

**РАБОЧАЯ ПРОГРАММА**  
модуля (курса)  
**«Анализ данных и элементы машинного обучения»**

## 1. АННОТАЦИЯ

Дисциплина «Анализ данных и элементы машинного обучения» включает учебный материал, способствующий формированию у слушателей теоретических знаний и практических навыков по методологии обработки данных средствами языка Python.

Цель дисциплины – освоение методов проведения полного цикла работ по анализу данных от сбора данных до интерпретации полученных результатов и подготовки соответствующих отчетов. Дисциплина рассматривается, как один из основных курсов программы профессиональной переподготовки «Анализ данных для прикладных областей».

## 2. СОДЕРЖАНИЕ

№ п/п	Наименование модуля, разделов и тем	Содержание обучения (по темам в дидактических единицах), наименование и тематика лабораторных работ, практических занятий (семинаров), самостоятельной работы с указанием кол-ва часов, используемых образовательных технологий и рекомендуемой литературы
1.	2.	3.
1	Лекции	8 ч
2	Тема 1. Роль математической статистики и теории вероятностей при анализе данных, средства статистического анализа данных	Типы статистических данных (числовые, ординальные, номинальные). Генеральная совокупность, выборка. Репрезентативность выборки. Способы обеспечения сопоставимости данных. Обзор существующих открытых библиотек данных. 2 ч
3	Тема 2. Подготовка данных к статистическому анализу	Одномерные и многомерные данные. Понятие многомерной случайной величины. Эмпирическая плотность и эмпирическая функция распределения, примеры их построения. Визуализация эмпирических распределений: построение гистограмм, графиков функций распределения. 2 ч
4	Тема 3. Возможности визуального анализа данных	Числовые характеристики одномерных случайных величин. Статистические числовые характеристики. Анализ данных на основе статистических числовых характеристик: характеристик центрального положения, разброса. 1 ч
5	Тема 4. Статистические числовые характеристики многомерных данных	Числовые характеристики многомерной случайной величины. Ковариация, коэффициент корреляции. Построение ковариационной матрицы. Корреляционный анализ.. Выявление зависимости между величинами на основе ковариационной матрицы. 1 ч
6	Тема 5. Выявление линейной зависимости. Элементы корреляционного и регрессионного анализа	Простая линейная регрессия. Многомерная регрессия. Полиномиальная регрессия. Построение простой и многомерной линейной регрессии. 1 ч

7	Тема 6. Кластерный анализ (кластеризация) данных	Кластеризация данных, задача и методы решения (метод к-средних, иерархическая кластеризация). Определение оптимального количества кластеров. 0,5 ч
8	Тема 7. Оценивание неизвестных параметров закона распределения данных	Оценивание неизвестных параметров распределения. Точечные оценки. Параметрическое и непараметрическое оценивание. Метод максимального правдоподобия и метод моментов. 0,5 ч
9	Практические занятия (семинары)	Библиотеки данных для анализа. Начало работы с Python. Чтение и обработка Excel файлов средствами Python . 2 ч
		Визуализация эмпирических распределений средствами Python. 2 ч
		Анализ многомерных данных: выявление зависимости. 2 ч
		Инструменты Python для построения регрессии. 2 ч
		Средства Python кластеризации данных. 4 ч
		Оценивание неизвестных параметров закона распределения данных. 4 ч
10	Самостоятельная работа	Выполнение домашних заданий по теме занятия 30 ч
11	Промежуточная аттестация	Лабораторная работа 2 ч

### 3. ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ МОДУЛЯ

Промежуточная аттестация представляет собой собеседование, которое проводится по результатам практических занятий (семинаров). Для сдачи зачета слушатель должен сделать все практические задания и дать ответ на два вопроса из нижеперечисленных вопросов по курсу:

