

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

**Краснодарский филиал
Кафедра «Математика и информатика»**

СОГЛАСОВАНО

ООО «Портал-Юг»
Генеральный директор



Е.В. Мостовой
«20» февраля 2024 г.

УТВЕРЖДАЮ
Краснодарский филиал
Финансового университета
Директор

Э.В. Соболев
«20» февраля 2024 г.

Коренева О.В.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
ПРИКЛАДНЫЕ МОДЕЛИ И МЕТОДЫ АНАЛИЗА
НЕЧИСЛОВОЙ ИНФОРМАЦИИ**

для студентов, обучающихся по направлению подготовки
01.04.02 Прикладная математика и информатика,

Направленность программы магистратуры:
Анализ больших данных и машинное обучение в экономике и финансах

*Рекомендовано Ученым советом Краснодарского филиала Финансового университета
(протокол № 12 от 20.02.2024)*

*Одобрено кафедрой «Математика и информатика»
(протокол № 13 от 13.02.2024)*

Краснодар 2024

УДК 303.732.4
ББК С5в63
К66

Рецензент: Кирий В.А., доцент кафедры «Математика и информатика», канд. физ.-мат. наук, Калайдин Е.Н., профессор кафедры «Математика и информатика», д-р физ.-мат. наук.

Коренева О.В. «Прикладные модели и методы анализа нечисловой информации». Рабочая программа дисциплины для студентов, обучающихся по направлению подготовки 01.04.02 «Прикладная математика и информатика», направленность программы магистратуры «Анализ больших данных и машинное обучение в экономике и финансах» - Краснодар: Краснодарский филиал Финансового университета при Правительстве Российской Федерации, Кафедра математики и информатики, 2024.

Дисциплина «**Прикладные модели и методы анализа нечисловой информации**» относится к Модулю дисциплин по выбору, углубляющих освоение программы магистратуры «Анализ больших данных и машинное обучение в экономике и финансах», направление подготовки 01.04.02 «Прикладная математика и информатика».

Рабочая программа дисциплины содержит цели и задачи дисциплины, требования к результатам освоения дисциплины, содержание дисциплины, тематику практических занятий и технологии их проведения, формы самостоятельной работы, систему оценивания, учебно-методическое и информационное обеспечение дисциплины.

Учебное издание

Коренева Ольга Владимировна

Прикладные модели и методы анализа нечисловой информации

*Формат 60*90/16. Гарнитура Times New Roman*

Усл. п.л. 4,7. Изд. № _____ от _____. Тираж 100 экз.
Заказ № _____

Отпечатано в Краснодарском филиале Финуниверситета

© Коренева О.В., 2024
© Краснодарский филиал Финуниверситета, 2024

СОДЕРЖАНИЕ

1. Наименование дисциплины.....	4
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине.....	4
3. Место дисциплины в структуре образовательной программы.....	6
4. Объем дисциплины(модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся.....	7
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий.....	7
6. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине.....	12
7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины.....	21
8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.....	22
9. Методические указания для обучающихся по освоению дисциплины.....	23
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем.....	24
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.....	24

1. Наименование дисциплины

«Прикладные модели и методы анализа нечисловой информации».

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

В совокупности с другими дисциплинами модуля учебная дисциплина «Прикладные модели и методы анализа нечисловой информации» обеспечивает формирование следующих компетенций: ПКН-3, ПКН-6, ПК-2, ПК-3.

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПКН-3	Способность проводить самостоятельные научные исследования в профессиональной области.	1. Демонстрирует знания в области проведения самостоятельных научных исследований в профессиональной области.	<u>Знать:</u> методы проведения самостоятельных научных исследований в профессиональной области <u>Уметь:</u> применять знания в области проведения самостоятельных научных исследований в профессиональной области
		2. Участвует в научных исследованиях в профессиональной области.	<u>Знать:</u> методы научных исследований в профессиональной области <u>Уметь:</u> применять знания для участия в научных исследованиях в профессиональной области
ПКН-6	Способность анализировать и оценивать эффективность применения методов прикладной математики и информатики.	1. Демонстрирует знание основных методов прикладной математики и информатики, применяемых в различных предметных областях.	<u>Знать:</u> основные методы прикладной математики и информатики, применяемых в различных предметных областях <u>Уметь:</u> применять знания об основных методах прикладной математики и информатики, применяемых в различных предметных областях

