


ФГБОУ ВО «БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
ФАКУЛЬТЕТ МАТЕМАТИКИ И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Утверждено:  
на заседании кафедры  
математического моделирования  
протокол от «17» апреля 2020 г.  
№ 9

Согласовано:  
Председатель УМК факультета /института

 / Ефимов А.М.

Зав. кафедрой  / С.И. Спивак

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

дисциплина *Open Source инструменты для анализа данных*

Цикл Б1.В.ДВ Дисциплины по выбору

**программа бакалавриата**

Направление подготовки

*01.03.02 Прикладная математика и информатика*

Направленность (профиль) подготовки

*"Математическое моделирование, численные методы и комплексы программ"*

Квалификация  
бакалавр

Разработчик (составитель) доцент, к.ф.-м.н., доцент	/ <u>Абдюшева С.Р.</u> 
--	--

Для приема: 2020

Уфа 2020 г.

Составитель к.ф.-м.н., доц. Абдюшева С.Р.

Рабочая программа дисциплины утверждена на заседании кафедры протокол от «\_17\_»  
\_\_апреля\_\_ 2020 \_\_г. № 9

## Список документов и материалов

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций
2. Цель и место дисциплины в структуре образовательной программы
3. Содержание рабочей программы (объем дисциплины, типы и виды учебных занятий, учебно-методическое обеспечение самостоятельной работы обучающихся)
4. Фонд оценочных средств по дисциплине
  - 4.1. Перечень компетенций и индикаторов достижения компетенций с указанием соотнесенных с ними запланированных результатов обучения по дисциплине. Описание критериев и шкал оценивания результатов обучения по дисциплине.
  - 4.2. Типовые контрольные задания или иные материалы, необходимые для оценивания результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине.
5. Учебно-методическое и информационное обеспечение дисциплины
  - 5.1. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины
  - 5.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет» и программного обеспечения, необходимых для освоения дисциплины, включая профессиональные базы данных и информационные справочные системы
6. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

**1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций**

По итогам освоения дисциплины обучающийся должен достичь следующих результатов обучения:

Категория (группа) компетенций (при наличии ОПК)	Формируемая компетенция (с указанием кода)	Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине
	ПК-1. Способен продемонстрировать базовые знания математических и естественных наук, программирования и информационных технологий	ПК-1.1. Обладает базовыми знаниями, полученными в области математических и (или) естественных наук, программирования и информационных технологий.	При ответах на вопросы демонстрирует базовые знания математических и естественных наук, программирования и информационных технологий.
		ПК-1.2. Умеет находить, формулировать и решать стандартные задачи в собственной научно-исследовательской деятельности в области программирования и информационных технологий.	Формулирует и решает стандартные задачи в собственной научно-исследовательской деятельности в области программирования и информационных технологий.
		ПК-1.3. Имеет практический опыт научно-исследовательской деятельности в области программирования и информационных технологий.	Обладает практическим опытом научно-исследовательской деятельности в области программирования и информационных технологий.
	ПК-4. Способен использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования.	ПК-4.1. Знает современные методы разработки и реализации алгоритмов математических моделей, управления информацией и данными.	Демонстрирует знание: современных методов разработки и реализации алгоритмов математических моделей, управления информацией и данными; интеллектуальных интернет-технологий, применимых в процессе принятия решений; ИТ-средств для принятия решений: регрессионный анализ, кластерный анализ, классификация, и др.
		ПК-4.2. Умеет разрабатывать и реализовывать алгоритмы математических моделей на базе языков и пакетов прикладных программ моделирования,	Демонстрирует умение: Разрабатывать и реализовывать алгоритмы математических моделей, искать источники информации и данные, воспринимать, анализировать, передавать

		<p>управлять информацией и данными.</p>	<p>информацию с использованием цифровых средств и алгоритмов, проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных.</p> <p>Применять методы принятия решений для задач прогнозирования, разделения наблюдений на классы, как по числовым, так и по категориальным признакам, формирования клиентских предложений, и т.п.</p>
		<p>ПК-4.3. Имеет практический опыт разработки и реализации алгоритмов на базе языков и пакетов прикладных программ моделирования, методами управления информацией и данными.</p>	<p>Обладает практическим опытом разработки и реализации алгоритмов на базе языков и пакетов прикладных программ моделирования.</p> <p>Владеет начальными навыками использования инструментов, поддерживающих реализацию методов машинного обучения – язык R, библиотеки Python, Excel для работы с данными.</p>

## 2. Цель и место дисциплины в структуре образовательной программы

Дисциплина «Open Source инструменты для анализа данных» относится к части, формируемой участниками образовательных отношений, является дисциплиной по выбору.

