

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

Факультет Санкт-Петербургская школа экономики и менеджмента  
Национального исследовательского университета  
«Высшая школа экономики»

Департамент прикладной математики и бизнес-информатики

**Рабочая программа дисциплины  
Интеллектуальный анализ данных**

для направления 01.06.01 «Математика и механика»  
подготовки научно-педагогических кадров в аспирантуре,  
образовательная программа «Вещественный, комплексный и функциональный анализ»

Разработчик программы

Сироткин А.В., к.ф.-м.н., доцент департамента прикладной математики и бизнес информатики,  
avsirotkin@hse.ru

Согласована Академическим советом Аспирантской школы по математике

«23» октября 2017 г., протокол № 8

Санкт-Петербург, 2017

*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 аспиранта, а также определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих дисциплину "Интеллектуальный анализ данных", и аспирантов направления подготовки 01.06.01 Математика и механика, профиль (направленность) «Вещественный, комплексный и функциональный анализ».

Рабочая программа дисциплины разработана в соответствии с:

- Образовательным стандартом НИУ ВШЭ по направлению подготовки 01.06.01 Математика и механика
- Образовательной программой по направлению подготовки 01.06.01 Математика и механика, профиль (направленность) «Вещественный, комплексный и функциональный анализ».
- Учебным планом образовательной программы «Вещественный, комплексный и функциональный анализ», утвержденным в 2017 г.

## 2. Цели освоения дисциплины

Целями дисциплины являются формирование у аспирантов системного представления о технологиях интеллектуального анализа данных (Data Mining), их применении и инструментах, изучение основных методов прикладного анализа данных, развитие навыков практического применения методов Data Mining для решения различных научных и технических задач.

В результате изучения курса аспиранты должны ознакомиться с основами построения систем поддержки принятия решения, принципами работы OLAP-систем, методологией многомерного анализа, технологией Data Mining, изучить методы, инструментальные средства и область применения Data Mining.

## 3. Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины аспирант должен

- Знать отличия Data Mining от классических статистических методов анализа и OLAP-систем, типы закономерностей и сферы применения Data Mining.
- Уметь квалифицировать задачи Data Mining, применять методы интеллектуального анализа данных.
- Иметь представление о тенденциях технологий интеллектуального анализа данных, стандартах и инструментах.

В результате освоения дисциплины аспирант осваивает следующие компетенции:

Компетенция	Код по ОС ВШЭ	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
способность выбирать и применять методы исследования, адекватно	УК-3	Способен формулировать и определять задачи для решения с использованием различных методов анализа	Лекции, практические занятия, самостоятельная работа при подготовке к практическим заня-	Аудиторная работа, самостоятельная работа, самостоятельная



Компетенция	Код по ОС ВШЭ	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
ватные предмету и задачам исследования		данных, осознанно выбирать методы и инструментальные средства	тиям	работа, реферат
способность собирать, анализировать, обрабатывать и хранить данные в соответствии с общепринятыми научными и этическими стандартами	УК-4	Владеет навыками написания программ для автоматизированного сбора и анализа информации из различных источников в глобальных компьютерных сетях; осуществляет оценку качества полученных решений и результатов	Лекции, практические занятия, самостоятельная работа при подготовке к практическим занятиям, консультации преподавателей, самостоятельная работа с литературой	Аудиторная работа, активность при обсуждении вопросов и заданий, предложенных преподавателем, Устный экзамен
способность осуществлять комплексные исследования, в том числе междисциплинарные, на основе целостного системного научного мировоззрения	УК-5	Умеет на основе комбинированного применения методов анализа данных строить принципиально новые модели	Лекции, практические занятия, самостоятельная работа при подготовке к практическим занятиям, консультации преподавателей	Аудиторная работа, самостоятельная работа

#### 4. Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к блоку дисциплин по выбору вариативной части для направления 01.06.01 Математика и механика, профиль (направленность) «Вещественный, комплексный и функциональный анализ».