		2. Владеет методологией математического моделирования для решения профессиональных задач.	<u>Знать:</u> методологию математического моделирования для решения профессиональных задач <u>Уметь:</u> применять методологию математического моделирования для решения профессиональных задач
ПК-2	Способность применять лингвистические и сетевые модели машинного обучения, предобученные на больших наборах данных в экономике и финансах.	1. Демонстрирует знание современных предобученных лингвистических моделей или инструментов создания предобучения сетевых моделей.	<u>Знать:</u> основные инструменты создания лингвистических и сетевых моделей или инструментов создания предобучения сетевых моделей <u>Уметь:</u> создавать лингвистические и сетевые модели или инструменты создания предобучения сетевых моделей
		2. Владеет методологией и инструментарием использования предобученных моделей для создания решения прикладных задач, разрабатывает на базе известных решений технологию решения прикладной задачи с помощью предобученных моделей.	<u>Знать:</u> методологию и инструменты использования предобученных моделей для создания решения прикладных задач, разрабатывает на базе известных решений технологию решения прикладной задачи с помощью предобученных моделей <u>Уметь:</u> разрабатывать на базе известных решений технологию решения прикладной задачи с помощью предобученных моделей

		3. Владеет практическим навыком применения технологии использования предобученных моделей для решения прикладных задач.	<u>Знать:</u> технологии использования предобученных моделей для решения прикладных задач <u>Уметь:</u> применять технологии использования предобученных моделей для решения прикладных задач
ПК-3	Способность самостоятельно собирать наборы текстовых или сетевых данных и выполнять их структурирование и предобработку в экономике и финансах.	1. Владеет методологией поиска и сбора наборов данных для решаемой прикладной задачи.	<u>Знать:</u> методологию поиска и сбора наборов данных для решаемой прикладной задачи <u>Уметь:</u> применять методологию поиска и сбора наборов данных для решаемой прикладной задачи
		2. Владеет способностью агрегировать и структурировать собранные первичные данные.	<u>Знать:</u> приемы собирать текстовые или сетевые данные и выполнять их структурирование и предобработку <u>Уметь:</u> агрегировать и структурировать собранные первичные данные
		3. Владеет практическим навыком предобработки наборов экономико-финансовых данных для их последующего использования в моделях машинного обучения.	<u>Знать:</u> методы предобработки наборов экономико-финансовых данных для их последующего использования в моделях машинного обучения <u>Уметь:</u> предобрабатывать наборы экономико-финансовых данных для их последующего использования в моделях машинного обучения

3. Место дисциплины в структуре образовательной программы

Дисциплина «Прикладные модели и методы анализа нечисловой информации» относится к Модулю дисциплин по выбору, углубляющих освоение программы магистратуры «Анализ больших данных и машинное обучение в экономике и финансах», направление подготовки 01.04.02 «Прикладная математика и информатика».

Изучение дисциплины «Прикладные модели и методы анализа нечисловой информации» базируется на знаниях, полученных в ходе изучения дисциплин «Методология поиска источников данных и подготовки данных для анализа» и «Построение и оценка моделей машинного обучения».

В процессе изучения дисциплины происходит знакомство с основными понятиями, методологиями, моделями, методами, методиками и технологиями анализа категориальных (или нечисловых) данных и их применением для улучшения качества моделей машинного обучения. Эти знания будут положены в основу формирования профессиональных компетенций выпускника направления «Прикладная математика и информатика». При этом студенты приобретают опыт практического использования изучаемых технологий в практических задачах, связанных с финансовыми технологиями.

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

Очная форма обучения, 2024 г.

Вид учебной работы по дисциплине	Всего (в з.е. и часах)	Модуль 5 (в часах)
Общая трудоёмкость дисциплины	3/108	108
Контактная работа- Аудиторные занятия	32	32
Лекции	8	8
Семинары, практические занятия	24	24
Самостоятельная работа	76	76
Вид текущего контроля	контрольная работа	контрольная работа
Вид промежуточной аттестации	зачет	зачет

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

5.1. Содержание дисциплины

Раздел 1. Шкалы измерения категориальных признаков

Основные шкалы измерения числовых признаков: интервальная шкала (шкала разностей), шкала отношений (абсолютная шкала). Категориальные (нечисловые) признаки: определение, классификация и примеры. Основные шкалы измерения категориальных признаков: номинальная и ранговая (ранговая). Бинарные (дихотомические) признаки.

Вероятностные модели бинарных и ранговых признаков.

Преобразование категориальных признаков в бинарные (one-hot encoding). Реализация инструментария one-hot encoding в Scikit-learn и Microsoft Azure ML Studio.

Раздел 2. Измерение связи между категориальными признаками с помощью ранговых коэффициентов корреляции

Коэффициент корреляции Пирсона как инструмент измерения связи между числовыми признаками.

Ранговый коэффициент корреляции Спирмена r . Вычисление рангового коэффициента корреляции Спирмена в случае отсутствия связанных рангов и в случае наличия связанных рангов.

Ранговый коэффициент корреляции Кендалла τ . Вычисление рангового коэффициента корреляции Кендалла в случае отсутствия связанных рангов и в случае наличия связанных рангов.

Реализация алгоритмов вычисления ранговых коэффициентов корреляции в scikit-learn.

Практическая интерпретация ранговых коэффициентов корреляции.

Оценка связи между двумя ранговыми признаками с помощью коэффициентов ранговой корреляции Спирмена и Кендалла. Оценка связи между ранговым и числовым признаками с помощью коэффициентов ранговой корреляции Спирмена и Кендалла. Оценка связи между двумя ранговым и дихотомическим признаками с помощью коэффициентов ранговой корреляции Спирмена и Кендалла.