- Перечислить способы представления выборочных значений.
- Назовите основные библиотеки Python, используемые для сбора, анализа данных.
- Дать определение одномерных и многомерных данных.
- Дать определение статистической (выборочной, эмпирической) функции распределения
- С помощью каких графиков в Python визуализируются многомерные данные?
- Определить, основные статистические числовые характеристик (выборочное среднее, дисперсию и стандарт)
- Привести способы вычисления статистической ковариации и выборочного коэффициента корреляции, определить понятие несмещенной выборочной ковариации
- Какая библиотека в Python содержит функции для вычисления ковариации и коэффициента корреляции
- Перечислить способы выявления статистическая зависимости двух случайных величин
- Определить простую линейную регрессию. Перечислить Ключевые различия между корреляцией и линейной регрессией
- Описать метод использования регрессии для прогнозирования. Определить, что такое подогнанные значения и остатки
- Какими средствами в Python строится линейная регрессия?
- Перечислить способы диагностики качества регрессионной модели
- Какими средствами в Python проверяется качество построенной регрессионной модели?
- Перечислить критерии качества статистических оценок

- Описать Метод моментов
- Описать Метод максимального правдоподобия
- Какими средствами реализуются в Python реализуется метод моментов?
- Описать метод интервального оценивания неизвестных параметров распределения.
- Описать методологию построения доверительных интервалов для математического ожидания
- Описать методологию построения доверительных интервалов для неизвестной дисперсии
- Описать методологию построения доверительных интервалов для неизвестной вероятности события

#### Формы и методы контроля и оценки результатов освоения модуля

№ п/п	Наименование процедуры	Основные показатели оценки	Формы и методы контроля и оценки
1	Промежуточный контроль. Модуль 4. Анализ данных и элементы машинного обучения	Владеет навыками проведения полного цикла работ по анализу данных от сбора данных до интерпретации полученных результатов и подготовки соответствующих отчетов	Дифференцированный зачет / Лабораторная работа

#### Критерии оценки

№ п/п	Наименование процедуры	Основные показатели оценки		Формы и методы контроля и оценки
1	Промежуточный контроль. Модуль 4. Анализ данных и элементы машинного обучения	Зачтено	Отлично. Высокий уровень подготовки, безупречное владение теоретическим материалом, слушатель демонстрирует творческий подход к решению нестандартных ситуаций. Слушатель дал полный и развернутый ответ на все теоретические вопросы билета, подтверждая теоретический материал практическими примерами из практики. Слушатель активно работал на практических занятиях.	Дифференцированный зачет / Лабораторная работа
			Хорошо. В целом хорошая подготовка с заметными ошибками или недочетами. Слушатель дает полный ответ на все теоретические вопросы билета, но имеются неточности в определениях понятий, процессов и т.п. Допускаются ошибки при ответах на дополнительные и уточняющие вопросы экзаменатора. Слушатель работал на практических занятиях.	
			Удовлетворительно. Минимально достаточный уровень подготовки. Слушатель показывает минимальный уровень теоретических знаний, делает существенные ошибки при характеристике нормативно-	

			правовой базы бухгалтерского учета, но при ответах на наводящие вопросы, может правильно сориентироваться и в общих чертах дать правильный ответ. Слушатель посещал практические занятия.	
		Не зачтено	Неудовлетворительно. Подготовка недостаточная и требует дополнительного изучения материала. Слушатель дает ошибочные ответы, как на теоретические вопросы билета, так и на наводящие и дополнительные вопросы экзаменатора. Слушатель пропустил большую часть практических занятий.	

#### 4. УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ МОДУЛЯ

##### 4.1 Учебно-методическое и информационное обеспечение программы:

Для эффективного освоения компетенций, формируемых учебной дисциплиной важно использование в учебном процессе активных и интерактивных форм проведения занятий.

Изучение учебной дисциплины предполагает наличие аудиторной и самостоятельной видов работ слушателей. В ходе практических занятий рассматриваются практические задачи из практики с целью наиболее полного овладения умениями и навыками.

Лекции по учебной дисциплине призваны формировать знания, предусмотренные учебной программой, и включают теоретическую базу обработки данных, на базе которой строятся прикладные аспекты.

Наряду с проработкой основной литературы (глав базового учебника) предусмотрено самостоятельное чтение дополнительной литературы (статей и других научных публикаций).

Практические занятия в малых группах и самостоятельная внеаудиторная работа направлены на выработку навыков анализа данных.