Дисциплина изучается на 4 курсе в 1 семестре.

Целями освоения дисциплины «Open Source инструменты для анализа данных» являются: ознакомление с современными инструментами анализа данных, основами теории принятия решений, методов прогнозирования, фундаментальная подготовка в области построения и анализа статистических моделей, овладение современным математическим аппаратом статистического анализа для дальнейшего использования в разнообразных приложениях. А также ознакомление со сквозными технологиями - ключевыми научно-техническими направлениями, которые оказывают наиболее существенное влияние на развитие рынков; более подробное изучение сквозной технологии **Нейротехнологии и искусственный интеллект** и ее субтехнологии **Рекомендательные системы и интеллектуальные системы поддержки принятия решений**; овладение основными методами машинного обучения и инструментами, поддерживающими реализацию методов машинного обучения - библиотеки языка R, Python, Excel для работы с данными.

### 3. Содержание рабочей программы (объем дисциплины, типы и виды учебных занятий, учебно-методическое обеспечение самостоятельной работы обучающихся)

Содержание рабочей программы представлено в Приложении 1.

Объем дисциплины «Open Source инструменты для анализа данных» составляет 6 ЗЕТ, или 216 академических часов, в том числе контактная работа с преподавателем 72,4 часа, самостоятельная работа студентов – 108 часов, контроль – 34,8 часа.

### 4. Фонд оценочных средств по дисциплине

#### 4.1. Перечень компетенций и индикаторов достижения компетенций с указанием соотнесенных с ними запланированных результатов обучения по дисциплине. Описание критериев и шкал оценивания результатов обучения по дисциплине.

Код и формулировка компетенции: ПК-1. Способен демонстрировать базовые знания математических и естественных наук, программирования и информационных технологий.

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Критерии оценивания результатов обучения			
		2 («Не удовлетворительно»)	3 («Удовлетворительно»)	4 («Хорошо»)	5 («Отлично»)
ПК-1.1. Обладает базовыми знаниями, полученными в области математических и (или) естественных наук, программирования и информационных технологий.	При ответах на вопросы демонстрирует базовые знания математических и естественных наук, программирования и информационных технологий.	Фрагментарные знания	Неполные знания	Сформированные, но содержащие отдельные пробелы знания	Сформированные систематические знания
ПК-1.2. Умеет находить, формулировать и решать стандартные задачи в собственной научно-исследовательской деятельности в области программирования и информационных технологий.	Формулирует и решает стандартные задачи в собственной научно-исследовательской деятельности в области программирования и информационных технологий.	Фрагментарные умения	Неполные умения	Сформированные, но содержащие отдельные пробелы умения	Сформированные систематические умения

ПК-1.3. Имеет практический опыт научно-исследовательской деятельности в области программирования и информационных технологий.	Обладает практическим опытом научно-исследовательской деятельности в области программирования и информационных технологий.	Фрагментарные владения	Неполные владения	Сформированные, но содержащие отдельные пробелы владения	Сформированные систематические владения
---	--	------------------------	-------------------	--	---

Код и формулировка компетенции ПК-4. Способен использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования.

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Критерии оценивания результатов обучения			
		2 («Не удовлетворительно»)	3 («Удовлетворительно»)	4 («Хорошо»)	5 («Отлично»)
ПК-4.1. Знает современные методы разработки и реализации алгоритмов математических моделей, управления информацией и данными.	Демонстрирует знание: современных методов разработки и реализации алгоритмов математических моделей, управления информацией и данными; интеллектуальных интернет-технологий, применимых в процессе принятия решений; ИТ-средств для принятия решений: регрессионный анализ, кластерный анализ, классификация, и др.	Фрагментарные знания	Неполные знания	Сформированные, но содержащие отдельные пробелы знания	Сформированные систематические знания
ПК-4.2. Умеет разрабатывать и реализовывать алгоритмы математических моделей на базе языков и пакетов прикладных программ моделирования, управлять информацией и данными.	Демонстрирует умение: Разрабатывать и реализовывать алгоритмы математических моделей, искать источники информации и данные, воспринимать, анализировать, передавать информацию с использованием цифровых средств и алгоритмов, проводить оценку информации, ее достоверность, строить	Фрагментарные умения	Неполные умения	Сформированные, но содержащие отдельные пробелы умения	Сформированные систематические умения