Оценка связи между несколькими ранговыми признаками. Коэффициент конкордации.

Раздел 3. Измерение связи между категориальными признаками с помощью таблиц сопряженности

Двумерные таблицы сопряженности. Проверка гипотезы о независимости двух категориальных признаков. Критерий «хи-квадрат» и критерий максимального правдоподобия. Оценка связи между двумя категориальными признаками на основе таблиц сопряженности с помощью меры «хи-квадрат».

Коэффициенты Чупрова, Крамера и Пирсона, их точечные и интервальные оценки. Практическая интерпретация коэффициентов Чупрова, Крамера и Пирсона. Декомпозиция таблиц сопряженности. Многомерные таблицы сопряженности.

Реализация инструментов отбора факторных признаков в scikit-learn и Microsoft Azure ML Studio.

Раздел 4. Теоретико-информационные меры связи

Понятие количества информации. Полная и условная энтропия распределения признака. Вычисление полной и условной энтропии по таблице сопряженности.

Направленные коэффициенты нормированной информации. Симметризованный коэффициент нормированной информации.

Направленные и симметризованные меры связи Гудмена и Краскала (X и T). Практическая интерпретация теоретико-информационных мер связи.

Раздел 5. Использование методов анализа нечисловой информации в машинном обучении

Обработка текстов - токенизация, нормализация, стемминг, лемматизация, извлечение признаков из текста (мешок слов, TD_IDF, Skip-граммы, N-граммы, хэширование).

Основные задачи анализа текстов: анализ эмоциональной окраски, определение тем, кластеризация текстов.

Продвинутые задачи: поиск имен в тексте, аннотация, чат-боты, «клонирование» текста, поиск синонимов.

Представление набора текстов в базах данных.

Применение нейросетей в задачах NLP.

5.2. Учебно-тематический план

№ п/ п	Наименование тем (разделов) дисциплины	Трудоёмкость в часах					Формы текущего контроля успеваемости
		Всего	Контактная работа — Аудиторные занятия			Самостоя тельная работа	
			Об щая, в т.ч.:	Лекции	Семинары, практичес кие занятия		
1	Шкалы измерения категориальных признаков	16	6	2	4	10	Участие в решении задач на практических занятиях. Обсуждения по результатам самостоятельной работы
2	Измерение связи между категориальными признаками с помощью ранговых коэффициентов корреляции	17	5	1	4	12	
3	Измерение связи между категориальными признаками с помощью таблиц сопряженности	21	5	1	4	16	
4	Теоретикоинформационные меры связи	24	6	2	4	18	
5	Использование методов анализа нечисловой информации в машинном обучении	30	10	2	8	20	
	В целом по дисциплине	108	32	8	24	76	Согласно учебному плану: контрольная работа
	Итого в %	100	30	25	75	70	

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
Шкалы измерения категориальных признаков	1. Основные шкалы измерения числовых признаков: интервальная шкала (шкала разностей), шкала отношений (абсолютная шкала). 2. Основные шкалы измерения категориальных признаков: номинальная и ранговая (ранговая). Бинарные (дихотомические) признаки. 3. Преобразование категориальных признаков в бинарные (one-hot encoding). Реализация инструментария one-hot encoding в Scikit-learn и Microsoft Azure ML Studio. <i>Рекомендуемые источники: [8:1,2,; 9:1-32]</i>	Решение задач в интерактивной форме, обсуждение результатов
Измерение связи между категориальными признаками с помощью ранговых коэффициентов корреляции	1. Оценка связи между двумя ранговыми случайными величинами. 2. Оценка связи между ранговой и категориальной случайными величинами. 3. Оценка связи между несколькими ранговыми случайными величинами. <i>Рекомендуемые источники: [8:1,2,; 9:1-32]</i>	Решение задач в интерактивной форме, обсуждение результатов
Измерение связи между категориальными признаками с помощью таблиц сопряженности	1. Оценка связи между двумя категориальными случайными величинами <i>Рекомендуемые источники: [8:1,2,; 9:1-32]</i>	Решение задач в интерактивной форме, обсуждение результатов
Теоретикоинформационные меры связи	1. Оценка связи между двумя категориальными случайными величинами 2. Оценка связи между тремя категориальными случайными величинами. <i>Рекомендуемые источники: [8:1,2,; 9:1-32]</i>	Решение задач в интерактивной форме, обсуждение результатов
Использование методов анализа нечисловой информации в машинном обучении	1. Токенизация, нормализация, извлечение признаков из текста 2. Определение тональности в социальных сетях. 3. Представление набора текстов в базах данных. <i>Рекомендуемые источники: [8:1,2,; 9:1-32]</i>	Решение задач в интерактивной форме, обсуждение результатов

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

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Прикладные модели и методы анализа нечисловой информации».

