

Федеральное государственное образовательное бюджетное
учреждение высшего образования
«Финансовый университет при Правительстве Российской Федерации»
(Финуниверситет)
Калужский филиал Финуниверситета
Кафедра «Бизнес-информатика и высшая математика»

«УТВЕРЖДАЮ»
Директор Калужского филиала
Финуниверситета
 В.А. Матчинов
«27» июня 2024 г.

И.В. Винокуров

ТЕХНОЛОГИИ ОБРАБОТКИ И АНАЛИЗА БОЛЬШИХ ДАННЫХ

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки
38.03.01 «Бизнес-анализ, налоги и аудит», профиль: «Учёт, анализ и аудит»

*Рекомендовано Ученым советом Калужского филиала Финуниверситета
(протокол № 16 от 27 июня 2024 г.)*

*Одобрено кафедрой «Бизнес-информатика и высшая математика»
(протокол № 12 от 27 июня 2024 г.)*

Калуга 2024

Рабочая программа предназначена для преподавания дисциплины «Технологии обработки и анализа больших данных» студентам, обучающимся по направлению подготовки 38.03.01 «Бизнес-анализ, налоги и аудит», профиль: «Учёт, анализ и аудит», по очной форме обучения.

В рабочей программе излагаются планируемые результаты освоения дисциплины, содержание дисциплины, тематика и содержание семинаров и практических занятий, технологии их проведения. Приводится перечень учебно-методического обеспечения для самостоятельной работы обучающихся, фонд оценочных средств для проведения промежуточной аттестации обучающихся, перечень основной и дополнительной литературы, а также ресурсов информационно-телекоммуникационной сети Интернет.

СОГЛАСОВАНО:


Заместитель директора
по учебно-методической работе
«27» июня 2024 г.

 /Орловцева О.М./

Начальник учебно-методического отдела
«27» июня 2024 г.

 /Толстикова В.С./

Заведующий кафедрой
«Бизнес-информатика и высшая математика»
«27» июня 2024 г.

 /Дробышева И.В./

Содержание

1. Наименование дисциплины	4
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения, соотнесённых с планируемыми результатами обучения по дисциплине	4
3. Место дисциплины в структуре образовательной программы.....	4
4. Объем дисциплины в зачётных 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	5
5.3. Содержание семинаров, практических занятий.....	6
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине.....	7
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы.....	7
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю успеваемости.....	8
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине.....	8
7.1 Перечень компетенций с указанием индикаторов их достижения в процессе освоения дисциплины	8
7.2 Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний	8
8. Перечень основной и дополнительной литературы, необходимой для освоения дисциплины	9
9. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины.....	10
10. Методические указания для обучающихся по освоению дисциплины.....	10
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем.....	10
11.1 Комплект лицензионного программного обеспечения	10
11.2 Современные профессиональные базы данных и информационные справочные системы	11
11.3 Сертифицированные программные и аппаратные средства защиты информации: не предусмотрены	11
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.....	11

1. Наименование дисциплины

«Технологии обработки и анализа больших данных»

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

В результате изучения дисциплины у студентов должны быть сформированы следующие компетенции:

Таблица 1

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (владения, умения и знания), соотнесённые с компетенциями/ индикаторами достижения компетенции
ПКП-5	Способность к использованию специальных программных продуктов, применяемых для выполнения бухгалтерско-аналитических и контрольных функций в экономическом субъекте	1. Использует специальные программные продукты для выполнения бухгалтерско-аналитических и контрольных функций в экономическом субъекте	Знать специализированные программные продукты для анализа данных Уметь применять специализированные программные продукты для анализа данных
		2. Демонстрирует владение специальными программными продуктами, применяемых для выполнения бухгалтерско-аналитических и контрольных функций в экономическом субъекте	Знать специальные программные продукты, применяемые для выполнения бухгалтерско-аналитических и контрольных функций Уметь решать задачи подготовки, обработки и анализа полученных результатов с использованием современных программных продуктов

3. Место дисциплины в структуре образовательной программы

Дисциплина «Технологии обработки и анализа больших данных» относится к элективной дисциплине модуля «Технологии анализа и прогнозирования данных», отражающего специфику ВУЗа по направлению 38.03.01 «Бизнес-анализ, налоги и аудит», профиль: «Учёт, анализ и аудит».