	логические умозаключения на основании поступающих информации и данных. Применять методы принятия решений для задач прогнозирования, разделения наблюдений на классы, как по числовым, так и по категориальным признакам, формирования клиентских предложений, и т.п.				
ПК-4.3. Имеет практический опыт разработки и реализации алгоритмов на базе языков и пакетов прикладных программ моделирования, методами управления информацией и данными.	Обладает практическим опытом разработки и реализации алгоритмов на базе языков и пакетов прикладных программ моделирования. Владеет начальными навыками использования инструментов, поддерживающих реализацию методов машинного обучения – язык R, библиотеки Python, Excel для работы с данными.	Фрагментарные владения	Неполные владения	Сформированные, но содержащие отдельные пробелы владения	Сформированные систематические владения

**4.2. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и опыта деятельности, характеризующих этапы формирования компетенций**

Код и формулировка компетенции: ПК-1. Способен демонстрировать базовые знания математических и естественных наук, программирования и информационных технологий.

<b>Код и наименование индикатора достижения компетенции</b>	<b>Результаты обучения по дисциплине</b>	<b>Оценочные средства</b>
ПК-1.1. Обладает базовыми знаниями, полученными в	При ответах на вопросы демонстрирует базовые знания	Групповой и индивидуальный опрос Зачет



области математических и (или) естественных наук, программирования и информационных технологий.	математических и естественных наук, программирования и информационных технологий.	
ПК-1.2. Умеет находить, формулировать и решать стандартные задачи в собственной научно-исследовательской деятельности в области программирования и информационных технологий.	Формулирует и решает стандартные задачи в собственной научно-исследовательской деятельности в области программирования и информационных технологий.	Домашние задания Лабораторные работы Зачет
ПК-1.3. Имеет практический опыт научно-исследовательской деятельности в области программирования и информационных технологий.	Обладает практическим опытом научно-исследовательской деятельности в области программирования и информационных технологий.	Зачет

Код и формулировка компетенции: ПК-4. Способен использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования.

<b>Код и наименование индикатора достижения компетенции</b>	<b>Результаты обучения по дисциплине</b>	<b>Оценочные средства</b>
ПК-4.1. Знает современные методы разработки и реализации алгоритмов математических моделей, управления информацией и данными.	Демонстрирует знание: современных методов разработки и реализации алгоритмов математических моделей, управления информацией и данными; интеллектуальных интернет-технологий, применимых в процессе принятия решений; ИТ-средств для принятия решений: регрессионный анализ, кластерный анализ, классификация, и др.	Групповой и индивидуальный опрос Зачет
ПК-4.2. Умеет разрабатывать и реализовывать алгоритмы математических моделей на базе языков и пакетов прикладных программ моделирования, управлять информацией и данными.	Демонстрирует умение: Разрабатывать и реализовывать алгоритмы математических моделей, искать источники информации и данные, воспринимать, анализировать, передавать информацию с использованием цифровых средств и алгоритмов, проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных. Применять методы принятия решений для задач прогнозирования, разделения наблюдений на классы, как по числовым, так и по категориальным признакам, формирования клиентских	Домашние задания Лабораторные работы Зачет

	предложений, и т.п.	
ПК-4.3. Имеет практический опыт разработки и реализации алгоритмов на базе языков и пакетов прикладных программ моделирования, методами управления информацией и данными.	Обладает практическим опытом разработки и реализации алгоритмов на базе языков и пакетов прикладных программ моделирования. Владеет начальными навыками использования инструментов, поддерживающих реализацию методов машинного обучения – язык R, библиотеки Python, Excel для работы с данными.	Лабораторные работы Зачет

### 4.3. Рейтинг-план дисциплины

Рейтинг–план дисциплины представлен в Приложении 2.

