

АВТНОМНАЯ НЕКОММЕРЧЕСКАЯ ОРГАНИЗАЦИЯ
ВЫСШЕГО ОБРАЗОВАНИЯ
«СЕВЕРО-КАВКАЗСКИЙ СОЦИАЛЬНЫЙ ИНСТИТУТ»



Утверждаю

Проректор по УР и КО

Ю.Е. Леденева

« 09 » 2023г.


РАБОЧАЯ ПРОГРАММА КУРСА
«Анализ и визуализация данных»
(наименование дисциплины)


Дополнительная профессиональная программа «Анализ и визуализация данных»


Форма обучения

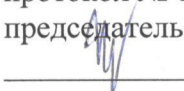
очная

(очная, очно-заочная, заочная)

Разработана
к.э.н., доцент., зав. кафедрой
информационных систем
 А.Ю. Орлова

Согласована
Начальник УМУ
 Д.В.Гришин

Рекомендована
на заседании кафедры ИС
от «31» августа 2023 г.
протокол № 1
зав. кафедрой  А.Ю. Орлова

Одобрена
на заседании учебно-методической
комиссии факультета
от «31» августа 2023 г.
протокол № 1
председатель УМК
 Ж.В. Игнатенко

Ставрополь, 2023 г.

Содержание

1. Цели освоения курса	3
2. Требования к результатам освоения содержания курса	3
3. Объем курса и виды учебной работы	4
4. Содержание и структура курса	4
4.1. Содержание курса.....	4
4.2. Структура курса.....	5
4.3. Практические занятия и семинары	6
4.4. Лабораторные работы	6
4.5. Курсовой проект (курсовая работа, расчетно-графическая работа, реферат, контрольная работа)	6
4.6. Внеаудиторная (самостоятельная) работа.....	6
5. Образовательные технологии.....	7
6. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации	7
7. Учебно-методическое и информационное обеспечение курса	8
7.1. Основная литература.....	8
7.2. Дополнительная литература.....	8
7.3. Программное обеспечение.....	9
7.4. Базы данных, информационно-справочные и поисковые системы, Интернет-ресурсы	9
7.5. Методические указания по видам занятий	9
8. Материально-техническое обеспечение курса	11

1. ЦЕЛИ ОСВОЕНИЯ КУРСА

Целью курса «Анализ и визуализация данных» является формирование компетенций в области информационных технологий в части машинного обучения и анализа данных.

Задачи курса:

1. Освоение современных технологий цифровизации.
2. Ознакомление с инновационными решениями использования больших данных.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ СОДЕРЖАНИЯ КУРСА

Процесс изучения курса направлен на формирование/совершенствование следующих компетенций:

Способен подготавливать данные для проведения аналитических работ по исследованию больших данных.

Способен проводить аналитические исследования с применением технологий больших данных в соответствии с требованиями заказчика

Слушатель должен знать:

- теоретические и прикладные основы анализа больших данных;
- современные методы и инструментальные средства анализа больших данных;
- технологии хранения и обработки больших данных в организации: базы данных, хранилища данных;
- содержание и последовательность выполнения этапов аналитического проекта;
- технологии анализа данных;
- методы и модели классификации;
- анализ данных в реальном времени.

Слушатель должен уметь:

- использовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников, в том числе в режиме реального времени;
- производить очистку данных для проведения аналитических работ;
- проводить интеграцию и преобразование больших объемов данных;
- использовать имеющуюся у исполнителя методологическую и технологическую инфраструктуру анализа больших данных для выполнения аналитических работ;
- программировать на языке DAX, ориентированном на работу с большими данными.

Слушатель должен владеть навыками:

- определения источников больших данных для анализа, идентификация внешних и внутренних источников данных для проведения аналитических работ;
- получения и фильтрации больших объемов данных из гетерогенных источников;
- извлечения, проверки и очистки больших объемов данных из гетерогенных источников;
- выбора средств представления результатов аналитики больших данных;
- подготовки отчета по результатам аналитических работ с использованием технологий больших данных.