Фонд оценочных средств включает контрольные материалы для проведения текущего контроля и промежуточной аттестации.

Вопросы для оценки знаний и умений, характеризующих формирование компетенций

Шифр компетенции	Вопросы	Правильный ответ
ПKN-3	1. Переменная, которая измеряется в реальных числах, например, возраст или доход	Количественные (непрерывные)
	2. Переменная, которая принимает значения из ограниченного множества, например, пол или цвет	Категориальные (дискретные)
	3. Категориальный признак, который может быть упорядочен, например, оценка продукта по шкале от 1 до 5	Порядковые
	4. Категориальный признак, который может принимать только два значения, например, есть/нет или правда/ложь	Бинарные
	5. Алгоритм классификации на основе вероятностного подхода, основанного на теореме Байеса	Байесовский классификатор
ПKN-6	6. Какие методы данных включают очистку данных, заполнение пропущенных значений, нормализацию, кодирование категориальных переменных, снижение размерности данных и др.	Предобработки, предварительной обработки
	7. Назовите метод машинного обучения, при котором модель обучается на основе обучения с учителем и используется для предсказания значений целевой переменной.	Регрессии
	8. Назовите метод, который используется для оценки важности признаков и выбора оптимального набора признаков для построения модели машинного обучения.	Отбора признаков
	9. Метод машинного обучения, при котором модель обучается на основе	Многоклассовая классификация

	обучения на примерах и используется для классификации объектов на несколько категорий это...	
	10. Методы _... используются для оптимизации параметров модели машинного обучения и выбора оптимальных гиперпараметров с целью достижения лучшей производительности модели.	Настройки параметров
ПК-2	11. Это базовая, в основном низкоуровневая библиотека, которая предназначена для анализа, классификации и обработки изображений. Широко используется в таких языках как C, C++, Python и Java.	OpenCV
	12. Методы обучения, которые используются для построения моделей машинного обучения, которые могут обрабатывать несколько типов входных данных, таких как изображения, тексты и звуки.	Мультимодального
	13. Назовите модели - метод машинного обучения, при котором модель обучается на основе примеров, а затем используется для генерации новых данных из той же области.	Генеративные
	14. Назовите методы, которые используются для оценки качества модели машинного обучения, сравнения моделей и выбора оптимальной модели.	Кросс-валидации, кроссвалидации, кросс валидации, перекрестной проверки
	15. Техника, которая применяется для борьбы с переобучением нейронных сетей, заключается в случайном исключении части нейронов в процессе обучения	Dropout
ПК-3	16. Задача, в которой необходимо определить тональность текста, например, позитивная или негативная	Классификация текстов (определение эмоциональной окраски)
	17. Задача, в которой необходимо находить и выделять в тексте имена, организации, даты и другие сущности.	Распознавание именованных сущностей
	18. Задача, в которой необходимо переводить текст с одного языка на другой.	Машинный перевод
	19. Задача, в которой необходимо генерировать текст на основе определенных правил и шаблонов.	Генерация текста
	20. На этом этапе определяется, какой алгоритм машинного обучения будет	Выбор метода и создание модели

	использоваться, настраиваются параметры модели, и происходит ее обучение на обучающей выборке	
--	---	--

Практико-ориентированные задания

Шифр компетенции	Практико-ориентированные задания	Правильный ответ
ПKN-3	1. Определить тональность текста «Госдума приняла в третьем, окончательном чтении законопроект о пенсионной реформе. «За» проголосовали 332 депутата, «против» - 83. Во время заседания принятие документа в очередной раз поддержала только «Единая Россия», все остальные партии - «Справедливая Россия», ЛДПР и КПРФ - выступили с его критикой. Во втором - ключевом - чтении реформа была принята накануне, 26 сентября. Рассмотрение и проекта, и поправок к нему заняло четыре часа. Документ предусматривает повышение пенсионного возраста для женщин с 55 до 60 лет, для мужчин - с 60 до 65. Законопроект был внесен в Госдуму в июне. В изначальной версии, предложенной правительством, для женщин возраст выхода на пенсию должен был повыситься до 63 лет, а для мужчин - до 65. В июле документ был принят в первом чтении. В августе президент России Владимир Путин выступил с телеобращением и объявил о смягчении условий реформы. Он предложил на переходный период считать предпенсионным возрастом пять лет до наступления срока выхода на пенсию и рекомендовал повысить пенсионный возраст для женщин не до 63, а до 60 лет».	Тематика Политика, экономика и бизнес. Тональность нейтральная
	2. Дан текст, который расположен на жестком диске компьютера в виде текстового документа text.txt: «Денежные рынки на основе товаров и товарные рынки в первоначальной форме, как полагают, возникли в Шумере между 4500 до н.э. и 4000 до н.э. Шумеры сначала использовали глиняные символы, запечатанные в сосуде из глины, затем глиняные таблетки письма, чтобы представлять сумму - например, число коз, которое должно быть поставлено. Эти обещания времени и дата поставки напоминают фьючерсный контракт. Ранние цивилизации по-разному использовали чушки, редкие ракушки или другие предметы, как товарные деньги. С этого времени трейдеры стремились найти пути упрощения и стандартизации торговых контрактов. Золотые и серебряные рынки развивались в	Естественность текста 90,64%