#### Примерный перечень вопросов для опроса на занятиях и к экзамену.

1. Методы сбора статистических данных.
2. Методы прогнозирования.
3. Сравнение применимости различных методов прогнозирования.
4. [Дисперсионный анализ](#)
5. [Корреляционный анализ](#)
6. Дискриминантный анализ.
7. Кластерный анализ.
8. Факторный анализ.
9. Открытые интернет-инструменты для анализа данных: R, Python, Excel и др.
10. Машинное обучение. Основные понятия и области применения.
11. Основные методы машинного обучения.
12. Методы обучения с учителем (линейная, полиномиальная, логистическая регрессия, деревья решений).
13. Методы обучения без учителя (k-средних, иерархическая кластеризация).
14. Работа с моделями МО (регуляризация, снижение размерности).

#### Формы и содержание контроля:

- текущий контроль: опрос на занятиях, домашние задания;
- итоговый контроль: лабораторные работы, тест, контрольные вопросы, экзамен.

#### Критерии оценивания ответа на экзамене (в баллах):

- **25-30 баллов** выставляется студенту, если студент дал полные, развернутые ответы на все теоретические вопросы билета, продемонстрировал знание функциональных возможностей, терминологии, основных элементов, умение применять теоретические знания при выполнении практических заданий. Студент без затруднений ответил на все дополнительные вопросы. Практическая часть работы выполнена полностью без неточностей и ошибок;

- **17-24 баллов** выставляется студенту, если студент раскрыл в основном теоретические вопросы, однако допущены неточности в определении основных понятий. При ответе на дополнительные вопросы допущены небольшие неточности. При выполнении практической части работы допущены несущественные ошибки;

- **10-16 баллов** выставляется студенту, если при ответе на теоретические вопросы студентом допущено несколько существенных ошибок в толковании основных понятий. Логика и полнота ответа страдают заметными изъянами. Заметны пробелы в знании основных методов. Теоретические вопросы в целом изложены достаточно, но с пропусками

материала. Имеются принципиальные ошибки в логике построения ответа на вопрос. Студент не решил задачу или при решении допущены грубые ошибки;

- **1-10 баллов** выставляется студенту, если ответ на теоретические вопросы свидетельствует о непонимании и крайне неполном знании основных понятий и методов. Обнаруживается отсутствие навыков применения теоретических знаний при выполнении практических заданий. Студент не смог ответить ни на один дополнительный вопрос.

#### **Групповой и индивидуальный опрос.**

Вопросы приведены выше. Опрос проводится в процессе занятий. Правильные и полные ответы оцениваются в 1 балл.

### **Примерные задания для домашних и лабораторных работ**

#### **Задачи на построение прогнозных значений, анализ и визуализацию результатов в Excel:**

1. Число кондиционеров, проданных за последние 24 месяца, приведено в таблице. Проанализируйте эти данные с точки зрения применимости метода скользящего среднего, а также примените метод экспоненциального сглаживания и метод линейной регрессии. Сравните результаты.

месяц	продажа	месяц	продажа	месяц	продажа	месяц	продажа
1	25	7	85	13	40	19	105
2	15	8	88	14	35	20	85
3	30	9	60	15	50	21	60
4	38	10	40	16	60	22	55
5	58	11	40	17	66	23	50
6	62	12	38	18	90	24	45

2. В таблице содержатся данные за десятилетний период о количестве людей в тысячах, посетивших туристическую зону на автомобиле и самолете. Проанализируйте эти данные с точки зрения применимости метода скользящего среднего, а также примените метод экспоненциального сглаживания и метод линейной регрессии. Сравните результаты.

Год	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Автомобиль	1042	1182	1224	1338	1455	1613	1644	1699	1790	1885
Самолет	500	522	540	612	715	790	840	900	935	980

3. Методом Монте-Карло смоделируйте ситуацию с пятью выигрышами или проигрышами в следующей игре в кости. Игрок бросает две симметричные игральные кости. Если выпавшая сумма равна 7 или 11, игрок выигрывает 10 долларов. Иначе он запоминает выпавшую сумму (называемую очком) и продолжает бросать кости до тех пор, пока выпавшая сумма не совпадет с очком, после чего игрок получает 10 долларов. Если же выпавшая сумма равна 7, игрок проигрывает 10 долларов.