Изучение данной дисциплины базируется на знаниях, умениях и навыках и компетенциях, сформированных на двух предшествующих уровнях высшего образования. Прежде всего, речь идет о владении современными методами и средствами получения, хранения и обработки информации. Для освоения учебной дисциплины аспиранты должны владеть знаниями и компетенциями, позволяющими подвергать первичному анализу данные эмпирических исследований, строить простейшие стохастические модели и уметь интерпретировать полученные результаты. Знания, навыки и умения, приобретенные в результате прохождения курса, будут востребованы при выполнении научных работ, связанных с интеллектуальной обработкой больших объемов информации.

#### 5. Тематический план учебной дисциплины

ОБЪЕМ ДИСЦИПЛИНЫ – 4 зачетных единиц.

№	Название раздела (темы)	Всего часов	Аудиторные часы			Самостоятельная работа
			Лекции	Семинары	Практические занятия	
1	Обзор задач интеллектуального анализа данных	24	2		2	20



2	Основные методы интеллектуального анализа данных	33	2		6	25
3	Процесс Data Mining	33	2		6	25
4	Инструментальные средства интеллектуального анализа данных	28	2		6	20
5	Практика применения интеллектуальных технологий	34	2		6	26
ИТОГО		152	10		26	116

## 6. Содержание дисциплины

### Тема 1. Обзор задач интеллектуального анализа данных

Понятие Интеллектуального анализа данных (Data Mining). Data Mining как часть рынка интеллектуальных технологий. Набор данных и их атрибутов. Измерения. Типы наборов данных. Форматы хранения данных. Метаданные.

### Тема 2. Основные методы интеллектуального анализа данных

Задача классификации. Процесс классификации. Методы, применяемые для решения задач классификации. Точность классификации: оценка уровня ошибок. Оценивание классификационных методов. Деревья решений. Процесс конструирования дерева решений. Метод опорных векторов. Метод «ближайшего соседа». Байесова классификация. Задача прогнозирования. Сравнение задач прогнозирования и классификации. Прогнозирование и временные ряды. Решение задачи прогнозирования. Задача кластеризации. Применение кластерного анализа. Иерархические методы. Итеративные методы. Методы поиска ассоциативных правил. Методы визуализации. Качество визуализации. Представление пространственных характеристик. Основные тенденции в визуализации.

### Тема 3. Процесс Data Mining

Средства извлечения данных: методы и возможности. Начальные этапы: анализ предметной области; постановка задачи, подготовка данных. Очистка данных. Инструменты очистки данных. Построение и использование модели. Стандарты Data Mining.

### Тема 4. Инструментальные средства интеллектуального анализа данных

Рынок инструментов Data Mining. Классификация инструментов Data Mining. Программное обеспечение для решения задач классификации. Программное обеспечение для решения задач кластеризации и сегментации. Программное обеспечение Data Mining для поиска ассоциативных правил. Программное обеспечение для решения задач оценивания и прогнозирования.

### Тема 5. Практика применения интеллектуальных технологий

Системы бизнес-интеллекта и управления знаниями. Сферы применения Data Mining. Применение Data Mining для бизнес-задач. Data Mining для научных исследований. Data Mining консалтинг. Data Mining услуги. Примеры решения. Техническое описание решения. Технологии лингвистического анализа бизнес-информации. Интеллектуальный поиск в интернете. Аналитическая обработка бизнес-информации. Комплексный подход к внедрению Data Mining, OLAP и хранилищ данных. Интеграция OLAP и Data Mining. Хранилища данных. Преимущества хранилища данных.

## 7. Образовательные технологии

Для реализации программы дисциплины используются лекционные занятия, практические занятия и задания для самостоятельной работы. В рамках курса применяются интерак-



тивные образовательные технологии, которые задействованы и в лекционных, и в практических занятиях. В ходе лекций и практических занятий обсуждаются основные проблемы, освещенные в ходе занятия, развиваются навыки использования информационных систем и технологий в процессе решения задач, сформулированных на практических занятиях, осуществляется разбор конкретных ситуаций. Самостоятельная работа аспирантов подразумевает освоение теоретического материала, подготовку к практическим занятиям, выполнение практических заданий и работу с литературой.