	<p>классических цивилизациях. Сначала драгоценные металлы ценили за их красоту и внутреннюю ценность и связывали их с королевской властью. В свое время, они использовались для торговли и обменивались на другие товары и сырье, или представляли собой оплату труда. Определенные количества золота затем стали деньгами. Дефицит золота, уникальная плотность и то, как оно может быть легко получено, обработано и измерено, сделали его естественным торговым активом.». Оценить качество контента по закону Ципфа.</p>	
	<p>3. Дан текст, который расположен на жестком диске компьютера в виде текстового документа text.txt: «В 1864 в США пшеница, зерно, рогатый скот и свиньи широко продавались и покупались при помощи стандартных инструментов, которыми начала торговать Чикагская товарная биржа (СВОТ), старейшая фьючерсная и опционная биржа в мире. Другие продовольственные товары были добавлены к Закону о товарной бирже и стали продаваться и покупаться через СВОТ в 1930-х и 1940-х годов, при этом расширенный список включал рис, молотые продукты, сливочное масло, яйца, ирландский картофель и сою. Успешные товарные рынки требуют широкого консенсуса в изменении продукта, чтобы сделать каждый товар приемлемым для торговли, таким, как золота в слитках определенной чистоты.». Определить количество и список стоп-слов.</p>	<p>Количество символов: 657 Количество символов без пробелов: 566 Количество слов: 91 Количество стоп-слов: 32</p>
	<p>4. Дан текст, который расположен на жестком диске компьютера в виде текстового документа text.txt: «Товарный рынок относится к рынкам, на котором торгуют первичным сырьем, а не продукцией обрабатывающей промышленности. Мягкими товарами являются сельскохозяйственные продукты, такие как пшеница, кофе, какао и сахар. Жесткими товарами называются, например, золото, резина и нефть. Инвесторы имеют доступ к примерно 50 крупным товарным рынкам по всему миру с объемом чисто финансовых операций намного превосходящим объем</p>	<p>Количество слов: 128 Количество уникальных слов: 113 Количество символов: 1042 Количество символов без пробелов: 914 Лемматизированные слова: торговлю 2 финансовым 2 являются 2</p>

	<p>физических торгов, по результатам которых товары поставляются. Фьючерсные контракты являются самым старым из способов инвестирования в сырьевые товары. Фьючерсы обеспечены материальными активами. Товарные рынки включают физическую торговлю и торговлю деривативами с использованием спотовых цен, форвардов, фьючерсов и опционов на фьючерсы. Фермеры использовали простую форму торговли производными инструментами на товарном рынке в течение многих столетий для управления ценовыми рисками. Производным финансовым инструментом (финансовым деривативом) является финансовый инструмент, стоимость которого происходит от товара, который называют базовым. Деривативы бывают либо биржевые, либо внебиржевые (ОТС)». Произвести его лемматизацию.</p>	<p>товарами 2 товары 2 рынкам 2</p>
	<p>5. Провести анализ текста по закону Ципфа и с помощью алгоритмов BM25 и BM25F: «Товарный индексный фонд является фондом, средства которого инвестируются в финансовые инструменты на основе или привязаны к товарному индексу. В почти каждом случае индекс на самом деле является Commodity Futures Index. Первый такой показатель был Индекс Бюро по изучению товаров (CRB), который стал рассчитываться в 1958 году. Его введение не сделало данный индекс полезным в качестве инвестиционного индекса. Первый практически инвестиционный индекс товарных фьючерсов – Goldman Sachs Commodity Index, созданный в 1991 году, и известен как "GSCI". Следующим был Dow Jones AIG Commodity Index. Он отличался от GSCI в первую очередь в весах, выделяемых для каждого товара. DJ AIG имел механизмы периодически ограничивать вес любого одного товара и удалять товары, веса которых стали слишком малы. После финансовых</p>	<p>Естественность текста: 67%</p> <p>Всего слов: 135 Отображено: 6 Отфильтровано стоп-слов: 55</p>