4. Цикл исполнения заказа на некоторую продукцию с равной вероятностью составляет 1 и 2 дня. Предполагается, что ежедневный спрос равен 0, 1 и 2 единицы этой продукции с вероятностями 0.2, 0.5 и 0.3 соответственно. Используйте значения случайных чисел из таблицы для оценки совместного распределения спроса и цикла исполнения заказа. Исходя из полученного совместного распределения, оцените плотность вероятности спроса в течение

цикла исполнения заказа. Спрос во время исполнения заказа может принимать значения 0, 1, 2, 3 и 4.

5. В небольшой цех с одним станком заказы поступают случайным образом. Время между поступлениями заказов распределено по экспоненциальному закону с математическим ожиданием 2 часа. Время, необходимое для выполнения заказа, является случайной величиной, равномерно распределенной на интервале  $[1, 2]$ , измеряемом в часах. Пусть первый заказ поступает в момент времени, равный нулю. Определите время поступления и выполнения первых пяти заказов, используя случайные числа из интервала  $[0, 1]$ .

6. Работы поступают в металлообрабатывающий цех в соответствии с распределением Пуассона с математическим ожиданием 6 работ в день. Цех имеет пять обрабатывающих центров, на которые контролер направляет полученные работы в строгом соответствии с очередностью. Методом сверток определите одно случайное значение интервала между получением работ на первом обрабатывающем центре.

### Построение линейной регрессии в Python

1. Загрузить данные об описаниях вакансий и соответствующих годовых зарплатах из файла `salary-train.csv`.
2. Провести его предобработку:
  - 2.1. Привести тексты к нижнему регистру (`text.lower()`).
  - 2.2. Заменить все, кроме букв и цифр, на пробелы для облегчения дальнейшего разделение текста на слова. Для такой замены в строке `text` подходит следующий вызов: `re.sub('[^a-zA-Z0-9]', ' ', text)`. Также можно воспользоваться методом `replace` у `DataFrame`, чтобы сразу преобразовать все тексты, например, так: `train['FullDescription'] = train['FullDescription'].replace('[^a-zA-Z0-9]', ' ', regex = True)`
  - 2.3. Применить `TfidfVectorizer` для преобразования текстов в векторы признаков. Оставить только те слова, которые встречаются хотя бы в 5 объектах. Для этого использовать параметр `min_df` у `TfidfVectorizer`.
  - 2.4. Заменить пропуски в столбцах `LocationNormalized` и `ContractTime` на специальную строку `'nan'` (листинг 27).
  - 2.5. Применить `DictVectorizer` для получения `one-hot`-кодирования признаков `LocationNormalized` и `ContractTime`.
  - 2.6. Объединить все полученные признаки в одну матрицу «объекты-признаки».
  - 2.7. Обучить гребневую регрессию с параметрами `alpha=1` и `random_state=241`, помня, что целевая переменная записана в столбце `SalaryNormalized`. Обучение производится вызовом метода `fit`.
3. Построить прогноз для двух примеров из файла `salary-test-mini.csv`. Прогноз строится методом `predict`.

### Построение логистической регрессии в Python. «Маркетинговая кампания банка»

**Критерии оценки (в баллах)** Каждое задание оценивается от 0 до 5 баллов в зависимости от трудоемкости.

- 4-5 баллов выставляется студенту, если задание полностью выполнено и аккуратно оформлено;
- 2-3 баллов выставляется студенту, если задание выполнено с ошибками и небрежно оформлено;
- 0-1 баллов выставляется студенту, если задание не выполнено.

**Тест по теме «Машинное обучение» (выбор из 4х вариантов)**

1. **Какая из задач не относится к типичным постановкам задач для машинного обучения:**
  - 1) Кластеризация
  - 2) Трансгрессия
  - 3) Регрессия
  - 4) Классификация
  
2. **Укажите взаимосвязь между недообучением и переобучением:**
  - 1) При усложнении модели, переобучение переходит в недообучение
  - 2) При упрощении модели, недообучение переходит в переобучение
  - 3) При усложнении модели, недообучение переходит в переобучение
  - 4) Модель может быть одновременно недообучена и переобучена
  
3. **Для чего используется регрессия в машинном обучении?**
  - 1) Формализация знаний экспертов и их перенос в компьютер в виде базы знаний.
  - 2) Обнаружение в обучающей выборке небольшого числа нетипичных объектов.
  - 3) Поиск набора признаков их значений, которые особенно часто встречаются в признаковых описаниях объектов.
  - 4) Предсказание каких-либо значений по набору признаков.
  