3. ОБЪЕМ КУРСА И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость курса составляет 72 академических часа.

Вид учебной работы	Всего часов	Периоды обучения
		14-19 календ. дней
Аудиторные занятия (всего)	40	40
в том числе:		
Лекции (Л)	16	16
Практические занятия (ПЗ)		-
Семинары (С)		-
Лабораторные работы (ЛР)	24	24
Внеаудиторные занятия (самостоятельная работа) (СР)	32	30
в том числе:		
Курсовой проект (работа)		-
Расчетно-графические работы		-
Контрольная работа		-
Реферат		-
Самоподготовка (самостоятельное изучение разделов, проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к лабораторным и практическим занятиям, коллоквиумам, рубежному контролю и т.д.)	30	30
Вид промежуточной аттестации	2	зачет
Общая трудоемкость, час	72	72

4. СОДЕРЖАНИЕ И СТРУКТУРА КУРСА

4.1. Содержание курса

№ раздела (темы)	Наименование раздела (темы)	Содержание раздела (темы)
1.	Технологии интеллектуального анализа данных	1. Знания и данные - сходства и отличия. 2. Принятие решений по прецедентам и моделям. 3. Анализ данных и анализ знаний. 4. Понятие о закономерностях.
2.	Бизнес-аналитика в прикладных статистических пакетах	1. Подготовка данных к анализу в SPSS. 2. Описательная статистика. Средние. Таблицы. 3. OLAP-кубы в SPSS. Модели анализа данных. 4. Т-критерии. Факторный, кластерный и регрессионный анализ в SPSS.
3.	Статистические основы анализа данных	1. Роль анализа данных в современном мире 2. Научные исследования 3. Построение системы анализа данных 4. Источники данных 5. Сбор данных 6. Подготовка данных Генеральная совокупность и выборка 7. Типы переменных. Количественные и номинативные переменные

		8. Меры центральной тенденции 9. Меры изменчивости 10. Квартили распределения и график box-plot1. 11. Нормальное распределение 12. Центральная предельная теорема
4.	Основы NumPy: массивы и векторные вычисления	1. Назначение пакета NumPy 2. Сравнение NumPy с MATLAB 3. Массивы NumPy 4. Векторизация вычислений в NumPy
5.	Основы работы с Pandas	1. Введение в структуры данных Pandas 2. Объект Series 3. Объект DataFrame 4. Индексные объекты 5. Базовая функциональность Pandas 6. Переиндексация DataFrame 7. Удаление элементов из оси 8. Доступ по индексу, выборка и фильтрация
6.	Переформатирование данных в Pandas	1. Комбинирование и слияние наборов данных 2. Слияние объектов DataFrame как в базах данных 3. Слияние по индексу 4. Конкатенация вдоль оси 5. Комбинирование перекрывающихся данных
7.	Визуализация данных в MS Power BI	1. Подготовка данных в Power BI Desktop 2. Загрузка данных в Power BI Desktop
8.	Анализ и прогнозирование с использованием языка DAX	1. Создание вычислений DAX в Power BI Desktop 2. Создание информационной панели Power BI 3. Выполнение анализа данных в Power BI Desktop

4.2. Структура курса

№ раздела (темы)	Наименование раздела (темы)	Количество часов				
		Всего	Л	ПЗ (С)	ЛР	СР
Тема 1	Технологии интеллектуального анализа данных	8	2		2	4
Тема 2	Бизнес-аналитика в прикладных статистических пакетах	10	2		4	4
Тема 3	Статистические основы анализа данных	8	2		2	4
Тема 4	Основы NumPy: массивы и векторные вычисления	10	2		4	4
Тема 5	Основы работы с Pandas	8	2		2	4
Тема 6	Переформатирование данных в Pandas	8	2		4	2
Тема 7	Визуализация данных в MS Power BI	8	2		2	4
Тема 8	Анализ и прогнозирование с использованием языка DAX	10	2		4	4

