно. Организации самостоятельной работы служит календарно-тематический план изучения дисциплины. В этом плане указана тематика лекций и практических занятий.

При подготовке к лекции целесообразно предварительно ознакомиться с ее содержанием по рекомендованным источникам и выделить наиболее трудные вопросы. Во время лекций следует конспектировать содержание лекции. После занятий следует провести работу с конспектом: отредактировать записи, оформить конспект. При оформлении целесообразно выделять специальным образом названия тем и формулировки вопросов, основные определения, формулировки теорем и примеры. Сделанные записи нужно сопоставить с учебниками и учебными пособиями и в случае расхождений проконсультироваться с преподавателем.

Методические указания по проведению практических занятий

Практические занятия структурно состоят из следующих компонент:

1. проверка наличия выполненного задания самостоятельной работы каждого студента;
2. выборочная проверка корректности выполнения домашнего задания;
3. разбор типичных ошибок, возникших в самостоятельной работе;
4. рассмотрение теоретических вопросов, связанных с текущим практическим занятием;
5. разбор методов выполнения практических заданий и решения задач;
6. корректировка заданий для самостоятельной работы студентов;
7. интерактивная форма – решение лабораторных работ по тематике занятия в малых группах (2-4 студента).

При подготовке к практическому занятию необходимо повторить или, если это требуется, изучить соответствующий теоретический материал. Во время занятия нужно точно следовать указаниям преподавателя к решению лабораторных работ. После занятий необходимо просмотреть полученные решения и восстановить имеющиеся пробелы или закончить решение. В случае затруднений отметить соответствующие задания и обратиться за консультацией к преподавателю.

Практические занятия проходят в интерактивной форме. Наряду с правильностью решения преподаватель учитывает также и активность студентов, направленную на решение предложенных задач.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем

11.1. Комплект лицензионного программного обеспечения:

- Пакет офисных программ;
- Антивирус Kaspersky;

11.2. Современные профессиональные базы данных и информационные справочные системы:

- Информационно-правовая система «Консультант Плюс»;
- Информационно-правовая система «Гарант»;
- Электронная энциклопедия: <http://ru.wikipedia.org/wiki/Wiki>;
- Система комплексного раскрытия информации «СКРИН»: <https://skrin.ru>;

11.3. Сертифицированные программные и аппаратные средства защиты информации: - не предусмотрены.

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Для освоения дисциплины необходимо любое вычислительное средство – компьютер, смартфон или планшет. Предпочтительно использование компьютера. При этом возможно использование компьютеров компьютерных классах университета. Все изучаемые технологии доступны на личных устройствах студентов в любой точке, где доступна сеть Интернет.