4. **В каком случае применять метод наименьших квадратов затруднительно?**
  - 1) Данные не помещаются в память
  - 2) В записях мало предикторов
  - 3) В данных мало записей
  - 4) Минимальная ошибка не равна нулю
  
5. **Для обучения модели линейной регрессии обычно используется**
  - 1) Среднеквадратическая ошибка
  - 2) Средняя абсолютная ошибка
  - 3) Перекрёстная энтропия
  - 4) Стандартное отклонение
  
6. **Регрессия с несколькими предикторами называется:**
  - 1) Многочисленной
  - 2) Множественной
  - 3) Разносторонней
  - 4) Разнообразной
  
7. **При увеличении степени полиномиальной регрессии, часто возникает проблема**
  - 1) Низкой точности на обучающей выборке
  - 2) Переобучения
  - 3) Недообучения
  - 4) Нехватки памяти компьютера

**8. Какие ограничения приходится накладывать на решение методом кластеризации?**

- 1) Использовать лишь некоторые метрики качества.
- 2) Использовать метрики, которые позволяют отклонять заведомо “плохие” варианты.
- 3) Использовать ограниченную модель.
- 4) Использовать итеративный подход.

**9. Какой подход называется обучением с учителем?**

- 1) Когда модели предоставляются входные данные и правильный ответ.
- 2) Когда нерегулярно происходит “наказание” модели за неправильный результат.
- 3) Ничего из вышеперечисленного.
- 4) Когда нерегулярно происходит “вознаграждение” модели за правильный результат.

**10. Какие бывают метрики машинного обучения?**

- 1) Площадь под кривой.
- 2) Точность.
- 3) Аккуратность.
- 4) Чувствительность.
- 5) Правильность.
- 6) Площадь под прямой.

**Критерии оценки (в баллах)**

Каждое задание оценивается от 0 до 1 баллов. Правильный ответ на вопрос тестового задания оценивается в 1 балл, неверный ответ – 0 баллов.

Количество вопросов в тесте 10. Максимальный балл – 10.

**5. Учебно-методическое и информационное обеспечение дисциплины**

**5.1. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины**

**Основная литература:**

1. Балдин, К.В. Теория вероятностей и математическая статистика [Электронный ресурс] : учебник / К.В. Балдин, В.Н. Башлыков, А.В. Рукосуев. — Электрон. дан. — Москва : Дашков и К, 2016. — 472 с. — Режим доступа: <https://e.lanbook.com/book/72438>.
2. Кобзарь, А.И. Прикладная математическая статистика. Для инженеров и научных работников [Электронный ресурс] : учебное пособие / А.И. Кобзарь. — Электрон. дан. — Москва : Физматлит, 2012. — 816 с. — Режим доступа: <https://e.lanbook.com/book/59747>.
3. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2015 – 400 с.

**Дополнительная литература:**

1. Грешилов, А.А. Компьютерные обучающие пособия для решения задач математической статистики и математического программирования [Электронный ресурс] : учебное пособие / А.А. Грешилов. — Электрон. дан. — Москва : МГТУ им. Н.Э. Баумана, 2013. — 191 с. — Режим доступа: <https://e.lanbook.com/book/106525>.