## 8. Оценочные средства для текущего, промежуточного и итогового контроля по дисциплине

### Формы контроля знаний аспирантов

Тип контроля	Форма контроля	2 год	Параметры
		1 полугодие	
Текущий	Аудиторная работа	*	Участие в обсуждении, вынесенных на рассмотрение вопросов и заданий, демонстрация знакомства с релевантной литературой
	Реферат	1	Реферат
Итоговый	Экзамен	1	Устный экзамен

Текущий контроль по дисциплине осуществляется путем оценки усвоения материала в ходе аудиторной работы на практических занятиях и лекциях, а также в форме оценки подготовленного реферата. Итоговый контроль по дисциплине осуществляется в форме устного экзамена.

**Текущий контроль** проводится во время интерактивного взаимодействия преподавателя и аспирантов на аудиторных занятиях. Оценивается активность аспирантов в обсуждении вынесенных на рассмотрение вопросов и заданий, демонстрация знакомства с рекомендованной литературой.

### Критерии оценивания и шкала оценки работы аспирантов на аудиторных занятиях

Оценка	Критерии выставления оценки
«Отлично» (8-10)	Аспирант обнаруживает всестороннее, систематическое и глубокое знание учебно-программного материала; принимает активное участие в обсуждении по теме занятия; усвоил основную и дополнительную литературу, рекомендованную рабочей программой дисциплины; проявляет творческие способности в понимании, изложении и использовании учебно-программного материала.
«Хорошо» (6-7)	Аспирант обнаруживает достаточное знание учебно-программного материала и основных категорий курса; усвоил основную литературу, рекомендованную в рабочей программе дисциплины, знаком с некоторыми источниками из списка дополнительной литературы.
«Удовлетворительно» (4-5)	Аспирант обнаруживает знания основного учебно-программного материала в объеме, необходимом для дальнейшей



	учебы, в целом знаком с основной литературой, рекомендованной рабочей программой дисциплины, участвует в обсуждении не достаточно активно, не задает вопросы.
«Неудовлетворительно» (1-3)	Аспирант не принимает участия в обсуждении на семинарском занятии, не обнаруживает знания основного учебно-программного материала. Не демонстрирует знакомства с основной литературой

### Реферат

#### Структура и требования к реферату:

- титульный лист;
- содержание;
- введение (0,5-1 стр.);
- основная часть (до 20 стр.);
- заключение (3-5 стр.);
- список литературы (не менее 5 источников);
- приложения (если есть).

#### Примерный перечень тем рефератов.

1. Системы бизнес-интеллекта и управление задач.
2. Применение Data Mining для решения бизнес-задач.
3. Data Mining для научных исследований.
4. Data Mining консалтинг.
5. Технологии лингвистического анализа бизнес-информации Text Mining.
6. Современные технологии выбора признаков
7. Комплексный подход к внедрению Data Mining, OLAP и хранилищ данных.
8. Web Mining на основе мультиагентных систем.
9. Основные тенденции в области визуализации данных.
10. Аналитическая обработка бизнес-информации в торговой, банковской и производственной сферах.

#### **Критерии оценивания и шкала оценки реферата**

Оценка	Критерии выставления оценки
«Отлично» (8-10)	Тема полностью раскрыта, четко выражена авторская позиция, имеются логичные и обоснованные выводы. Автор свободно ориентируется в материале, ссылается на других авторов, разработывавших тему. Использованная литература релевантна. Список литературы содержит не менее 5 источников. Работа оформлена на высоком уровне.
«Хорошо» (6-7)	Тема в целом раскрыта, прослеживается авторская позиция, сформулированы необходимые выводы, автор уверенно ориентируется в материале. Имеются замечания / неточности в части изложения и отдельные недостатки по оформлению работы. Использованная литература релевантна. Список литературы содержит не менее 5 источников.
«Удовлетворительно» (4-5)	Тема раскрыта недостаточно полно, авторская позиция выражена слабо / выводы не обоснованы; материал изложен



	непоследовательно, без соответствующей аргументации и необходимого анализа. Список литературы не достаточно релевантный/ полный. Имеются недостатки в оформлении.
«Неудовлетворительно» (1-3)	Тема не раскрыта; материал изложен без собственной оценки и выводов. Имеются недостатки в оформлении работы. Автор плохо ориентируется в представленном материале. Список литературы не релевантный/ полный. Имеются недостатки в оформлении работы.

