Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Максимов Алексей Борисович
Должность: директор департамента по образовательной политике
Дата подписания: 02.11.2023 11:08:12
Уникальный программный ключ:
8db180d1a3f02ac9e60521a5672742735c18b1d6

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное автономное образовательное учреждение
высшего образования
«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
Факультет информационных технологий

УТВЕРЖДАЮ
Декан факультета
«Информационные технологии»
_____/Д.Г.Демидов/
«10» мая 2022

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Анализ данных»

Направление подготовки/специальность
09.03.02 Информационные системы и технологии

Профиль/специализация
«Автоматизированные системы обработки информации и управления»

Квалификация
Бакалавр

Формы обучения
Очная

Москва, 2022 г.

Разработчик(и):
д.т.н., профессор
каф. «Информатика и
информационные технологии», / Д.И. Попов /

Согласовано:
Заведующий кафедрой
«Информатика и информационные технологии»,
к.т.н. / Е.В. Булатников /

Содержание

1. Цели, задачи и планируемые результаты обучения по дисциплине....4
2. Место дисциплины в структуре образовательной программы....4
3. Структура и содержание дисциплины....4
3.1. Виды учебной работы и трудоемкость(по формам обучения)....5
3.2. Тематический план изучения дисциплины(по формам обучения)....6
3.3. Содержание дисциплины....6
3.4. Тематика семинарских/практических и лабораторных занятий....7
3.5 Тематика курсовых проектов (курсовых работ)....7
4. Учебно-методическое и информационное обеспечение....7
4.1. Нормативные документы и ГОСТы....7
4.2. Основная литература....7
4.3. Дополнительная литература....7
4.4. Электронные образовательные ресурсы....8
4.5 Лицензионное и свободно распространяемое программное обеспечение....9
4.6 Современные профессиональные..... базы данных иинформационные справочные системы 9
5. Материально-техническое обеспечение....9
6. Методические рекомендации....9
6.1. Методические рекомендации для преподавателя по организацииобучения....9
6.2. Методические указания для обучающихся по освоению дисциплины....9
7. Фонд оценочных средств....10
7.1 Методы контроля и оценивания результатов обучения....10
7.2. Шкала и критерии оценивания результатов обучения....10
7.3. Оценочные средства....11

1. Цели, задачи и планируемые результаты обучения по дисциплине

В результате изучения дисциплины учащийся получит знания, умения и навыки, связанные с применением в своей профессиональной деятельности современных технологий в области анализа данных, научится решать задачи сбора, организации, хранения и анализа данных. На практике получит навыки разработки алгоритмов, программных модулей и моделей для анализа данных. Полученные знания могут быть использованы в профессиональной деятельности для анализа данных в любых предметных областях

Код компетенции	**В результате освоения образовательной программы обучающийся должен обладать**	**Перечень планируемых результатов обучения по дисциплине**
ПК-5	Способен оптимизировать функционирование БД	ИПК-5.1. Знает способы оптимизации запросов, индексов, хранимых процедур ИПК-5.2. Умеет выявлять проблемные ситуации в работе БД по обработке информации ИПК-5.3. Имеет навыки разработки и применения программного обеспечения для мониторинга работы БД по обработке информации
ПК-6	Способен предотвращать потери и повреждения данных	ИПК-6.1. Знает способы и методы резервного копирования и восстановления данных ИПК-6.2. Умеет производить резервное копирование и восстановление данных ИПК-6.3. Имеет навыки применения программного обеспечения для резервного копирования и восстановления данных

2. Место дисциплины в структуре образовательной программы

Дисциплина «Анализ данных» относится к части, формируемой участниками образовательных отношений Блока 1 «Дисциплины (модули)» учебного плана программы бакалавриата по направлению 09.03.02 «Информационные системы и технологии».

- Дисциплина взаимосвязана логически и содержательно-методически со следующими дисциплинами и практиками ОПОП:
- Распознавание образов в информационных и автоматизированных системах;
- Теория принятия решений.
- Теория систем и системный анализ;
- Преддипломная практика;
- Государственная итоговая аттестация (выполнение и защита ВКР).

3. Структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 3 зачетных единиц, т.е. 108

академических часов.

3.1. Виды учебной работы и трудоемкость(по формам обучения)

3.1.1 Очная форма обучения

№ п/п	Вид учебной работы	Количество часов	Семестры
			6
1	**Аудиторные занятия**	**54**	**54**
	В том числе:	-	-
1.1	Лекции	18	18
1.2	Семинарские/практические занятия	-	-
1.3	Лабораторные занятия	36	36
2	**Самостоятельная работа**	**54**	**54**
	В том числе:	-	-
2.1	Подготовка и выполнение лабораторных работ	-	-
3	**Курсовое проектирование**	**-**	**-**
3	**Промежуточная аттестация**	**-**	**-**
	Зачет/экзамен/диф.зачет	-	зачет
	Итого:	**108**	**108**

3.2. Тематический план изучения дисциплины(по формам обучения)

3.2.1 Очная форма обучения

№ п/п	Разделы/темы дисциплины	Трудоемкость, час					
		Всего	Аудиторная работа				Самостоятельная работа
			Лекции	Семинарские/практические занятия	Лабораторные занятия	Практическая подготовка	
1.	Основные понятия и категории анализа данных	4	2				2
2.	Использование языка R в анализе данных	4	2				2
3.	Лабораторная работа №1. Основы языка R для анализа данных	10			6		4
4.	Особенности языка R, основные библиотеки и фреймворки	4	2				2
5.	Лабораторная работа №2. Работа с пакетами CARET и BORUTA на языке R	10			6		4
6.	Использования языка Python для анализа данных	6	2				4
7.	Лабораторная работа №3. Анализ данных с помощью библиотек языка Python	10			6		4
8.	Разведочный анализ данных. Оценки центрального положения, вариабельности и распределения данных	6	2				4
9.	Обследование двоичных и категориальных данных	6	2				4
10.	Лабораторная работа №4. Способы подготовки исходных данных	10			6		4
11.	Корреляционный и регрессионный анализ данных	6	2				4
12.	Основы дисперсионного анализа	6	2				4
13.	Лабораторная работа №5. Кластерный и регрессионный анализ данных с использованием языка R	10			6		4
14.	Повторный отбор в анализе данных – бутстрап	6	2				4
15.	Лабораторная работа №6. Реализация бутстрапа на языках R и Python	10			6		4
Итог:		**108**	**18**		**36**		**54**

3.3. Содержание дисциплины

Основные понятия и категории анализа данных

Использование языка R в анализе данных

Особенности языка R, основные библиотеки и фреймворки

Использования языка Python для анализа данных
Разведочный анализ данных. Оценки центрального положения, вариабельности и распределения данных
Обследование двоичных и категориальных данных
Корреляционный и регрессионный анализ данных
Основы дисперсионного анализа
Повторный отбор в анализе данных – бутстрап

3.4. Тематика семинарских/практических и лабораторных занятий

4. 3.4.1 Семинарские/практические занятие
5. Не предусмотрены.
6.
7. 3.4.2 Лабораторные занятия
8. Лабораторная работа №1. Основы языка R для анализа данных
9. Лабораторная работа №2. Работа с пакетами CARET и BORUTA на языке R
10. Лабораторная работа №3. Анализ данных с помощью библиотек языка Python
11. Лабораторная работа №4. Способы подготовки исходных данных
12. Лабораторная работа №5. Кластерный и регрессионный анализ данных с использованием языка R
13. Лабораторная работа №6. Реализация бутстрапа на языках R и Python

3.5 Тематика курсовых проектов (курсовых работ)

Курсовой проект (работа) не предусмотрен.

4. Учебно-методическое и информационное обеспечение

4.1. Нормативные документы и ГОСТы

1. Федеральный закон от 29 декабря 2012 года No 273-ФЗ «Об образовании в Российской Федерации» (с изменениями и дополнениями);
2. Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 09.03.02 Информационные системы и технологии, утвержденный Приказом Министерства образования и науки РФ от 19 сентября 2017 г. N 929 "Об утверждении федерального... Редакция с изменениями N 1456 от 26.11.2020;
3. Приказ Министерства образования и науки РФ от 05 апреля 2017 г. No 301 «Об утверждении Порядка организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры.

4.2. Основная литература

1. Data Science и интеллектуальный анализ данных : учебное пособие / Д. М. Назаров, С. В. Бегичева, Д. Б. Ковтун, А. Д. Назаров. — Москва : Ай Пи Ар Медиа, 2022. — 304 с. — ISBN 978-5-4497-1931-7. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: https://www.iprbookshop.ru/127201.html (дата

обращения: 16.01.2022)

2. Протодьяконов, А. В. Алгоритмы Data Science и их практическая реализация на Python : учебное пособие / А. В. Протодьяконов, П. А. Пылов, В. Е. Садовников. — Вологда : Инфра-Инженерия, 2022. — 392 с. — ISBN 978-5-9729-1006-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/281189 (дата обращения: 30.06.2022)

3. Агалаков, С. А. Анализ данных в среде R : учебное пособие / С. А. Агалаков. — Омск : ОмГУ, 2020. — 52 с. — ISBN 978-5-7779-2438-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/136348 (дата обращения: 30.06.2022)

4.3. Дополнительная литература

Алексеев, Д. С. Технологии интеллектуального анализа данных / Д. С. Алексеев, О. В. Щекочихин. — 2-е изд., стер. — Санкт-Петербург : Лань, 2024. — 176 с. — ISBN 978-5-507-48763-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/362915 (дата обращения: 31.10.2022)

4.4. Электронные образовательные ресурсы

Анализ данных LMS Московского политеха URL: https://online.mospolytech.ru/enrol/index.php?id=12581

1. Документация по языку R [Электронный ресурс]. – URL: https://cran.r-project.org/manuals.html (дата обращения: 05.06.2022).

2. Наглядная статистика. Используем R! [Электронный ресурс]. – URL: http://ashipunov.info/shipunov/software/r/r-ru.htm (дата обращения: 05.06.2022).

3. Анализ данных с использованием Python [Электронный ресурс]. – URL: https://habr.com/ru/articles/353050/(дата обращения: 05.06.2022).

4. Как анализировать данные в Pandas: первые шаги [Электронный ресурс]. – URL: https://blog.skillfactory.ru/kak-nachat-analizirovat-dannye-v-pandas-pervye-shagi/ (дата обращения: 05.06.2022).

5. Учимся анализировать — полный цикл [Электронный ресурс]. – URL: https://habr.com/ru/articles/685764/ (дата обращения: 05.06.2022).

6. Анализ данных: Определение, типы и примеры [Электронный ресурс]. – URL: https://hr-portal.ru/story/analiz-dannyh-opredelenie-tipy-i-primery (дата обращения: 05.06.2022).

7. Процесс анализа данных [Электронный ресурс]. – URL: https://pythonru.com/baza-znanij/process-analiza-dannyh (дата обращения: 05.06.2022).

4.5 Лицензионное и свободно распространяемое программное обеспечение

1. Microsoft Windows 7
2. Microsoft Office 2007
3. Python

4.6 Современные профессиональные базы данных и информационные справочные системы

1. ОП "Юрайт" https://urait.ru/
2. IPR Smart https://www.iprbookshop.ru/
3. ЭБС "Лань" https://e.lanbook.com/

5. Материально-техническое обеспечение

Лекционные аудитории общего фонда.
Аудитории для проведения практических занятий общего фонда.
Настенный/ переносной экран.
Переносной/ стационарный проектор для демонстрации слайдов.
Ноутбук для демонстрации слайдов.
Компьютерный класс для самостоятельной работы.
Библиотека, читальный зал.

6. Методические рекомендации

6.1. Методические рекомендации для преподавателя по организации обучения

Лекционные занятия проводятся в соответствии с содержанием настоящей рабочей программы.

Посещение лекционных занятий является обязательным.

Допускается конспектирование лекционного материала письменным или компьютерным способом.

Регулярная проработка материала лекций по каждому разделу в рамках подготовки к промежуточным и итоговым формам аттестации, а также выполнение и подготовка к защите лабораторных работ по дисциплине является одним из важнейших видов самостоятельной работы обучающегося в течение семестра.

6.2. Методические указания для обучающихся по освоению дисциплины

Посещение лекционных занятий является обязательным. Пропуск лекционных занятий без уважительных причин в объеме более 40% от общего количества предусмотренных учебным планом на семестр лекций влечет за собой невозможность аттестации по дисциплине, так как обучающийся не набирает минимально допустимого для получения итоговой аттестации по дисциплине количества баллов за посещение лекционных занятий.

Допускается конспектирование лекционного материала письменным или компьютерным способом.

Регулярная проработка материала лекций по каждому разделу в рамках подготовки к промежуточным и итоговым формам аттестации, а также выполнение и подготовка к защите лабораторных работ по дисциплине является одним из важнейших видов самостоятельной работы студента в течение семестра.

7. Фонд оценочных средств

В процессе обучения используются следующие оценочные формы самостоятельной работы студентов, оценочные средства текущего контроля успеваемости и промежуточных аттестаций:

подготовка к работе на лабораторных занятиях

7.1 Методы контроля и оценивания результатов обучения

В результате освоения дисциплины (модуля) формируются следующие компетенции:

Код компетенции	В результате освоения образовательной программы обучающийся должен обладать
ПК-5	Способен оптимизировать функционирование БД
ПК-6	Способен предотвращать потери и повреждения данных

В процессе освоения образовательной программы данные компетенции, в том числе их отдельные компоненты, формируются поэтапно в ходе освоения обучающимися дисциплин (модулей), практик в соответствии с учебным планом и календарным графиком учебного процесса.

7.2. Шкала и критерии оценивания результатов обучения

Форма промежуточной аттестации: зачет.

Промежуточная аттестация обучающихся в форме зачёта проводится по результатам выполнения всех видов учебной работы, предусмотренных учебным планом по данной дисциплине (модулю), при этом учитываются результаты текущего контроля успеваемости в течение семестра. Оценка степени достижения обучающимися планируемых результатов обучения по дисциплине (модулю) проводится преподавателем, ведущим занятия по дисциплине (модулю) методом экспертной оценки. По итогам промежуточной аттестации по дисциплине (модулю) выставляется оценка «зачтено» или «не зачтено».

К промежуточной аттестации допускаются только студенты, выполнившие все виды учебной работы, предусмотренные рабочей программой по дисциплине «Анализ данных» (посетили более 40% лекционных занятий, выполнили лабораторные работы, прошли промежуточный контроль в форме 3 контрольных работ).

Шкала оценивания	Описание

Зачтено	Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует соответствие знаний, умений, навыков приведенных в таблицах показателей, оперирует приобретенными знаниями, умениями, навыками, применяет их в ситуациях различной сложности. При этом могут быть допущены незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.
Не зачтено	Не выполнен один или более видов учебной работы, предусмотренных учебным планом. Студент демонстрирует неполное соответствие знаний, умений, навыков приведенных в таблицах показателей, допускаются значительные ошибки, проявляется отсутствие знаний, умений, навыков по ряду показателей, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.

7.2.2. Критерии оценки работы обучающегося на лабораторных занятиях

«5» (отлично): выполнены все практические задания, предусмотренные практическими занятиями, обучающийся четко и без ошибок ответил на все контрольные вопросы, активно работал на практических занятиях.

«4» (хорошо): выполнены все практические задания, предусмотренные практическими занятиями, обучающийся с корректирующими замечаниями преподавателя ответил на все контрольные вопросы, достаточно активно работал на практических занятиях.

«3» (удовлетворительно): выполнены все практические задания, предусмотренные практическими занятиями с замечаниями преподавателя; обучающийся ответил на все контрольные вопросы с замечаниями.

«2» (неудовлетворительно): обучающийся не выполнил или выполнил неправильно практические задания, предусмотренные практическими занятиями; студент ответил на контрольные вопросы с ошибками или не ответил на контрольные вопросы.

7.2.3. Критерии оценки тестирования

Тестирование оценивается в соответствии с процентом правильных ответов, данных студентом на вопросы теста.

Стандартная шкала соответствия результатов тестирования выставляемой балльной оценке:

− «отлично» - свыше 85% правильных ответов;
− «хорошо» - от 70,1% до 85% правильных ответов;
− «удовлетворительно» - от 55,1% до 70% правильных ответов;
− от 0 до 55% правильных ответов – «неудовлетворительно»

«5» (отлично): тестируемый демонстрирует системные теоретические знания, владеет терминами и обладает способностью быстро реагировать на вопросы теста.

«4» (хорошо): тестируемый в целом демонстрирует системные теоретические знания, владеет большинством терминов и обладает способностью быстро реагировать на вопросы теста.

«3» (удовлетворительно): системные теоретические знания у тестируемого отсутствуют, он владеет некоторыми терминами и на вопросы теста реагирует достаточно медленно.

«2» (неудовлетворительно): системные теоретические знания у тестируемого отсутствуют, терминологией он не владеет и на вопросы теста реагирует медленно.

7.3. Оценочные средства

Вопросы к зачету по дисциплине «Анализ данных»

1. Microsoft SQL Server
2. Бизнес-проблемы для интеллектуального анализа данных
3. Задачи интеллектуального анализа данных
4. Классификация
5. Кластеризация
6. Взаимосвязи
7. Регрессия
8. Прогнозирование
9. Анализ последовательностей
10. Анализ отклонений
11. Проектный цикл интеллектуального анализа данных
12. Формулирование бизнес-проблемы
13. Сбор данных
14. Очистка и преобразование данных
15. Создание модели
16. Оценка модели
17. Отчетность и прогнозирование
18. Интеграция приложения
19. Управление моделями
20. Настройка инструментов Table Analysis
21. Настройка аналитических служб с административными правами
22. Настройка аналитических служб без прав администратора
23. Подготовка к работе надстройки
24. Инструмент Analyze Key Influencers
25. Инструмент Detect Categories
26. Инструмент Fill From Example
27. Инструмент прогнозирования Forecast
28. Инструмент Highlight Exceptions
29. Инструмент Scenario Analysis
30. Инструмент Goal Seek
31. Инструмент Prediction Calculator
32. Инструмент анализа потребительской корзины Shopping Basket Analysis
33. Технический обзор инструментов Table Analysis
34. История DMX
35. Процесс интеллектуального анализа данных
36. Атрибут
37. Состояние
38. Вариант
39. Ключи
40. Входы и выходы
41. Объекты DMX
42. Структура интеллектуального анализа
43. Модель интеллектуального анализа
44. Синтаксис запросов DMX

45. Создание структур интеллектуального анализа
46. Дискретизированные столбцы
47. Вложенные таблицы
48. Разделение на обучающий и проверочный наборы данных
49. Создание моделей интеллектуального анализа
50. Вложенные таблицы
51. Сложные сценарии вложения
52. Фильтры
53. Заполнение структур интеллектуального анализа
54. Заполнение вложенных таблиц
55. Запрос данных структуры
56. Запрос данных модели
57. Прогнозирование
58. Прогнозирующее соединение
59. Синтаксис прогнозирующих запросов
60. Вложенные исходные данные
61. Прогноз в реальном времени
62. Вырожденные прогнозы
63. Прогнозирующие функции
64. Внешние или определяемые пользователем функции
65. Прогнозы по вложенным таблицам
66. Прогноз вложенных столбцов значений
67. Режимы Offline и Immediate
68. Создание объектов интеллектуального анализа данных
69. Источники данных
70. Создание источника данных MovieClick
71. Работа с именованными вычислениями
72. Исследование данных
73. Создание и редактирование моделей
74. Структуры и модели
75. Использование мастера Data Mining Wizard
76. Создание структуры и модели интеллектуального анализа MovieClick
77. Использование визуального конструктора Data Mining Designer
78. Работа с редактором Mining Structure Editor
79. Работа с редактором моделей интеллектуального анализа
80. Создание и модификация дополнительных моделей
81. Обработка структуры интеллектуального анализа MovieClick
82. Диаграмма точности анализа Mining Accuracy Chart
83. Выбор проверочных данных
84. Использование матрицы классификации
85. Использование CrossValidation
86. Использование построителя Mining Model Prediction Builder
87. Импорт данных с помощью Data Mining Client
88. Исследование и подготовка данных
89. Исключение нетипичных значений
90. Балансировка данных
91. Моделирование
92. Моделирование на основе задач
93. Сложное моделирование в Data Mining Client
94. Использование алгоритма Naive Bayes

95. Создание прогнозирующей модели
96. Использование деревьев решений
97. Модель регрессии
98. Взаимосвязи
99. Алгоритм Microsoft Time Series
100. Прогнозирование взаимозависимых рядов
101. Гипотетические сценарии
102. Авторегрессия
103. Периодичность
104. Деревья авторегрессии
105. Алгоритм Microsoft Clustering
106. Кластеризация как фаза анализа
107. Выявление аномалий при помощи кластеризации
108. Принципы кластеризации
109. Жесткая и мягкая кластеризация
110. Дискретная кластеризация
111. Масштабируемая кластеризация
112. Кластерное прогнозирование
113. Алгоритм Microsoft Sequence Clustering
114. Выполнение кластерных прогнозов
115. Выполнение прогнозов последовательности
116. Извлечение вероятности для прогнозов последовательностей
117. Использование гистограммы прогнозов последовательности
118. Выявление необычных шаблонов последовательностей
119. Матрица переходов состояний
120. Кластеризация с использованием цепи Маркова
121. Разбиение кластеров
122. Алгоритм взаимосвязей Microsoft Association Rules
123. Алгоритмы Microsoft Neural Network и Logistic Regression
124. Обратное распространение, функция ошибки и сопряженные градиенты
125. Нормализация и отображение
126. Топология сети
127. Обучение условия окончания
128. Нелинейно разделяемые классы
129. Интеллектуальный анализ кубов OLAP
130. Схема "звезда" и схема "снежинка"
131. Измерения и иерархии
132. Меры и группы мер
133. Обработка куба и его хранение
134. Использование упреждающего кэширования
135. Запросы к куб.

Примеры тестовых заданий

1. Пирамида Data Science включает следующие уровни: ...
a. принятие решений
b. мудрость, информация, данные
c. исследование данных
d. машинное обучение

2. На этапе “Внедрения” в CRISP-DM лучшими считаются модели: ...

a. ориентированные на конкретных пользователей, столкнувшихся с четко обозначенной проблемой
b. ориентированные на широкий круг пользователей, столкнувшихся разными проблемами
c. которые плавно вписываются в существующую практику

3. Функция в языке R, которая используется для конвертации матрицы или коллекции векторов в блок данных называется …
a. matrix
b. data.frame
c. dim
d. row.names

4. Во фреймворке tidyverse пакет “Tidymodels” предназначен для: ...
a. моделирования
b. работы с временем
c. машинного обучения
d. веб-скреппинга

5. Дисперсионный анализ с повторными измерениями реализуется для …
a. независимых выборок
b. зависимых выборок
c. многофакторного непараметрического вычисления
d. независимых непараметрических факторов