2. Самсонова, С.А. Практикум по математической статистике [Электронный ресурс] : учебное пособие / С.А. Самсонова. — Электрон. дан. — Архангельск : САФУ, 2015. — 97 с. — Режим доступа: <https://e.lanbook.com/book/96567>.
3. Верховный алгоритм: как машинное обучение изменит наш мир / Педро Домингос ; пер. с англ. В. Горохова— М. : Манн, Иванов и Фербер, 2016 — 336 с.
4. «Python и машинное обучение» Себастьян Рашка, 2017. <https://cdnpdf.com/embed/38379-python-i-mashinnoe-obuchenie-2017-rashka-s#dfliip-flipbookContainer/7/>
5. Джеймс Г., Уиттон Д., Хастис Т., Тибширани Р. Введение в статистическое обучение с примерами на языке R <https://djvu.online/file/hEs2yql45ir8q>

## **5.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет» и программного обеспечения, необходимых для освоения дисциплины, включая профессиональные базы данных и информационные справочные системы**

Информационно-образовательные ресурсы в сети «Интернет»

1. Открытые информационные научные ресурсы ведущих научных центров и научных журналов.
2. Международный электронный архив научных статей <http://arxiv.org/>.
3. Открытый образовательный видеопортал UniverTV.ru. Образовательные фильмы на различные темы. Лекции в ведущих российских и зарубежных вузах. Научная конференция или научно-популярная лекция по интересующему вас вопросу. <http://univertv.ru/video/matematika/>
4. Научная электронная библиотека eLIBRARY.RU. Крупнейший российский информационный портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты более 12 млн. научных статей и публикаций. На платформе eLIBRARY.RU доступны электронные версии более 1400 российских научно-технических журналов, в том числе более 500 журналов в открытом доступе. <http://elibrary.ru>
5. Общероссийский математический портал. <http://www.mathnet.ru>
6. Информационно-аналитический центр по параллельным вычислениям. <http://parallel.ru/>
7. <https://www.tadviser.ru/index.php/>  
[Статья:Сквозные технологии цифровой экономики#](#)
8. Лучшие системы для анализа данных. <https://soware.ru/categories/data-mining-systems>
9. Большие данные: бесплатные источники. <https://habr.com/ru/post/510082/>

Перечень программного обеспечения и информационных справочных систем:

- ЭБС «Университетская библиотека онлайн»;
  - ЭБС издательства «Лань»;
  - ЭБС «Электронный читальный зал»;
  - БД периодических изданий на платформе EastView: «Вестники Московского университета», «Издания по общественным и гуманитарным наукам»;
  - Научная электронная библиотека;
  - БД диссертаций Российской государственной библиотеки.
- Также доступны следующие зарубежные научные ресурсы баз данных:
- Web of Science;
  - Scopus;
  - Издательство «Taylor&Francis»;
  - Издательство «Annual Reviews»;

- «Computers & Applied Sciences Complete» (CASC) компании «EBSCO»
- Архивы научных журналов на платформе НЭИКОН (Cambridge University Press, SAGE Publications, Oxford University Press);
- Информационная система «Единое окно доступа к образовательным ресурсам» (<http://window.edu.ru>);
- справочно-правовая система Консультант Плюс;
- справочно-правовая система Гарант.

**6. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине**

<i>Наименование специализированных аудиторий, кабинетов, лабораторий</i>	<i>Вид занятий</i>	<i>Наименование оборудования, программного обеспечения</i>
<i>1</i>	<i>2</i>	<i>3</i>
<i>Аудитория 501, 502, 528, 530, 531, 515</i>	<i>Лекции</i>	<i>Мультимедийный проектор, экран, доска</i>
<i>Аудитория 531, 527, 526, 503, 511, 523, 509</i>	<i>Лабораторные работы</i>	<i>Мультимедийный проектор, экран, доска</i>
<i>Компьютерный класс 426, 520а, 521, 522, 525</i>	<i>Самостоятельная работа</i>	<i>Компьютеры, имеющие выход в сеть Internet, имеющие необходимое программное обеспечение: пакет MS Office. Электронная библиотека</i>

Приложение № 1

ФГБОУ ВО «БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
ФАКУЛЬТЕТ МАТЕМАТИКИ И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

**СОДЕРЖАНИЕ РАБОЧЕЙ ПРОГРАММЫ**



дисциплины Прогнозирование и многомерный статистический анализ

на 7 семестр  
(наименование дисциплины)

очная  
форма обучения

<b>Вид работы</b>	<b>Объем дисциплины</b>
Общая трудоемкость дисциплины (ЗЕТ / часов)	216/6
Учебных часов на контактную работу с преподавателем:	
лекций	36
практических/ семинарских лабораторных	36
ФКР	1,2
Учебных часов на самостоятельную работу обучающихся (СР)	108
Учебных часов на подготовку к экзамену (Контроль)	34,8