	проблем AIG с в 2008 года права Индекс были проданы UBS и в настоящее время индекс известен как DJUBS».	
ПKN-6	6. Время решения тестовой задачи составило (сек): 59, 50, 54, 52, 54, 56, 58, 52, 51, 48, 59, 62, 53, 60, 43, 48, 45, 59. Вычислите точечные оценки математического ожидания и дисперсии.	$m=53,5;$ $S^2=29,323$
	7. Вычислить выборочное среднее, выборочную дисперсию, моду и медиану выборки: 7, 3, 3, 6, 4, 5, 1, 2, 1, 3.	$\bar{x} = 3,5;$ $D_x = 3,65$ $Mo = 3; Me = 3$
	8. Предполагая время безотказной работы устройства распределенным нормально, построить 99%-й доверительный интервал для его математического ожидания, если по результатам испытаний 25 устройств среднее время оказалось 1240 ч. Дисперсия генеральной совокупности известна и равна 70 ч ² .	$1235,69 < m < 1244,31$
	9. Измерялось сопротивление партии резисторов в кОм: 56, 68, 75, 80, 73, 59, 66, 77, 64. Предполагая распределение нормальным, построить 95%-й доверительный интервал для математического ожидания.	$62,35 < m < 74,98$
	10. Построить 90%-й и 95%-й доверительные интервалы для дисперсии диаметра вала, если $n = 16$, $\bar{x} = 29$ мм, $s^2 = 4,5$ мм ² .	90%: $2,7 < \sigma^2 < 9,3$ 95%: $2,45 < \sigma^2 < 10,78$
ПК-2	11. Фиксировалось время выполнения задания в минутах, получены следующие данные: 25, 20, 31, 27, 24, 28, 22, 26, 25, 22. Проверить гипотезу о том, что среднее время выполнения задания составляет 25 минут.	Гипотеза принимается, можно считать, что среднее время выполнения задания составляет 25 минут
	12. Время безотказной работы приборов первой группы составило 1200, 1240, 1310, 1320, 1330, 1360 ч, стандартное отклонение генеральной совокупности для этой группы 150 ч. Для второй группы получены следующие данные: 1260, 1270, 1270, 1280, 1290, 1290,	Время безотказной работы в двух группах приборов можно считать одинаковым

	1300, 1300 ч, стандартное отклонение – 80 ч. Можно ли считать время безотказной работы в двух группах приборов одинаковым?																					
	13. Технология производства некоторого вещества дает в среднем 1000 кг вещества в сутки со среднеквадратическим отклонением среднего, равным 80 кг. Новая технология производства в среднем дает 1100 кг вещества с тем же среднеквадратическим отклонением. Можно ли считать, что новая технология обеспечивает повышение производительности, если $\alpha = 0,05$.	Новая технология не обеспечивает повышение производительности.																				
	14. Точность наладки станка автомата, производящего некоторые детали, характеризуется дисперсией длины деталей. Если эта 30 величина будет больше 400 мкм ² , станок останавливается для наладки. Выборочная дисперсия длины 15 случайно отобранных деталей из продукции станка оказалась равной 680 мкм ² . Нужно ли производить наладку станка, если уровень значимости $\alpha = 0,1$.	Наладку станка производить нужно.																				
ПК-3	15. Определить количество информации и энтропию сообщения из 5 букв, если число букв в алфавите равно 32 и все сообщения равновероятны.	25 бит																				
	16. Найти выборочный коэффициент сопряжённости Крамера для признаков А и В. Проверить при $\alpha=0,05$ гипотезу о независимости этих признаков. <table border="1"><tr><td></td><td>A_1</td><td>A_2</td><td>A_3</td><td>Σ</td></tr><tr><td>B_1</td><td>15</td><td>18</td><td>6</td><td>39</td></tr><tr><td>B_2</td><td>3</td><td>5</td><td>3</td><td>11</td></tr><tr><td>Σ</td><td>18</td><td>23</td><td>9</td><td>50</td></tr></table>		A_1	A_2	A_3	Σ	B_1	15	18	6	39	B_2	3	5	3	11	Σ	18	23	9	50	Полученное значение говорит о явном наличии связи между признаками. Коэффициент Крамера, имеющий значение больше 0,1 также свидетельствует об этом.
		A_1	A_2	A_3	Σ																	
B_1	15	18	6	39																		
B_2	3	5	3	11																		
Σ	18	23	9	50																		
17. Кодировка Unicode использует для кодирования одного символа 16 бит. Сколько байт потребуется для кодирования следующего сообщения, набранного случайно на клавиатуре несмышлёным ребенком (пробелов в сообщении нет):	36 байтов																					

	ЖЖШ&550@{RK+M^G8?Q	
	18. Проводилась одноканальная (моно) звукозапись с частотой дискретизации 16 кГц и 32-битным разрешением. В результате был получен файл размером 20 Мбайт, сжатие данных не производилось. В течение какого времени проводилась запись?	328 секунд или 5,5 минут

Тесты

Шифр компетенции	Тестовые задания	Правильный ответ
ПКН-3	1. Из нижеперечисленного, относится к обучающей выборке: А) примеры с известными ответами Б) классификация данных В) алгоритм решающий функцию Г) алгоритм решающий выборку	А
	2. Задача классификации - это ... А) поиск группы, в которую можно объединить объекты, имеющих схожие характеристики Б) определение зависимости между двумя числовыми переменными В) предсказание числового значения целевой переменной на основе значений независимых переменных Г) поиск класса, к которому относится объект, на основе его признаков	Г
	3. Задача регрессии - это ... А) предсказание числового значения целевой переменной на основе значений независимых переменных Б) поиск класса, к которому относится объект, на основе его признаков В) описание связи между двумя категориальными переменными Г) сравнительный анализ нескольких числовых переменных с помощью графиков	А
ПКН-6	4. Задача ранжирования - это: А) определение порядка признака согласно рангу Б) множество объектов, разделенных на классы В) исследование влияние одного или нескольких признаков на объект Г) множество объектов, разделенных на ранги	А
	5. Преимущество библиотеки numpy перед	Б