	Итоговая аттестация	2				
	Общая трудоемкость	72	16		24	30

4.3. Практические занятия и семинары

Не предусмотрено

4.4. Лабораторные работы

№ п/п	№ раздела (темы)	Наименование лабораторных работ	Количество часов
1.	1.	Лабораторная работа «Технологии бизнес-аналитики: olap-технологии»	2
2.	2.	Лабораторная работа «Бизнес-аналитика в прикладных статистических пакетах (spss)»	4
3.	3.	Лабораторная работа «Исследование средств разработки программ Jupyter Notebook»	2
4.	4.	Лабораторная работа «Исследование методов обработки данных с применением массивов NumPy»	4
5.	5.	Лабораторная работа «Исследование базовых функций Pandas»	2
6.	6.	Лабораторная работа «Исследование методов преобразования данных в Pandas»	4
7.	7.	Лабораторная работа «Создание DAX вычислений в Power BI Desktop»	2
8.	8.	Лабораторная работа «Выполнение анализа данных в Power BI Desktop»	4

4.5. Курсовой проект (курсовая работа, расчетно-графическая работа, реферат, контрольная работа)

Не предусмотрено

4.6. Внеаудиторная (самостоятельная) работа

Виды самостоятельной работы	Количество часов
Проработка текущего материала по конспектам лекций и рекомендуемой литературе по теме «Технологии интеллектуального анализа данных»	4
Проработка текущего материала по конспектам лекций и рекомендуемой литературе по теме «Бизнес-аналитика в прикладных статистических пакетах»	4
Проработка текущего материала по конспектам лекций и рекомендуемой литературе по теме «Статистические основы анализа данных»	4
Проработка текущего материала по конспектам лекций и рекомендуемой литературе по теме «Основы NumPy: массивы и векторные вычисления»	4
Проработка текущего материала по конспектам лекций и рекомендуемой литературе по теме «Основы работы с Pandas»	4
Проработка текущего материала по конспектам лекций и рекомендуемой литературе по теме «Переформатирование данных в Pandas»	2
Проработка текущего материала по конспектам лекций и рекомендуемой литературе по теме «Визуализация данных в MS Power BI»	4
Проработка текущего материала по конспектам лекций и рекомендуемой литературе по теме «Анализ и прогнозирование с использованием языка DAX»	4

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Информационные технологии, используемые при осуществлении образовательного процесса по курсу:

При чтении лекций используется компьютерная техника для демонстрации слайдов с помощью программного приложения Microsoft Power Point. На лабораторных занятиях слушатели представляют результаты выполнения самостоятельной работы, подготовленные с помощью программного продукта Microsoft Word. При выполнении лабораторных работ на практических занятиях, студентами используется программное обеспечение: MS Word, Google Chrome, Visual Studio, MS Power BI на удаленном сервере.

6. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Промежуточная аттестация проводится в форме зачета.

1. Перечень типовых заданий к лабораторным работам

1. С использованием библиотеки Pandas на языке Python создайте код для обработки данных, загруженных из Localhost.

2. Загрузите данные в Power BI Desktop с Localhost (DimEmployee; DimEmployeeSalesTerritory; DimProduct; DimReseller; DimSalesTerritory; FactResellerSales) и создайте на языке DAX меру, которая бы считала общее количество проданного товара по клиентам.

3. Загрузите данные в Power BI Desktop с Localhost (DimEmployee; DimEmployeeSalesTerritory; DimProduct; DimReseller; DimSalesTerritory; FactResellerSales) и в рабочей области Данные сделайте связи таким образом, чтобы данные по всем продавцам и товарам можно было использовать для анализа данных.