Форма(ы) контроля:  
экзамен 7 семестр

№ п/п	Тема и содержание	Форма изучения материалов: лекции, практические занятия, семинарские занятия, лабораторные работы, самостоятельная работа и трудоемкость (в часах)				Основная и дополнительная литература, рекомендуемая студентам (номера из списка)	са ра
		ЛК	ПР/СЕМ	ЛР	СР		
1	2	3	4	5	6	7	
1.	Методы сбора статистических данных. Методы прогнозирования.	4		6	12	[1],[2] Доп.литра [1], [2]	Про лек мат лит ист Вып дом зад
2.	Методы статистического анализа. Дисперсионный анализ. Корреляционный анализ Дискриминантный, кластерный, факторный анализ.	4			12	[1],[2] Доп.литра [1], [2]	Про лек мат лит ист Вып дом зад
3.	Сквозные технологии цифровой экономики				8	Информационно- образовательные ресурсы в сети «Интернет» [7]	Изу скв тех хар суб

							при эко.
4.	ИИ в принятии решений				8	Информационно-образовательные ресурсы в сети «Интернет» [7]	Изу сов, сре, реш при Озн дос обл реш пер раз
5.	Открытые системы для анализа данных	4		4	20	Доп. лит-ра [4-5] Информационно-образовательные ресурсы в сети «Интернет» [8-9]	Озн док осн фун воз биб Rutl
6.	Машинное обучение. Основные понятия и области применения	6		4	12	[1],[2], [3] Доп. лит-ра [1], [2], [3]	Про лек мат лит ист Вып дом зад
7.	Методы обучения с учителем (линейная, полиномиальная, логистическая регрессия, деревья решений).	6		6	12	[1],[2], [3] Доп. лит-ра [1], [2], [3]	Про лек мат лит ист

							Вып дом зада
8.	Методы обучения без учителя (к-средних, иерархическая кластеризация).	6		8	12	[1],[2], [3] Доп. лит-ра [1], [2], [3]	Пр лек мат лит ист Вып дом зада
9.	Работа с моделями МО (регуляризация, снижение размерности).	6		8	12	[1],[2], [3] Доп. лит-ра [1], [2], [3]	Пр лек мат лит ист Вып дом зада
	Экзамен			1,2	34,8	[1],[2] Доп.литра [1], [2]	Пр лек мат лит ист
	<b>Всего часов:</b>	36		37,2	142,8		

**Рейтинг-план дисциплины****Прогнозирование и многомерный статистический анализ***(название дисциплины согласно рабочему учебному плану)*

направление подготовки Направление 01.03.02 Прикладная математика и информатика

курс 4, семестр 1(7)

Виды учебной деятельности студентов	Балл за конкретное задание	Число заданий за семестр	Баллы	
			Минимальный	Максимальный
<b>Модуль 1.</b>				
<b>Методы сбора статистических данных и прогнозирования</b>			<b>0</b>	<b>50</b>
<b>Текущий контроль</b>				
1. Аудиторная работа	2	5		<b>10</b>
2. Домашняя работа	3	5		<b>15</b>
<b>Рубежный контроль</b>				
1. Лабораторная работа	5	5		<b>25</b>
<b>Модуль 2.</b>				
<b>Анализ данных методами машинного обучения</b>			<b>0</b>	<b>50</b>
<b>Текущий контроль</b>				
1. Аудиторная работа	2	5		<b>10</b>
2. Домашняя работа	3	5		<b>15</b>
<b>Рубежный контроль</b>				
1. Лабораторная работа	5	5		<b>25</b>
<b>Поощрительные баллы</b>				
1. Студенческая олимпиада или конкурс рефератов				<b>10</b>
2. Публикация статей				
3. Работа со школьниками (кружок, конкурсы, олимпиады)				
<b>Посещаемость (баллы вычитаются из общей суммы набранных баллов)</b>				
1. Посещение лекционных занятий			<b>0</b>	<b>-6</b>
2. Посещение практических (семинарских, лабораторных занятий)			<b>0</b>	<b>-10</b>
<b>Итоговый контроль</b>				
Зачет			<b>60</b>	<b>100</b>