	<p>использованием обычных списков в Python при работе с большими массивами данных:</p> <p>А) Легко добавлять/удалять элементы из списка</p> <p>Б) Позволяет быстро выполнять математические операции над массивами</p> <p>В) Обеспечивает простой доступ к элементам массива</p> <p>Г) Выполняет конвертацию между различными форматами данных.</p>	
	<p>6. Отличие тензоров от numpy-массивов заключается в том, что ...</p> <p>А) Тензоры неизменяемы и могут использоваться на аппаратных ускорителях (GPU), в то время как numpy-массивы могут использоваться для широкого спектра задач.</p> <p>Б) Numpy-массивы поддерживают больше типов данных по сравнению с тензорами.</p> <p>В) Тензоры работают только на GPU.</p> <p>Г) Numpy-массивы могут быть только двумерными.</p>	А
ПК-2	<p>7. Типы слоев, которые есть в TensorFlow:</p> <p>А) Dense, Convolutional, RNN</p> <p>Б) Dense, Pooling, Dropout</p> <p>В) BatchNorm, Convolutional, Dropout</p> <p>Г) Все вышеперечисленное</p>	Г
	<p>8. Задачей восстановления регрессии называют ...</p> <p>А) Задачу обучения по прецедентам при $Y \neq R$</p> <p>Б) Задачу поиска вектора параметров a^*</p> <p>В) Задачу поиска минимума среднего квадрата ошибки</p> <p>Г) Задачу модели регрессии</p>	А
	<p>9. Верные утверждения:</p> <p>А) класс — это множество всех объектов с определенным значением</p> <p>Б) в задачах регрессии допустимым ответом является действительное число или числовой вектор</p> <p>В) в задачах ранжирования ответы получают сразу на множестве объектов</p> <p>Г) области минимального объёма с достаточно гладкой границей являются основной составляющей задач ранжирования</p>	А,Б,В
ПК-3	<p>10. Задачи, из нижеперечисленных, являются задачами ранжирования:</p> <p>А) задачи поискового вывода</p> <p>Б) определение наиболее целесообразного способа лечения</p> <p>В) обнаружение спама</p> <p>Г) задачи поискового вывода</p>	А,Б
	<p>11. Задачи, из нижеперечисленных, являются задачами прогнозирования:</p> <p>А) математический прогноз даты сильных</p>	А,Б

	землетрясений Б) прогнозирование вероятности летального исхода В) определение длительности и исхода заболевания Г) обнаружение спама	
	12. Методы, использующиеся для борьбы с переобучением в нейронной сети: А) Регуляризация Б) Уменьшение количества параметров модели В) добавление слоя Dropout Г) Все вышеперечисленное	Г

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

7.1. Основная литература

1. Келлехер, Д. Наука о данных: базовый курс / Джон Келлехер, Брендан Тирни; пер. с англ. - Москва: Альпина Паблишер, 2020. - 222 с. - ЭБС ZNANIUM.com. - URL: <https://znanium.com/catalog/product/1221800>. - Текст : электронный.

2. Крылов, В. Е., Математические методы в экономике. : учебник / В. Е. Крылов. — Москва : КноРус, 2022. — 216 с. — ISBN 978-5-406-08727-5. — URL: <https://book.ru/book/940661> (дата обращения: 26.11.2024). — Текст : электронный.

7.2. Дополнительная литература:

1. Ганегедара, Т. Обработка естественного языка с TensorFlow : руководство / Т. Ганегедара ; перевод с англ. В. С. Яценкова. — Москва : ДМК Пресс, 2020.—382 с. — ЭБС Лань. — URL: <https://e.lanbook.com/book/140584>. — Текст : электронный.

2. Павленков М. Н. Математические методы в экономике и управлении: учебное пособие / Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского -2022,-233.

3. Попова, В. Б. Статистический анализ и прогнозирование с использованием пакетов прикладных программ : учебное пособие / В. Б. Попова, И. В. Фецович. — Воронеж : Мичуринский ГАУ, 2021. — 147 с. — ISBN 978-5-94664-432-7.— Текст: электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/253565>. — Режим доступа: для авториз. пользователей.

4. Устюгов, В. А. Сложные системы: целостность, иерархия, идентичность : монография / В. А. Устюгов, В. И. Кудашов, М. А. Петров. — Красноярск : СФУ, 2020. — 204 с. — ISBN 978-5-7638-4329-3. — Текст : электронный //

Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/181630>. — Режим доступа: для авториз. пользователей.

