

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

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

УТВЕРЖДАЮ

Проректор по учебной
и методической работе

_____ Е.А. Каменева

23.05.2023 г.

В. А. Иванюк

Обработка данных и моделирование в табличном редакторе

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки
01.03.02 - Прикладная математика и информатика, ОП «Анализ данных»

*Рекомендовано Ученым советом
Факультета информационных технологий и анализа больших данных
(протокол № 32 от 16.05.2023г.)*

*Одобрено Советом учебно-научного
Департамента анализа данных и машинного обучения
(протокол №3 от 26.04.2023г.)*

Москва 2023

Содержание

1. Наименование дисциплины	2
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине	2
3. Место дисциплины в структуре образовательных программ	3
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	3
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	6
5.3. Содержание семинаров, практических занятий.....	7
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине	8
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы.....	8
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю	8
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине.....	11
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	15
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	16
10. Методические указания для обучающихся по освоению дисциплины	17
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем.....	18
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.....	19

1. Наименование дисциплины

«Обработка данных и моделирование в табличном редакторе»

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
УК-4	Способность использовать прикладное программное обеспечение при решении профессиональных задач	1. Использует основные методы и средства получения, представления, хранения и обработки данных.	Знать: основные методы и средства получения, представления, хранения и обработки данных. Уметь: использовать основные методы и средства получения, представления, хранения и обработки данных.
		2. Демонстрирует владение профессиональными пакетами прикладных программ.	Знать: профессиональные пакеты прикладных программ. Уметь: владеть профессиональными пакетами прикладных программ.
		3. Выбирает необходимое прикладное программное обеспечение в зависимости от решаемой задачи.	Знать: виды прикладного программного обеспечения. Уметь: выбирать необходимое прикладное программное обеспечение в зависимости от решаемой задачи.
		4. Использует прикладное программное обеспечение для решения конкретных прикладных задач.	Знать: прикладное программное обеспечение для решения конкретных прикладных задач. Уметь: использовать прикладное программное обеспечение для решения конкретных прикладных задач.

3. Место дисциплины в структуре образовательных программ

Дисциплина «Обработка данных и моделирование в табличном редакторе» относится к Циклу математики и информатики по направлению подготовки 01.03.02 – Прикладная математика и информатика, ОП «Анализ данных».

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 5 (в часах)
Общая трудоемкость 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	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

Тема 1. Обработка данных в MS Excel

1.1 Анализ качества данных. Репрезентативность данных. Обработка некачественных данных (пропуски, выбросы). Робастность.

1.2 Методы обработки данных. Нормирование.

1.3 Меры описательной статистики.

1.4 Правило мажорантности.

Тема 2. Моделирование в MS Excel

2.1 Анализ данных на основе диаграмм и гистограмм.

2.2 Временные ряды.

2.3 Дисперсионный анализ.

2.4 Корреляционно-регрессионный анализ.

2.5 Методы прогнозирования финансовых временных рядов.

Тема 3. Нейронные сети

3.1 Обучение нейронных сетей.

3.2 Практическая реализация самоорганизующегося свёрточного ансамбля нейронной сети.

3.3 Использование нейронного анализа, для оценки кредитных рисков при одобрении кредита.

3.4 Глубокое обучение.

3.5 Прогнозирование финансовых временных рядов на основе нейронного анализа.

5.2. Учебно-тематический план

очная форма обучения

№ п/п	Наименование тем (разделов) дисциплины	Трудоемкость в часах				Самостоя- тельная работа	Формы текущего контроля успеваемости
		Всего	Контактная работа - Аудиторная работа				
			Общая, в т.ч.:	Лекции	Семинары, практиче- ские занятия		
1	Обработка дан- ных в MS Excel	29	14	4	10	15	Самостоятель- ные работы. Участие в реше- нии задач на практических занятиях.
2	Моделирование в MS Excel	38	18	6	12	20	
3	Нейронные сети	41	18	6	12	23	
	В целом по дис- циплине	108	50	16	34	58	Согласно учеб- ному плану: контрольная работа
	Итого в %		46	32	68	54	