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

Таблица 2

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 6 (в часах)
Общая трудоемкость дисциплины	3 з/е, 108 ч.	108 ч.
Контактная работа – аудиторные занятия	34	34
Лекции	16	16
Семинары, практические занятия	18	18
Самостоятельная работа	74	74
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	зачёт	зачёт

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. Программные продукты для создания аналитических моделей

Среда для формирования и исследования аналитических моделей Google Collaboratory. Интегрированная среда разработки приложений на языке Python – Visual Studio Code. Способы представления данных аналитических моделей и их визуализация с использованием библиотеки Matplotlib. Работа с Google Disc. Сохранение и загрузка моделей.

Тема 2. Регрессионные модели

Прогнозирование значений наблюдаемой величины с использованием линейной и полиномиальной регрессий. Регрессор LinearRegression библиотеки Scikit-Learn. Оценка точности полученных результатов. Построение графиков.

Тема 3. Модели кластеризации

Прогнозирование принадлежности одному или нескольким классам с использованием логистической регрессии. Регрессор LogisticRegression библиотеки Scikit-Learn. Кластеризация методом k-средних. Регрессор KMeans библиотеки Scikit-Learn. Кластеризация с использованием методов дерева принятия решений и случайного леса. Регрессоры DecisionTreeClassifier и RandomForestClassifier библиотеки Scikit-Learn. Классификация методом опорных векторов. Регрессор SVM библиотеки Scikit-Learn.

5.2. Учебно-тематический план

Таблица 3

№	Наименование тем(разделов) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Всего	Аудиторная работа			Самостоятельная работа	
			Общая	Лекции	Семинары, практические занятия		

Тема 1. Программные продукты для создания аналитических моделей							
1	Изучение среды для создания и исследования аналитических моделей Google Colaboratory	12	4	2	2	8	Выполнение и защита практических работ
2	Изучение среды разработки программных моделей на языке Python – Visual Studio Code	12	4	2	2	8	Выполнение и защита практических работ
Тема 2. Регрессионные модели							
3	Формирование и исследование моделей линейной и полиномиальной регрессий.	16	4	2	2	12	Выполнение и защита практических работ
4	Оценка точности регрессионных моделей. Построение графиков	16	4	2	2	12	Выполнение и защита практических работ
Тема 3. Модели кластеризации							
5	Проектирование и реализация модели логистической регрессии	14	4	2	2	10	Выполнение и защита практических работ
6	Проектирование и реализация модели кластеризации методом k-средних	18	6	2	4	12	Выполнение и защита практических работ
7	Проектирование и реализация моделей дерева принятия решений и случайного леса	20	8	4	4	12	Выполнение и защита практических работ
В целом по дисциплине		108	34	16	18	74	Контрольная работа

5.3. Содержание семинаров, практических занятий

Таблица 4

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8, 9	Формы проведения занятий
--	--	--------------------------

Тема 1. Программные продукты для создания аналитических моделей	<ul style="list-style-type: none"> Изучение сред разработки и исследования аналитических моделей Google Colaboratory и Visual Studio Code <p>Основная литература: 1-7 Дополнительная литература: 8,9</p>	Компьютерный практикум
Тема 2. Регрессионные модели	<ul style="list-style-type: none"> Проектирование и реализация регрессионных моделей для обработки экспериментальных данных Оценка точности моделей <p>Основная литература: 1-7 Дополнительная литература: 8,9</p>	Компьютерный практикум
Тема 3. Модели кластеризации	<ul style="list-style-type: none"> Проектирование и реализация моделей кластеризации для обработки экспериментальных данных Вычисление и визуализация метрик точности <p>Основная литература: 1-7 Дополнительная литература: 8,9</p>	Компьютерный практикум