Если при проверке работы на плагиат обнаружен процент заимствований более 20%, за реферат выставляется оценка "0".

### **Оценочные средства для итогового контроля по дисциплине (устный экзамен)**

Итоговый контроль по дисциплине проводится в форме устного экзамена. Экзаменационный билет содержит два вопроса. Время на подготовку – 40 мин.

### **Примеры вопросов для итогового контроля по дисциплине (устный экзамен)**

1. Понятие Интеллектуального анализа данных (Data Mining). Data Mining как часть рынка интеллектуальных технологий
2. Интеллектуальный анализ данных. Отличия от других видов анализа данных
3. Типы наборов данных. Форматы хранения данных. Метаданные.
4. Особенности использования данных при интеллектуальном
5. анализе данных. Данные, информация и знания.
6. Классификация задач Data Mining. Сферы применения Data Mining.
7. Процесс Data Mining
8. Задачи интеллектуального анализа данных. Классификация и кластеризация
9. Задачи интеллектуального анализа данных. Прогнозирование и визуализация
10. Методы прогнозирования и классификации: деревья решений
11. Методы прогнозирования и классификации: метод опорных векторов
12. Методы прогнозирования и классификации: метод «ближайшего соседа»
13. Методы прогнозирования и классификации: нейронные сети.
14. Методы поиска ассоциативных правил
15. Байесова классификация
16. Поставщики Data Mining. Классификация инструментов.
17. Программное обеспечение Data Mining для поиска ассоциативных правил.
18. Программное обеспечение для решения задач кластеризации и сегментации.
19. Программное обеспечение для решения задач классификации.
20. Программное обеспечение Data Mining для решения задач оценивания и прогнозирования
21. Средства извлечения данных
22. Хранилища данных
23. Сферы применения Data Mining
24. Очистка данных. Инструменты очистки данных
25. Стандарты Data Mining
26. Интеграция Data Mining и OLAP

### **Критерии оценивания и шкала оценки устного экзамена**

Оценка	Критерии выставления оценки
--------	-----------------------------





«Отлично» (8-10)	Дан полный ответ на вопрос. Имеются логичные и аргументированные выводы (приведено несколько аргументов). Даны ссылки на использованную при подготовке к экзамену литературу. Приведены примеры из практики. Продемонстрирован междисциплинарный подход к освещению вопроса. Ответы на дополнительные вопросы демонстрируют глубокое знание проблемы.
«Хорошо» (6-7)	Дан полный ответ на вопрос. Выводы в целом логичные и аргументированные ( не менее 2-х аргументов). Даны ссылки на использованную при подготовке к экзамену литературу. Приведены примеры из практики. Продемонстрирован междисциплинарный подход к освещению вопроса. Ответы на дополнительные вопросы демонстрируют достаточно глубокое знание проблемы.
«Удовлетворительно» (4-5)	Ответ на вопрос не является полным. Выводы не достаточно логичны, аргументы не достаточны . Примеры из практики не достаточны для подтверждения теоретических выводов. Ответы на дополнительные вопросы демонстрируют поверхностное знание проблемы.
«Неудовлетворительно» (1-3)	Ответ на вопрос не является полным/ является неправильным. Выводы не логичны, аргументы не достаточны, даны ссылки на не релевантные источники /не даны. Не приведены примеры из практики. Не продемонстрирован междисциплинарный подход к освещению вопроса. Даны неправильные/не даны ответы на дополнительные вопросы.

Оценкой за экзамен является среднее арифметическое оценок за каждый вопрос. Оценка за каждый вопрос является целым числом.

## 9. Порядок формирования оценок по дисциплине

Оценки по всем формам контроля выставляются по 10-балльной шкале.