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
Тема 1. Обработка данных в MS Excel	Объектный подход. Достоинства и недостатки алгорит- мического и объектного подходов. Организация данных. Шкалирование. Меры описательной статистики. Постро- ение гистограммы, отражающей количество элементов в карманах. Гистограмма по Стерджесу. Диаграмма рассея- ния, диаграмма размаха. Правило мажорантности. Выбро- сы. Пропуски. Подходы и методы обработки ошибок. Ме- тоды удаления выбросов. Методы удаления пустот. Рекомендуемые источники: 8.[1]-[2]	Интерактивная форма, работа на компьютере
Тема 2. Моделиро- вание в MS Excel	Анализ временных рядов. Стационарность временных ря- дов. Методы и модели прогнозирования Рекомендуемые источники: 8.[1]	Интерактивная форма, работа на компьютере
Тема 3. Нейронные сети	Множественная линейная регрессия. Нейронная сеть. Стохастический метод (обучение по Кохонену). Гради- ентный метод (обратное распространение ошибки). По- строение автокорреляционного нейронного прогноза. Рекомендуемые источники: 8.[1]	Интерактивная форма, работа на компьютере

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
Тема 1. Обработка данных в MS Excel	Виды представления данных (деревья, таблицы). Принципы сбора и обработки информации.	Работа с учебной литературой. Решение задач. Разбор вопросов по теме занятия. Выполнение домашних заданий.
Тема 2. Моделирование в MS Excel	Построение моделей прогнозирования на основе корреляционно-регрессионного анализа.	Работа с учебной литературой. Решение задач. Разбор вопросов по теме занятия. Выполнение домашних заданий.
Тема 3. Нейронные сети	Построение имитационных моделей в MS Excel.	Работа с учебной литературой. Решение задач. Разбор вопросов по теме занятия. Выполнение домашних заданий.

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

Примерные задания для подготовки к контрольной работе

1. Постройте биномиальную гистограмму в Excel.
2. Постройте гистограмму по Стерджесу в Excel.
3. Постройте диаграмму размаха в Excel.
4. Соберите недельные данные о ценах закрытия и объемах торгов по десяти акциям, обращающимся на Московской бирже с 01.01.2018 г. по сегодняшний день. Удалите строки с пропущенными и нулевыми данными в одном из столбцов.
5. Рассчитайте для двух акций недельные доходности и логдоходности, а также натуральные логарифмы цен и объемов торгов.

6. Создайте индикаторы для пары USD/RUB за период 2014–2019 гг.:
- a) Определяющий находится ли цена актива в промежутке между 40 и 55;
 - b) Вычисляющий среднее геометрическое значение последних 90 дней тор-гов;
 - c) Вычисляющий достигнутое среднее гармоническое значение;
 - d) Определяющий точки разворота вниз (верхние экстремумы) среднего геометрического за 90 дней.
7. Проведите нормирование нескольких рядов данных различными методами, постройте их общий график.
8. Рассчитайте меры вариативности для временного ряда EUR/RUB.
9. Создайте скользящий двадцатидневный индикатор направления тренда на основе среднего кубического отклонения.
10. Выберите произвольный актив Блумберг с временным рядом, включающим некризисный период (2004–2007 годы), и создайте для него индикатор кризиса, основанный на правиле трех сигм, имеющий регулирующую чувствительность и отражающий кризис 2008 года.
11. Рассчитайте доли в динамическом портфеле из трёх активов, существующем пять лет с ежегодной ребалансировкой. Постройте графики.
12. Рассчитайте доли в динамическом VaR-портфеле из двух активов, существующем год с ребалансировкой каждые 20 дней.
13. Постройте автокорреляционный нейронный прогноз для курса доллара США.
14. Постройте многофакторный прогноз на основе однослойной нейронной сети.
15. Алгоритмы обучения с учителем.
16. Ансамбль методов в статистике и обучении машин.
17. Глубокое обучение нейронных сетей.

Примерный вариант контрольной работы

Пример №

Постройте в EXCEL гистограмму распределения роста студентов вашей группы:

1. Методом опроса соберите данные о росте студентов;
2. Определите шкалу и диапазон значений;
3. Рассчитайте количество групп по формуле Стёрджеса

$$n = 1 + \lfloor \log_2 N \rfloor$$

4. Вычислите границы карманов;
5. Подсчитайте попадания в карманы;
6. Постройте диаграмму количества попаданий (гистограмму).