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

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

Таблица 5

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Тема 1 Программные продукты для создания аналитических моделей	<ul style="list-style-type: none"> Изучение основных функциональных возможностей библиотек Matplotlib и Plotly для визуализации результатов исследования аналитических моделей в средах Google Colaboratory и Visual Studio Code <p>Основная литература: 1,2 Дополнительная литература: 8,9</p>	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников. Подготовка к практическим работам
Тема 2. Регрессионные модели	<ul style="list-style-type: none"> Матрица ошибок confusion matrix Вычисление и способы визуализации <p>Основная литература: 1,2 Дополнительная литература: 8,9</p>	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников. Подготовка к практическим работам
Тема 3. Модели класте-	<ul style="list-style-type: none"> Метрики точности регрес- 	Изучение методических мате-

ризации	сионных моделей и моделей кластеризации – Accurasy, Precision, F1-score и AUC-ROC Основная литература: 1,2 Дополнительная литература: 8,9	риалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников. Подготовка к практическим работам
---------	---	---

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

Примерные темы для контрольной работы:

Проектирование и исследование модели аналитики для <заданной предметной области>.

Критерии балльной оценки по контрольной работе содержатся в соответствующих методических рекомендациях кафедры.

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

7.1 Перечень компетенций с указанием индикаторов их достижения в процессе освоения дисциплины

Перечень компетенций представлен в разделе 2, который характеризует перечень планируемых результатов обучения по дисциплине, соотнесенных спланируемыми результатами освоения образовательной программы.

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

Таблица 6

Компетенция	Типовые задания
ПКП-5 Способность к использованию специальных программных продуктов, применяемых для выполнения бухгалтерско-аналитических и контрольных функций в экономическом субъекте	1. Использует специальные программные продукты для выполнения бухгалтерско-аналитических и контрольных функций в экономическом субъекте Задание 1. Создайте модель кластерного анализа в среде Google Colaboratory Задание 2. Реализуйте классификацию по экспериментальным данным методом оперных векторов
	2. Демонстрирует владение специальными программными продуктами, применяемых для выполнения бухгалтерско-аналитических и контрольных функций в экономическом субъекте Задание 1. Постройте график линейной регрессии для заданной предметной области Задание 2. Вычислите метрики точности моделей кластеризации для заданной предметной области

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях кафедры.

Примерные вопросы к зачету:

1. Опишите основные этапы работы с моделями данных в Google Colaboratory и Visual Studio Code.
2. Раскройте понятие линейной регрессии. Приведите типы задач, в которых используется линейная регрессия. Опишите модель линейной регрессии, реализованную в библиотеке Scikit-Learn.
3. Раскройте понятие логистической регрессии. Приведите типы задач, в которых используется логистическая регрессия. Опишите модель логистической регрессии, реализованную в библиотеке Scikit-Learn.
4. Раскройте понятие дерева принятия решений. Приведите типы задач, в которых используется дерево принятия решений. Опишите модель дерева принятия решений, реализованную в библиотеке Scikit-Learn.
5. Раскройте понятие кластеризации. Приведите типы задач, в которых используется кластеризация. Опишите реализацию модели кластеризации в библиотеке Scikit-Learn.
6. Раскройте понятие бинарной классификации. Приведите типы задач, в которых используется бинарная классификация. Опишите реализации моделей бинарной классификации в библиотеке Scikit-Learn.
7. Раскройте понятие многоклассовой классификации. Приведите типы задач, в которых используется многоклассовая классификация. Опишите модели многоклассовой классификации, реализованные в библиотеке Scikit-Learn.
8. Раскройте основные достоинства и недостатки известных вам моделей кластеризации.
9. Раскройте понятие метрики точности модели. Приведите известные вам метрики точности и опишите реализацию их вычисления на языке Python.
10. Приведите основные библиотеки Python, используемые для обработки экспериментальных данных. Опишите их назначения и основные функциональные возможности.