2. Перечень заданий к самостоятельной работе слушателя

1. Проанализируйте способы загрузки данных в Power BI Desktop.
2. Изучите способы загрузки данных при работе с библиотекой Pandas.
3. Изучите математические методы и модели, которые необходимо использовать при прогнозировании в Power BI Desktop.
4. Изучите возможности OLAP-кубов в SPSS. Виды моделей анализа данных.

3. Типовые вопросы и задания к зачету (итоговой аттестации)

Итоговая аттестация включает зачёт. Зачет проводится после выполнения учебного плана программы в части установленного объема различных видов учебной деятельности. Слушатель допускается к сдаче зачета после защиты всех отчетов по лабораторным работам, установленным учебным планом. Слушатель отвечает на 2 вопроса из перечня.

Вопросы:

1. Знания и данные - сходства и отличия.
2. Принятие решений по прецедентам и моделям.
3. Анализ данных и анализ знаний.
4. Понятие о закономерностях.
5. Подготовка данных к анализу в SPSS.
6. Описательная статистика. Средние. Таблицы.
7. OLAP-кубы в SPSS. Модели анализа данных.
8. Т-критерии. Факторный, кластерный и регрессионный анализ в SPSS.
9. Роль анализа данных в современном мире

10. Научные исследования
11. Построение системы анализа данных
12. Источники данных
13. Сбор данных
14. Подготовка данных Генеральная совокупность и выборка
15. Типы переменных. Количественные и номинативные переменные
16. Меры центральной тенденции
17. Меры изменчивости
18. Квартили распределения и график box-plot1.
19. Нормальное распределение
20. Центральная предельная теорема
21. Назначение пакета NumPy
22. Сравнение NumPy с MATLAB
23. Массивы NumPy
24. Векторизация вычислений в NumPy
25. Введение в структуры данных Pandas
26. Объект Series
27. Объект DataFrame
28. Индексные объекты
29. Базовая функциональность Pandas
30. Переиндексация DataFrame
31. Удаление элементов из оси
32. Доступ по индексу, выборка и фильтрация
33. Комбинирование и слияние наборов данных
34. Слияние объектов DataFrame как в базах данных
35. Слияние по индексу
36. Конкатенация вдоль оси
37. Комбинирование перекрывающихся данных
38. Подготовка данных в Power BI Desktop
39. Загрузка данных в Power BI Desktop
40. Создание вычислений DAX в Power BI Desktop
41. Создание информационной панели Power BI
42. Выполнение анализа данных в Power BI Desktop

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ КУРСА

7.1. Основная литература

1. Бизнес-статистика : учебник и практикум для вузов / И. И. Елисеева [и др.] ; под редакцией И. И. Елисеевой. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2022. — 444 с. — (Высшее образование). — ISBN 978-5-534-14822-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/490172>.

2. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2022. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/489100>

7.2. Дополнительная литература

1. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2022. — 174 с. — (Высшее образование). — ISBN 978-5-

7.3. Программное обеспечение

MicrosoftOffice или Яндекс 360, GoogleChrome или Яндекс браузер, MicrosoftPowerBI.

7.4. Базы данных, информационно-справочные и поисковые системы, Интернет-ресурсы

1. База данных IT специалиста» <http://info-comp.ru/>
2. База данных веб-технологий <http://www.php.su>
3. База данных программного обеспечения Oracle <https://www.oracle.com/ru/index.html>

7.5. Методические указания по видам занятий

Методические указания для подготовки к лекции

Аудиторные занятия планируются в рамках такой образовательной технологии, как проблемно-ориентированный подход с учетом профессиональных и личностных особенностей обучающихся. Это позволяет учитывать исходный уровень знаний обучающихся, а также существующие технические возможности обучения.

Методологической основой преподавания курса являются научность и объективность.

Лекция является первым шагом подготовки обучающихся к практическим занятиям. Проблемы, поставленные в ней, на практическом занятии приобретают конкретное выражение и решение.