Пример №

1. Проведите графический анализ входных и выходных данных.

Постройте в подходящем масштабе графики прогнозируемой величины и величин, которые, предположительно влияют на неё. Рассчитайте величины их корреляции. Определите предполагаемые зависимости.

2. Определите подходящую архитектуру нейронной сети.

Можно использовать следующие виды архитектуры:

- Для явных, зависимостей с высокой (>0.75) корреляцией: 1 нейрон с несколькими входами (до 20 входов) – 21 коэффициент.
 - Для неявных, зависимостей с умеренной (>0.5) корреляцией: 1 слой до 10 нейронов с 1 входом + выходной нейрон – 21 коэффициент.
 - Для неявных, зависимостей с низкой (<0.5) корреляцией: 2 слоя до 3-х нейронов с 3 входами + выходной нейрон – 22 коэффициента.
3. Задайте формулу сети, формулу ошибки сети и определите обучающую выборку.
 4. Рассчитайте значения границ входных и выходных коэффициентов.

Для сетей с тангенциальной сигма-функцией:

- границы входных коэффициентов: $\left[\frac{-3}{x_{max}}, \frac{3}{x_{max}}\right]$
- границы выходных коэффициентов: $[[x_{min}], [6x_{max}]]$

Для сетей с рациональной сигма-функцией для $\alpha = 1$:

- границы входных коэффициентов: $\left[\frac{-6}{x_{max}}, \frac{6}{x_{max}}\right]$
- границы выходных коэффициентов: $[-12x_{max}, 12x_{max}]$

5. Задайте границы и проведите обучение сети эволюционным алгоритмом.
6. Получите от обученной сети прогноз для текущих данных.

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях Департамента анализа данных и машинного обучения.

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	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Примеры типовых контрольных заданий

УК-4 Способность использовать прикладное программное обеспечение при решении профессиональных задач	1. Использует основные методы и средства получения, представления, хранения и обработки данных.	Знать: основные методы и средства получения, представления, хранения и обработки данных. Уметь: использовать основные методы и средства получения, представления, хранения и обработки данных.	Изучите различные типы данных и способы их представления. Сделайте импорт данных в табличный редактор.
	2. Демонстрирует владение профессиональными пакетами прикладных программ.	Знать: профессиональные пакеты прикладных программ. Уметь: владеть профессиональными пакетами прикладных программ.	С помощью табличного редактора «Мой офис» рассчитайте коэффициенты в нейронной сети. С помощью табличного редактора «Мой офис» постройте модель на основе корреляционно-регрессионного анализа
	3. Выбирает необходимое прикладное программное обеспечение в зависимости от решаемой задачи.	Знать: виды прикладного программного обеспечения. Уметь: выбирать необходимое прикладное программное обеспечение в зависимости от решаемой задачи.	В табличном редакторе «Мой офис» постройте графики и диаграммы цен закрытия различных активов финансового рынка. На основе статистического пакета Statista (http://library.fa.ru/resource.asp?id=809) и табличного редактора «Мой офис» постройте модель прогнозирования курса акций.

	4. Использует прикладное программное обеспечение для решения конкретных прикладных задач.	Знать: прикладное программное обеспечение для решения конкретных прикладных задач. Уметь: использовать прикладное программное обеспечение для решения конкретных прикладных задач.	В табличном редакторе «Мой офис» постройте инвестиционный портфель Фишера В табличном редакторе «Мой офис» решите задачу коммивояжера
--	---	---	--

Примерные вопросы для подготовки к зачету

1. Виды диаграмм. Диаграмма размаха. Диаграмма рассеяния
2. Анализ качества данных. Репрезентативность данных.
3. Робастность
4. Подходы и методы обработки ошибок
5. Индикаторный метод обработки ошибок
6. Ситуативный «ad-hoc» метод обработки ошибок
7. Метод исключения для обработки ошибок
8. Метод исключения по квартилям
9. Статистическая подстановка
10. Компрессионная подстановка
11. Регрессионная подстановка
12. Прогностическая подстановка
13. Условное форматирование данных
14. Основные статистические показатели
15. Инструмент «Описательная статистика» в MS Excel
16. Нормализация данных