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

$$\text{Онакопленная} = 0,7 \cdot \text{Ореф} + 0,3 \cdot \text{Оауд}, \text{ где}$$

*Ореф* – оценка за реферат;

*Оауд* – оценка за аудиторная работа.

Способ округления накопленной оценки текущего контроля арифметический.

Результирующая оценка по дисциплине рассчитывается следующим образом:

$$\text{Результ} = 0,5 \cdot \text{Онакопленная} + 0,5 \cdot \text{Оэкз}, \text{ где}$$

*Онакопленная* – накопленная оценка по дисциплине;

*Оэкз* – оценка за экзамен.

Способ округления экзаменационной и результирующей оценок – арифметический.





## 10. Учебно-методическое и информационное обеспечение дисциплины

### 10.1 Основная литература

1. Информационные аналитические системы [Электронный ресурс] : учебник / Т. В. Алексеева, Ю. В. Амириди, В. В. Дик и др.; под ред. В. В. Дика. - М.: МФПУ Синергия, 2013. - 384 с. - Режим доступа по паролю: <http://znanium.com/bookread2.php?book=451186> (ЭБС Znanium)
2. Вейнберг Р. Р. Интеллектуальный анализ данных и систем управления бизнес-правилами в телекоммуникациях: Монография [Электронный ресурс] / Р.Р. Вейнберг. - М.: НИЦ ИНФРА-М, 2016. - 173 с.: 60x90 1/16. - (Научная мысль). - Режим доступа по паролю: <http://znanium.com/bookread2.php?book=520998> (ЭБС Znanium)
3. Интеллектуальный анализ временных рядов: Учебное пособие [Электронный ресурс] / Н.Г. Ярушкина, Т.В. Афанасьева, И.Г. Перфильева. - М.: ИД ФОРУМ: ИНФРА-М, 2012. - Режим доступа по паролю: <http://znanium.com/bookread2.php?book=249314> (ЭБС Znanium)

### 10.2 Дополнительная литература

1. Дорогов В.Г. Введение в методы и алгоритмы принятия решений [Электронный ресурс]: учебное пособие / В.Г. Дорогов, Я.О. Теплова. – Москва: ИД ФОРУМ: ИНФРА-М, 2012. – Режим доступа по паролю: <http://znanium.com/bookread2.php?book=241287> (ЭБС Znanium)
2. Миркин, Б. Г. Введение в анализ данных : учебник и практикум [Электронный ресурс] / Б. Г. Миркин. — М. : Издательство Юрайт, 2017. - Режим доступа по паролю: <https://www.biblio-online.ru/viewer/46A41F93-BC46-401C-A30E-27C0FB60B9DE#page/1> (ЭБС ЮРАЙТ)
3. Станкевич, Л. А. Интеллектуальные системы и технологии : Учебник и практикум для бакалавриата и магистратуры [Электронный ресурс] / Л. А. Станкевич. — М. : Издательство Юрайт, 2017. - Режим доступа по паролю: <https://www.biblio-online.ru/viewer/A45476D8-8106-487A-BA38-2943B82B4360#page/1> (ЭБС ЮРАЙТ)
4. Bassan, A. S., & Sarkar, D. (2014). Mastering SQL server 2014 data mining: Master selecting, applying, and deploying data mining models to build powerful predictive analysis frameworks [Electronic Resource]. - Birmingham, UK: Packt Publishing. - Authorized access: <https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=1911531> (Online Digital Library "Ebrary")
5. Hsing, Tailen, and Randall L. Eubank. Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operations [Electronic resource]. -Wiley, 2015. - Authorized access: <https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=1991870> (Online Digital Library "Ebrary")

### 10.3 Ресурсы информационно-телекоммуникационной сети «Интернет»

1. <http://www.machinelearning.ru/> - коллекция материалов по машинному обучению

### 10.4 Программные средства

Для успешного освоения дисциплины, аспирант использует следующие программные средства:

- Rstudio;
- MS Word, MS Excel, MS Power Point(при необходимости);
- браузеры Google Chrome и др.



## **11. Материально-техническое обеспечение дисциплины**

В рамках курса используется компьютер, проектор.