Преподаватель на вводной лекции определяет структуру курса, поясняет цели и задачи изучения курса, формулирует основные вопросы и требования к результатам освоения. При проведении лекций, как правило, выделяются основные понятия и определения. При описании закономерностей обращается особое внимание на сравнительный анализ конкретных примеров.

Рекомендуя литературу для самостоятельного изучения, преподаватель поясняет, каким образом максимально использовать возможности, предлагаемые библиотекой СКСИ, в том числе ее электронными ресурсами, а также сделает акцент на привлечение ресурсов сети Интернет и профессиональных баз данных для изучения практики.

Выбор методов и форм обучения по курсу определяется:

- общими целями образования, воспитания, развития и психологической подготовки обучающихся;
- особенностями учебного курса и спецификой его требований к отбору дидактических методов;
- целями, задачами и содержанием материала конкретного занятия;
- временем, отведенным на изучение того или иного материала;
- уровнем подготовленности обучающихся;
- уровнем материальной оснащенности, наличием оборудования, наглядных пособий, технических средств.

Лекции дают обучающимся систематизированные знания по курсу, концентрируют их внимание на наиболее сложных и важных вопросах.

Лекции обычно излагаются в традиционном или в проблемном стиле (интерактивном). Интерактивный стиль позволяет стимулировать активную познавательную деятельность обучающихся и их интерес к дисциплине, формировать творческое мышление, прибегать к противопоставлениям и сравнениям, делать обобщения, активизировать внимание обучающихся путем постановки проблемных вопросов, поощрять дискуссию. Во время лекционных занятий рекомендуется вести

конспектирование учебного материала, обращать внимание на формулировки и категории, раскрывающие суть того или иного явления или процессов, выводы и практические рекомендации.

В конце лекции делаются выводы и определяются задачи на самостоятельную работу. Во время лекционных занятий рекомендуется вести конспектирование учебного материала, обращать внимание на формулировки и категории, раскрывающие суть того или иного явления или процессов, научные выводы и практические рекомендации. В случае недопонимания какой-либо части предмета следует задать вопрос в установленном порядке преподавателю.

Конспект – это систематизированное, логичное изложение материала источника. Различаются четыре типа конспектов:

План-конспект – это развернутый детализированный план, в котором достаточно подробные записи приводятся по тем пунктам плана, которые нуждаются в пояснении.

Текстуальный конспект – это воспроизведение наиболее важных положений и фактов источника.

Свободный конспект – это четко и кратко сформулированные (изложенные) основные положения в результате глубокого осмысливания материала. В нем могут присутствовать выписки, цитаты, тезисы; часть материала может быть представлена планом.

Тематический конспект – составляется на основе изучения ряда источников и дает более или менее исчерпывающий ответ по какой-то схеме (вопросу).

Подготовленный конспект и рекомендуемая литература используются при подготовке к и практическим занятиям. Подготовка сводится к внимательному прочтению учебного материала, к выводу с карандашом в руках всех утверждений, к решению примеров, задач, к ответам на вопросы. Примеры, задачи, вопросы по теме являются средством самоконтроля.

Методические указания по подготовке к лабораторным работам

Целью лабораторных работ является углубление и закрепление теоретических знаний, полученных обучающимися на лекциях и в процессе самостоятельного изучения учебного материала, а, следовательно, формирование у них определенных умений и навыков.

В ходе подготовки к лабораторным работам необходимо прочитать конспект лекции, изучить основную литературу, ознакомиться с дополнительной литературой, выполнить выданные преподавателем задания. При этом учесть рекомендации преподавателя и требования программы. Дорабатывать свой конспект лекции, делая в нем соответствующие записи из литературы. Желательно при подготовке к лабораторным работам по курсу одновременно использовать несколько источников, раскрывающих заданные вопросы.