17. Инструмент «Анализ данных» в MS Excel
18. Инструмент «Поиск решения» в MS Excel
19. Корреляционно-регрессионный анализ
20. Дисперсионный анализ
21. Методы обучения нейронных сетей
22. Ансамбль методов в статистике и обучении машин
23. Использование нейронного анализа, для оценки кредитных рисков при одобрении кредита

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 анализ данных в Microsoft Excel : учебник / В. И. Соловьев. — Москва : КноРус, 2021. — 497 с. — ЭБС BOOK.ru. — URL: <https://book.ru/book/938856> (дата обращения: 16.05.2023). — Текст : электронный.
2. Зададаев, С. А. Математика на языке R: учебник / С. А. Зададаев; Финансовый университет при Правительстве РФ, Департамент анализа данных, принятия решений и финансовых технологий. — Москва : Прометей, 2018. — 324 с. — Текст : непосредственный. - То же. — ЭБС Университетская библиотека онлайн. — URL: <https://biblioclub.ru/index.php?page=book&id=494941> (дата обращения: 16.05.2023). — Текст : электронный.

Дополнительная литература:

3. Гармаш, А. Н. Экономико-математические методы и прикладные модели : учебник для бакалавриата и магистратуры / А. Н. Гармаш,

И. В. Орлова, В. В. Федосеев ; под редакцией В. В. Федосеева. — 4-е изд., перераб. и доп. — Москва : Юрайт, 2022. — 328 с. — (Бакалавр и магистр. Академический курс). — ЭБС Юрайт. — URL: <https://urait.ru/bcode/507819> (дата обращения: 16.05.2023). - Текст: электронный.

4. Глотова, М. Ю. Математическая обработка информации : учебник и практикум для академического бакалавриата:/ М. Ю. Глотова , Е. А. Самохвалова. - Москва: Юрайт, 2023. - 347 с. - Бакалавр. Академический курс – ЭБС Юрайт. - URL: <https://www.biblio-online.ru/bcode/511062> (дата обращения: 16.05.2023). – Текст : электронный.

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

1. Образовательный портал Финансового университета. Ссылка на ЭУК
<https://campus.fa.ru/course/view.php?id=10233>
<https://campus.fa.ru/course/view.php?id=20774>
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.znanium.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. Деловая онлайн-библиотека Alpina Digital <http://lib.alpinadigital.ru/>
10. Электронная библиотека Издательского дома «Гребенников»
<https://grebennikon.ru/>
11. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
12. Национальная электронная библиотека <http://нэб.рф/>
13. <http://rts.micex.ru/>
14. <http://www.gks.ru/>
15. <http://wwwcbr.ru/>
16. Яндекс-практикум «Excel для работы» <https://practicum.yandex.ru/excel-for-work>

10. Методические указания для обучающихся по освоению дисциплины

Самостоятельная работа студентов проходит аудиторно и внеаудиторно. Организации самостоятельной работы служит учебно-тематический план изучения дисциплины. В этом плане указана тематика лекций, практических занятий, вопросы и задания для самостоятельного изучения.

При подготовке к лекции целесообразно предварительно познакомиться с ее содержанием по рекомендованным пособиям и выделить наиболее трудные вопросы. Во время лекций следует конспектировать содержание лекции. После занятий следует провести работу с конспектом: отредактировать записи, оформить конспект.

При подготовке к практическому занятию необходимо повторить или, если это требуется, изучить соответствующий теоретический материал. Практические занятия проходят, как правило, в интерактивной форме и преподаватель учитывает активность студентов, направленную на решение предложенных задач и в поиске ответов на вопросы.

Домашние задания следует выполнять регулярно при подготовке к практическим занятиям. Контроль за выполнением домашних заданий осуществляется в ходе практических занятий и выборочного собеседования

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. Пакет офисных программ. Табличный редактор «Мой офис»: <https://myoffice.ru>.
2. Антивирус Kaspersky

11.2. Современные профессиональные базы данных и информационные справочные системы:

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

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

11.4. Microsoft Azure

11.5. Hadoop

11.6. HDInsight

11.7. Spark

11.8. Дистрибутив языка Python 3.4 (или более поздней версии) Anaconda 3

11.9. Программа Deductor Academic версии 5.3 или выше.

11.10. Платформа Statista <http://library.fa.ru/resource.asp?id=809>

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

Компьютерный класс, оснащённый системой динамического проецирования.