Для достижения поставленных целей преподавания дисциплины реализуются следующие средства, способы и организационные мероприятия:

- изучение теоретического материала дисциплины на лекции с использованием компьютерных технологий;
- самостоятельное изучение теоретического материала дисциплины с использованием Internet-ресурсов, информационных баз, электронных библиотек, методических разработок, специальной и научной литературы;
- закрепление теоретического материала при проведении практических занятий с использованием учебного и научного оборудования, выполнения проблемно-ориентированных, поисковых, творческих заданий.

Самостоятельная работа слушателей включает:

1. Изучение учебной литературы по курсу.
2. Решение практических ситуаций и задач
3. Изучение источников управленческой информации
4. Работу с ресурсами Интернет
5. Решение практических ситуаций в виде творческих заданий
6. Изучение практических материалов деятельности конкретных предприятий
7. Изучение статистической информации
8. Подготовку к экзамену по курсу «Анализ данных и элементы машинного обучения».

Цель самостоятельной работы - подготовка современного компетентного специалиста и формирование способностей и навыков к непрерывному самообразованию и профессиональному совершенствованию.

##### 4.2. Содержание комплекта учебно-методических материалов.

Слушателям предоставляются презентации и раздаточный материал.

### 4.3. Используемые образовательные технологии. Краткое описание.

Лекции и практические занятия проводятся с использованием возможностей мультимедийного класса. Использование в учебном процессе активных и интерактивных форм проведения занятий (компьютерных симуляций, деловых и ролевых игр, разбор конкретных ситуаций).

### 4.4. Литература

#### Основная литература

1. Федоткин М.А. Основы прикладной теории вероятностей и статистики. — М.: Высшая школа. 2006. - 168 с.
2. Теория вероятностей и математическая статистика. Авторы: Федоткин М.А., Пройдакова Е.В.: Электронный управляемый курс. – Нижний Новгород: Нижегородский госуниверситет, 2014. (Идентификационный номер в электронном каталоге фонда электронных образовательных ресурсов ННГУ: 789Е.14.08)
3. Золотых Н.Ю. Машинное обучение. Курс лекций. Нижний Новгород: ННГУ, 2007. <http://www.uic.nnov.ru/~zny/m>
4. Курс «Машинное обучение» – <https://www.intuit.ru/studies/courses/13844/1241/info>
5. Практическая статистика для специалистов Data Science: Пер. с англ./ П. Брюс, Э. Брюс. — СПб.: БХВ-Петербург, 2018..

#### Программное обеспечение и Интернет-ресурсы

1. Операционная система Microsoft Windows.
2. Open-source среда Spyder.
3. Веб-интерактивная вычислительная среда Jupyter Notebook (для поддержки языка Рут

### 4.5. Материально-технические условия реализации программы:

#### Материально-техническая база

№ п.п.	Наименование модуля (тем, разделов)	Материально-технические условия для реализации программ (наличие лабораторий, производственных участков и т.п. по профилю программы профессиональной переподготовки)
1	Тема 1. Роль математической статистики и теории вероятностей при анализе данных, средства статистического анализа данных	Реализация дисциплины предполагает наличие: - аудиторий для лекционных и практических занятий с необходимым мультимедийным оборудованием; - The R Project for Statistical Computing <a href="https://www.r-project">https://www.r-project</a> - Welcome to Python.org <a href="https://www.python.org/">https://www.python.org/</a> - scikit-learn: machine learning in Python <a href="https://scikit-learn.org/">scikit-learn.org/</a> . В ходе проведения занятий рекомендуется использовать компьютерные иллюстрации для поддержки различных видов занятий, подготовленные с использованием MicrosoftOffice или других средств визуализации материала.
2	Тема 2. Подготовка данных к статистическому анализу	
3	Тема 3. Возможности визуального анализа данных	
4	Тема 4. Статистические числовые характеристики многомерных данных	
5	Тема 5. Выявление линейной зависимости. Элементы корреляционного и регрессионного анализа	
6	Тема 6. Кластерный анализ (кластеризация) данных	
7	Тема 7. Оценивание неизвестных параметров закона распределения данных	
8	Практические занятия (семинары)	