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

Основная литература:

1. Протоdjаконов, А. В. Алгоритмы Data Science и их практическая реализация на Python: учебное пособие / Протоdjаконов А. В., Пылов П. А., Садовников В. Е. – М., Вологда: Инфра-Инженерия, 2022. – 392 с. – URL: <https://www.iprbookshop.ru/124000.html>
2. Маккинли, У. Python и анализ данных / Уэс Маккинли. – Саратов: Профобразование, 2019. – 482 с.
3. Чару, А. Нейронные сети и глубокое обучение. Учебный курс / Аггарвал Чару. – СПб.: Вильямс, 2020. – 752 с.
4. Гудфеллоу, Я., Бенджио, И., Курвилль, А. Глубокое обучение / пер. с англ. А. А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с. – URL: <https://e.lanbook.com/book/107901.html>

5. Хливненко, Л. В. Практика нейросетевого моделирования: учебное пособие для вузов / Л. В. Хливненко, Ф. А. Пятакович. – СПб.: Лань, 2021. – 200 с. – URL: <https://e.lanbook.com/book/173811.html>.
6. Мюллер, А, Гвидо, С. Введение в машинное обучение с помощью Python. Руководство для специалистов по обработке данных. / Андреас Мюллер, Сара Гвидо. – СПб.: ООО “Альфа-книга”, 2017. – 480 с.
7. Николаенко, С., Кадури, А., Архангельская, Е. Глубокое обучение. Погружение в мир нейронных сетей. – СПб.: Питер, 2018. – 480 с. – URL: <https://djvu.online/file/JA192EZJDBIE>

Дополнительная литература:

8. Грас, Дж. Data Science. Наука о данных с нуля / Джоэл Грас. – СПб: БХВ-Петербург, 2020. – 411 с.
9. Chollet, Francois Deep Learning with Python, Second Edition. – NY: Simon and Schuster, 2021. – 478 с.
10. Клетте, Р. Компьютерное зрение. Теория и алгоритмы: учебник / Рейнхальд Клетте; перевод с англ. А. А. Слинкина. – М.: ДМК Пресс, 2019. – 506 с. – URL: <https://e.lanbook.com/book/131691.html>

9. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины

1. Электронная библиотека Финансового университета <http://elib.fa.ru/>
2. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
3. Электронно-библиотечная система «Университетская библиотека ОН-ЛАЙН» <http://biblioclub.ru/>
4. Электронно-библиотечная система издательства «ЮРАЙТ» <https://www.urait.ru/>
5. Электронно-библиотечная система издательства «Лань» <https://e.lanbook.com/>
6. Библиотека Scikit-Learn <https://scikit-learn.org>
7. Библиотека Keras <https://keras.io>

10. Методические указания для обучающихся по освоению дисциплины

Рекомендации по освоению дисциплины приведены в «Методических рекомендациях для студентов бакалавриата по освоению дисциплин образовательных программ высшего образования», утвержденных приказом № 1040 ректора Финуниверситета от 11 мая 2021 г.

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:

1. Операционная система – Windows 8 или выше, Linux.
2. Среда разработки – Google Colaboratory и Visual Studio Code.
3. Офисный пакет – Microsoft Office или LibreOffice.

11.2 Современные профессиональные базы данных и информационные справочные системы:

№	Название рекомендуемых технических и компьютерных средств обучения	Наименование разделов и тем
1	Правовая база данных «КонсультантПлюс»	Все темы
2	Справочно-правовая система «Гарант»	Все темы
3	www.skrin.ru – Система комплексного раскрытия информации «СКРИН»	Все темы
4	http://www.iteam.ru/publications/strategy – Технологии корпоративного управления	Все темы
5	Информационная система СПАРК	Все темы
6	Информационная система Bloomberg	Все темы
7	Информационная система Thomson Reuters	Все темы

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

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

Помещения для проведения лекций, семинарских занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы.