

**Санкт-Петербургский филиал федерального государственного
автономного образовательного учреждения высшего образования
"Национальный исследовательский университет
"Высшая школа экономики"**

Факультет Санкт-Петербургская школа социальных наук и востоковедения

Департамент государственного администрирования

**Рабочая программа дисциплины
«Аналитические платформы и управление данными»**

для образовательной программы «Управление и аналитика в государственном секторе»
направления подготовки 38.03.04. «Государственное и муниципальное управление»
уровень бакалавр

Разработчик программы
Нестеров Сергей Александрович, к.т.н., доцент

Согласована начальником ОСУП
«24» августа 2018 г.

Н.Э. Орешенкова _____

Утверждена Академическим советом образовательной программы
«27» августа 2018 г.
Протокол № 1

Академический руководитель образовательной программы

В.П. Кайсарова _____

Санкт-Петербург, 2018

Настоящая программа не может быть использована другими подразделениями университета и другими вузами без разрешения кафедры-разработчика программы.



1 Область применения и нормативные ссылки

Настоящая рабочая программа дисциплины устанавливает минимальные требования к знаниям и умениям студента, а также определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих дисциплину «Аналитические платформы и управление данными», учебных ассистентов и студентов направления подготовки 38.03.04. «Государственное и муниципальное управление», обучающихся по образовательной программе «Управление и аналитика в государственном секторе».

Рабочая программа дисциплины разработана в соответствии с:

- Образовательным стандартом НИУ ВШЭ, утвержденным Ученым советом Национального исследовательского университета «Высшая школа экономики»
- Образовательной программой «Управление и аналитика в государственном секторе» направления подготовки 38.03.04 "Государственное и муниципальное управление"
- Объединенным учебным планом университета по образовательной программе «Управление и аналитика в государственном секторе».

2 Цели освоения дисциплины

Целью курса «Аналитические платформы и управление данными» является получение студентами знаний, умений и навыков в области обработки и анализа данных с использованием технологий баз данных и интеллектуального анализа данных (data mining).

3 Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате изучения дисциплины студент осваивает следующие компетенции:

Компетенция	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
Способен выявлять научную сущность проблем в профессиональной области	Демонстрирует: - понимание роли и значения рыночной экономики; - способность всестороннего и критического анализа мер социально-экономической политики	Решение задач, поиск ответов на тестовые задания, анализ моделей и проблемных вопросов.	Устный зачёт, тест, экзамен
Способен работать с информацией: находить, оценивать и использовать информацию из различных источников, необходимую для решения научных и профессиональных задач (в том числе на основе системного подхода)	Владеет навыками поиска и использования информационных и статистических источников, характеризующих проблемы микроэкономической теории	Консультации по работе с источниками и интернет-ресурсами по курсу в процессе подготовки анализа моделей и обсуждения проблемных вопросов в домашних заданиях	Презентация по теме 1
Способен применять адекватные современные методы	Демонстрирует применение современных методов	Решение задач, поиск ответов на тестовые за-	Решение задач, выполнение



ды исследования для решения научных и аналитических задач в профессиональной деятельности	анализа деятельности субъектов рыночной экономики для решения аналитических задач	дания, анализ моделей и проблемных вопросов	тестов
Способен анализировать данные социальных, экономических, социологических исследований с использованием количественных и качественных методов	Владеет умением использовать социально-экономические исследования для изучения проблем микроэкономики	Анализ моделей и проблемных вопросов при подготовке домашних заданий	Выполнение тестов

4 Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к вариативной части профессионального цикла дисциплин, обеспечивающих подготовку бакалавра для направления подготовки 38.03.04. «Государственное и муниципальное управление».

Изучение данной дисциплины базируется на знаниях, полученных при изучении следующих дисциплин:

- «Математика».

Знания, полученные в ходе изучения дисциплины, могут быть использованы в дальнейшем в практической и исследовательской работе студентов.



5 Тематический план учебной дисциплины

ОБЪЕМ ДИСЦИПЛИНЫ - 3 зачетные единицы

№	Название раздела	Всего часов	Аудиторные часы		Самостоятельная работа
			Лекции	Семинары	
	1 часть «Технологии баз данных»				
1	<i>Тема 1.1</i> Основы реляционных баз данных	32	4	2	26
2	<i>Тема 1.2</i> Язык SQL	42	4	8	30
	2 часть «Основы интеллектуального анализа данных»				
3	<i>Тема 2.1.</i> Основные задачи интеллектуального анализа данных и алгоритмы их решения	40	4	6	30
ИТОГО		114	12	16	86

6 Содержание дисциплины

Темы и содержание занятий

Тема 1.1 Основы реляционных баз данных (лекц. 4 ауд. час, семинары 2 ауд. час)

Вводятся основные понятия, кратко рассматривается реляционная модель данных, реляционная алгебра и нормализация реляционных баз данных.

Семинар. Решение задач на реляционную алгебру и нормализацию. Создание первой базы данных в соответствии с предложенным образцом. Настройка ограничений целостности.

Самостоятельная работа включает изучение основной и дополнительной литературы, подготовку к семинару, выполнение практических заданий.

Тема 1.2 Язык SQL (лекц. 4 ауд. час, семинары 8 ауд. час)

История SQL; операторы DDL - создание таблиц и ограничений; операторы DML - добавление, изменение, удаление, записей; оператор SELECT - составление простых запросов, выборка данных из нескольких таблиц, подзапросы.

Семинар. Проектирование структуры базы данных на основе исходного набора данных. Добавление, изменение и удаление данных. Написание запросов на выборку данных на языке SQL.

Самостоятельная работа включает изучение основной и дополнительной литературы, подготовку к семинару, выполнение практических заданий.

Тема 2.1. Основные задачи интеллектуального анализа данных и алгоритмы их решения (лекц. 4 ауд. час, семинары 6 ауд. час)

Вводится понятие интеллектуального анализа данных, дается обзор задач интеллектуального анализа. Рассматривается ряд широко используемых алгоритмов - упрощенный алгоритм Байеса, деревья решений, алгоритмы кластеризации.

Семинар. На учебных наборах данных демонстрируется решение задач классификации и кластеризации, обсуждаются способы оценки качества построенной модели интеллектуального анализа.



Самостоятельная работа включает изучение основной и дополнительной литературы, подготовку к семинару, выполнение практических заданий.

7 Оценочные средства

7.1 Формы контроля знаний студентов

Тип контроля	Форма контроля	Модуль	Параметры
		4	
Текущий	Выполнение практических заданий на семинарских занятиях	*	Решение задач по темам курса (в том числе с использованием компьютерных технологий)
	Домашнее задание	*	Самостоятельное выполнение практических заданий по темам курса (в том числе с использованием компьютерных технологий)
Итоговый	Экзамен	*	Экзамен проводится в форме теста. Для теста рекомендуемая продолжительность – 60 минут

7.2 Критерии оценки знаний и навыков

Все работы оцениваются по 10-ти балльной шкале. При выполнении практических заданий на семинарских занятиях студент должен продемонстрировать умение применять количественные методы анализа данных, а также интерпретировать полученные результаты.

Критерии оценивания практических заданий

Балл	Критерии выставления оценки
8-10	Все задачи решены и дана интерпретация показателей; 85% и более правильных ответов на вопросы тестов
6-7	В основном задачи решены, но допущены незначительные ошибки, либо недостаточно полно проанализированы результаты расчетов; 65-85% правильных ответов на вопросы тестов
4-5	Часть задач не решена, либо решена неправильно. Недостаточно полно проанализированы результаты расчетов; 50-65% правильных ответов на вопросы тестов
0-3	Более 50% задания не выполнено; менее 50% правильных ответов на вопросы тестов

Результирующая оценка по дисциплине «Количественные методы анализа данных» складывается из накопленной оценки, полученной по результатам текущего контроля, аудиторной и самостоятельной работы студентов и оценки, полученной на экзамене (итоговый контроль).

Критерии оценивания домашней работы

Оценка	Критерии выставления оценки
«Отлично» (8-10)	Все задания выполнены правильно.
«Хорошо» (6-7)	В основном задания выполнены правильно, но допущены незначительные ошибки, либо недостаточно полно проанализированы результаты расчетов.
«Удовлетворительно» (4-5)	Часть заданий не выполнена, либо выполнена неправильно. Недостаточно полно проанализированы результаты расчетов.
«Неудовлетворительно» (0-3)	Более 50% задания не выполнено. Отсутствует анализ полученных результатов.



Примерный перечень вопросов к экзамену

1. Основные понятия баз данных (система баз данных, база данных, СУБД), файл-серверная и клиент-серверная архитектуры многопользовательских систем баз данных.
2. Реляционная модель данных: допустимые структуры данных, типы отношений и ограничения целостности.
3. Реляционная алгебра: отношения совместимые по типу, реляционные операции объединения, пересечения, вычитания, декартова произведения.
4. Реляционная алгебра: операции выборки, проекции, соединения, реляционного деления.
5. Нормализация реляционных баз данных: 1- 3 нормальные формы.
6. Нормализация реляционных баз данных: НФБК и старшие нормальные формы.
7. Язык SQL: типы данных, операторы определения данных (создание, изменение и удаление таблиц).
8. Язык SQL: операторы добавления, изменения, удаления данных.
9. Язык SQL: оператор SELECT.
10. Язык SQL: вложенные запросы, коррелированная и некоррелированная обработка вложенных запросов. Соединение таблиц при выполнении запроса.
11. Основные задачи интеллектуального анализа.
12. Этапы интеллектуального анализа данных.
13. Соответствие задач интеллектуального анализа и алгоритмов, применяемых для их решения.
14. Упрощенный алгоритм Байеса.
15. Алгоритмы, основанные на деревьях решений.
16. Алгоритмы кластеризации: k-средних, алгоритмы иерархической кластеризации.

Примеры заданий итогового контроля в форме теста

- 1) Файл-серверная архитектура многопользовательских систем баз данных предполагает, что СУБД находится:
 - на стороне клиента
 - на стороне сервера приложений
 - на стороне сервера баз данных
 - на стороне файлового сервера



- 2) Укажите наиболее младшую нормальную форму, требующую, чтобы все неключевые атрибуты неприводимо зависели от первичного ключа.
- 1НФ
 - **2НФ**
 - 3НФ
 - Ни одна из перечисленных
- 3) Имеется таблица Book со следующими полями:
ID – уникальный числовой идентификатор издания, первичный ключ;
Title – название книги;
Author – автор книги (принимается допущение, что у каждой книги только один автор);
Publisher – издательство, в котором вышла книга;
Year – год издания.
Нужно написать запрос, возвращающий всю информацию о книгах, изданных в издательствах «Азбука» или «Политехника». Среди перечисленных ниже вариантов один является неверным, выберите его.
- `SELECT * FROM Book WHERE Publisher IN ('Политехника','Азбука');`
 - `SELECT * FROM Book WHERE Publisher = 'Политехника' OR Publisher = 'Азбука';`
 - **`SELECT * FROM Book WHERE Publisher = 'Политехника', Publisher = 'Азбука';`**
 - `(SELECT * FROM Book WHERE Publisher='Политехника')`
`UNION`
`(SELECT * FROM Book WHERE Publisher='Азбука');`
- 4) Задача интеллектуального анализа, решение которой позволяет выявить группы среди исследуемого множества объектов это:
- Классификация
 - **Кластеризация**
 - Прогнозирование временных рядов
 - Анализ взаимосвязей

7.3 Порядок формирования оценок по дисциплине

Накопленная оценка по дисциплине рассчитывается с помощью взвешенной суммы оценок за отдельные формы текущего контроля знаний следующим образом:

$$O_{\text{накопл}} = 0,5 O_{\text{текущий}1} + 0,5 O_{\text{текущий}2}$$

где:

$O_{\text{текущий}1}$ – оценка за выполнение практических заданий на семинарских занятиях

$O_{\text{текущий}2}$ – оценка за домашнее задание.

Способ округления накопленной оценки текущего контроля: арифметический.

Результирующая оценка по дисциплине рассчитывается следующим образом:

$$O_{\text{результ}} = 0,6 \cdot O_{\text{накопл}} + 0,4 O_{\text{экзамен}}, \text{ где}$$

$O_{\text{накопл}}$ – накопленная оценка по дисциплине

$O_{\text{экзамен}}$ – оценка за экзамен

Способ округления экзаменационной и результирующей оценок: арифметический



На передаче студенту не предоставляется возможность получить дополнительный балл для компенсации оценки за текущий контроль.

8 Образовательные технологии

В преподавании дисциплины используется сочетание традиционных образовательных технологий (лекции и семинары/практические занятия) и информационно-коммуникационных технологий, в том числе, с использованием систем дистанционного обучения.

9 Учебно-методическое и информационное обеспечение дисциплины

9.1 Основная литература:

- С.А. Нестеров, Базы данных: учебник и практикум для академического бакалавриата. – М.: Издательство Юрайт, 2016. – 230 с. – Серия: Бакалавр. Академический курс. [электронная версия в ЭБС издательства Юрайт <https://biblio-online.ru/book/bazy-dannyh-433369>, доступ по подписке]
- Замятин А.В. Интеллектуальный анализ данных : учеб. пособие. – Томск: Издательский Дом Томского государственного университета, 2016. – 120 с. [электронная версия в ЭБС Томского гос. ун-та <http://vital.lib.tsu.ru/vital/access/manager/Repository/vtls:000529594/SOURCE1>, свободный доступ]

9.2 Дополнительная литература:

- С. А. Нестеров, Базы данных [Электронный ресурс]: учебное пособие – СПб: Санкт-Петербургский государственный политехнический университет, 2013. [электронная версия в ЭБС СПбПУ <http://elibr.spbstu.ru/dl/2/3405.pdf>, свободный доступ]

9.3 Ресурсы информационно-телекоммуникационной сети «Интернет»

- С. А. Нестеров, Н. В. Андреева Управление данными: онлайн-курс. URL: <https://openedu.ru/course/spbstu/DATAM/>
- Intro to SQL for Data Science: interactive course. URL: <https://www.datacamp.com/courses/intro-to-sql-for-data-science>
- Data Mining with Weka: online course. URL: <https://www.futurelearn.com/courses/data-mining-with-weka>
- More Data Mining with Weka: online course. URL: <https://www.futurelearn.com/courses/more-data-mining-with-weka>

10 Рекомендации для самостоятельной работы студентов

Самостоятельная работа предполагает индивидуальную подготовку к занятиям (изучение лекционного материала, чтение рекомендуемой литературы, подготовку к промежуточной и итоговой аттестации) и выполнение домашних заданий.

Этапы самостоятельной работы:

- изучение учебной и дополнительной литературы по теме курса;
- выполнение практических заданий выданных для самостоятельного решения (домашняя работа)
- изучение рекомендованных преподавателем онлайн-курсов.



Реализация программы предполагает активную самостоятельную работу студентов.

11 Материально-техническое и программное обеспечение дисциплины

Лекции проводятся в аудиториях, оснащенных мультимедийным оборудованием. Практические занятия проводятся в компьютерных классах. Для успешного освоения дисциплины, студенты на разных этапах изучения курса будут использовать следующие программные средства: Microsoft Access или другую настольную СУБД, SQL Server или другую СУБД корпоративного уровня, программные средства интеллектуального анализа данных (пакет программ WEKA или пакеты для языка R).

12 Особенности организации обучения для лиц с ограниченными возможностями здоровья

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося) могут предлагаться следующие варианты восприятия учебной информации с учетом их индивидуальных психофизических особенностей, в том числе с применением электронного обучения и дистанционных технологий:

1) *для лиц с нарушениями зрения:* в печатной форме увеличенным шрифтом; в форме электронного документа; в форме аудиофайла (перевод учебных материалов в аудиоформат); индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания и консультации.

2) *для лиц с нарушениями слуха:* в печатной форме; в форме электронного документа; видеоматериалы с субтитрами; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания и консультации.

3) *для лиц с нарушениями опорно-двигательного аппарата:* в печатной форме; в форме электронного документа; в форме аудиофайла; индивидуальные задания и консультации.