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Портал искусственного интеллекта [Электронный ресурс]: сайт. - Режим доступа: <http://www.aiportal.ru>
2. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/>
3. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
4. Электронно-библиотечная система «Университетская библиотека ОНЛАЙН» <http://biblioclub.ru/>
5. Электронно-библиотечная система Znanium <http://www.znaniy.com>
6. Электронно-библиотечная система издательства «ЮРАЙТ» <https://urait.ru/>
7. Электронно-библиотечная система издательства Проспект <http://ebs.prospekt.org/books>
8. Электронно-библиотечная система издательства «Лань» <https://e.lanbook.com/>
9. Электронная библиотека Издательского дома «Гребенников» <https://grebennikon.ru/>
10. Деловая онлайн-библиотека Alpina Digital <http://lib.alpinadigital.ru/>
11. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
12. Национальная электронная библиотека ИЭП://НЭБ.РФ/
13. Финансовая справочная система «Финансовый директор» <http://www.1fd.ru/>
14. Ресурсы информационно-аналитического агентства по финансовым рынкам Cbonds.ru <https://cbonds.ru/>
15. СПАРК <https://spark-interfax.ru/>
16. Academic Reference <http://ar.cnki.net/ACADREF>
17. Bank Focus <http://library.fa.ru/resource.asp?id=527>
18. Пакет баз данных компании EBSCO Publishing, крупнейшего агрегатора научных ресурсов ведущих издательств мира <http://search.ebscohost.com>
19. Электронные продукты издательства Elsevier <http://www.sciencedirect.com>
20. Emerald: Management eJournal Portfolio <https://www.emerald.com/insight/>
21. Информационно-аналитическая база данных EMIS Global <https://www.emis.com/php/companies/overview/index>
22. Реферативная база данных по математике MathSciNET <https://mathscinet.ams.org/mathscinet/>
23. Oxford Scholarship Online <https://oxford.universitypressscholarship.com/>
24. Коллекция научных журналов Oxford University Press <https://academic.oup.com/journals/>
25. ProQuest: База данных Business Ebook Subscription на платформе Ebook

Central <https://search.proquest.com/>

26. ProQuest Dissertations & Theses A&I <https://search.proquest.com/>

27. База данных RUSLANA компании Bureau van Dijk

<https://ruslana.bvdep.com/>

28. Scopus <https://www.scopus.com>

29. Электронная коллекция книг издательства Springer: Springer eBooks

<http://link.springer.com/>

30. Интерактивная финансовая информационная система компании Bloomberg
Система Thomson Reuters Eikon

Web of Science <http://apps.webofknowledge.com>

9. Методические указания для обучающихся по освоению дисциплины

При изучении теоретического материала необходимо опираться на рабочую программу дисциплины, материалы лекций и литературу из основного списка. Кроме этого, необходимо активно работать с Интернет-источниками и пособиями других авторов, помогающими усвоить материал отдельных разделов программы.

Необходимо конспектировать лекции, пометая сложные и непонятные моменты с тем, чтобы задать вопросы лектору в конце лекции или же на консультации.

При подготовке к семинарским занятиям необходимо изучить вопросы, вынесенные на самостоятельное изучение, так как семинарские занятия предполагают их обсуждение и дискуссию по теме; кроме того, задания для самостоятельной работы необходимы для того, чтобы успешно выполнить самостоятельные задания на семинарах.

Индивидуальные задания для работы на компьютере, файлы с выполненными заданиями необходимо хранить в личной сетевой папке в компьютерной сети вуза.

10. 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:

1. ОС Astra Linux
2. Libre Office
3. Антивирус Kaspersky

11.2. Современные профессиональные базы данных и информационные

справочные системы:

1. Информационно-правовая система «Гарант»
2. Информационно-правовая система «Консультант Плюс»
3. Система комплексного раскрытия информации «СКРИН» -

<http://www.skrin.ru/>

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

11.4. Microsoft Azure Machine Learning Studio

11.5. Jupiter Notebooks и Microsoft Azure Notebooks

11.6. Scikit-learn.

11.7. Vowpal Wabbit

11.8. Дистрибутив языка Python 3.4 (или более поздней версии) Anaconda

3

11.9. Kaggle.

11.10. Виртуальные машины Oracle VirtualBox

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

Материально-техническая база Краснодарского филиала Финансового университета соответствует действующим противопожарным правилам и нормам, обеспечивает проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Образовательный процесс обеспечивается специальными помещениями, которые представляют собой аудитории для проведения занятий лекционного типа, занятий семинарского типа, выполнения курсовых работ, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, помещения для самостоятельной работы студентов и помещения для хранения и профилактического обслуживания учебного оборудования. Специальные помещения укомплектованы специализированной мебелью и техническими средствами обучения, обеспечивающими представление учебной информации большой аудитории, демонстрационным оборудованием.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой, обеспечивающей доступ к сети Интернет и электронной информационно-образовательной среде Краснодарского филиала Финансового университета.