Методические указания для выполнения самостоятельной работы

Самостоятельная работа обучающихся заключается:

В целях наиболее эффективного изучения дисциплины подготовлены различные задания, различающиеся по преследуемым целям.

Задания самостоятельной внеаудиторной работы обучающихся заключаются в продолжении изучения теоретического материала дисциплины и в развитии навыков самостоятельного анализа литературы.

I. Самостоятельное теоретическое обучение предполагает освоение слушателем во внеаудиторное время рекомендуемой преподавателем основной и дополнительной литературы. С этой целью обучающимся рекомендуется постоянно знакомиться с классическими теоретическими источниками по темам дисциплины, а также с новинками литературы, статьями в периодических изданиях, справочных правовых системах.

Для лучшего понимания материала целесообразно осуществлять его конспектирование с возможным последующим его обсуждением на лабораторных работах. Формы конспектирования материала могут быть различными:

1) обобщение – при подготовке такого конспекта слушателем осуществляется анализ и обобщение всех существующих в доктрине подходов по выбранному дискуссионному вопросу раздела, в том числе, дореволюционных ученых, ученых советского и современного периода развития. Основная задача обучающегося заключается не только в изложении точек зрения по исследуемому вопросу, но и в выражении собственной позиции с соответствующим развернутым теоретическим обоснованием.

2) рецензия – при подготовке такого конспекта студентом осуществляется рецензирование выбранного источника по изучаемому дискуссионному вопросу, чаще всего, статьи и периодическом издании, тезисов выступления на конференции либо главы из монографии. Для этого слушателем дается оценка содержанию соответствующего источника по следующим параметрам: актуальность выбранной темы, в том числе убедительность обоснования актуальности исследования автором; соответствие содержания работы ее названию; логичность, системность и аргументированность (убедительность) выводов автора; научная добросовестность (наличие ссылок на использованные источники, самостоятельность исследования, отсутствие фактов недобросовестных заимствований текстов, идей и т.п.); научная новизна и др.

Формой контроля за самостоятельным теоретическим обучением является защита отчета по выполненным лабораторным работам.

II. Ключевую роль в планировании индивидуальной траектории обучения по дисциплине играет *опережающая самостоятельная работа* (ОПС). Такой тип обучения предлагается в замену традиционной репродуктивной самостоятельной работе (самостоятельное повторение учебного материала и рассмотренных на занятиях алгоритмов действий, выполнение по ним аналогичных заданий). ОПС предполагает следующие виды самостоятельных работ:

познавательно-поисковая самостоятельная работа, предполагающая подготовку докладов, выступлений на практических занятиях, подбор литературы по конкретной проблеме, написание рефератов и др.;

творческая самостоятельная работа, к которой можно отнести выполнение специальных творческих и нестандартных заданий. Задача преподавателя на этапе планирования самостоятельной работы – организовать ее таким образом, чтобы максимально учесть индивидуальные способности каждого обучающегося, развить в нем познавательную потребность и готовность к выполнению самостоятельных работ все более высокого уровня. Слушатели, приступая к изучению тем, должны применить свои навыки работы с библиографическими источниками и рекомендуемой литературой, умение четко формулировать свою собственную точку зрения и навыки ведения научных дискуссий. Все подготовленные и представленные тексты должны являться результатом самостоятельной информационно-аналитической работы обучающихся. На их основе студенты готовят материалы для выступлений в ходе практических занятий.

8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ КУРСА

Для реализации дисциплины требуется следующее материально-техническое обеспечение:

- для проведения занятий лекционного типа – аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории: учебная мебель, экран, проектор, ноутбук.

- для проведения лабораторных работ – аудитория, оснащенная компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в

электронную информационно-образовательную среду организации, а также с установленным ПО, согласно п. 5.

- для проведения текущего контроля и промежуточной аттестации – аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории: учебная мебель, экран, проектор, ноутбук.

- для самостоятельной работы обучающихся – аудитория, оснащенная компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации.