

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ
ФЕДЕРАЦИИ

ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ

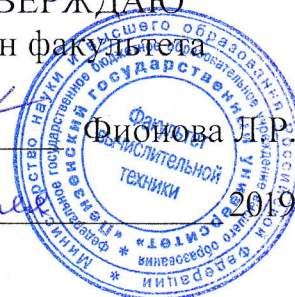
УТВЕРЖДАЮ
Декан факультета



Феонова Д.Р.

« 08 » июля

2019 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.ДВ.01.02 Статистические основы анализа больших данных

Направление подготовки 01.03.02 — «Прикладная математика и информатика»

Направленность (профиль подготовки) - «Компьютерные технологии»

Квалификация выпускника – бакалавр

Форма обучения – очная

Пенза, 2019

1. Цели освоения дисциплины

Целями освоения дисциплины «Статистические основы анализа больших данных» являются теоретическая и практическая подготовка студентов к анализу и обработке больших данных, т.е. данных, которые не поддаются или очень трудно поддаются обработке традиционными методами. Сюда относятся структурированные (традиционный формат БД) и неструктурированные данные, медиа и случайные объекты.

Задачи освоения дисциплины:

- приобретение студентами знаний о технологиях подготовки, хранения, обработки и анализа больших данных;
- применение математических и статистических методов для анализа больших объемов информации.

2. Место дисциплины в структуре ОПОП бакалавриата

Учебная дисциплина «Статистические основы анализа больших данных» в учебном плане содержится в части формируемой участниками образовательных отношений Блока 1 «Дисциплины (модули)» ОПОП. Данная дисциплина имеет логическую и содержательно-методологическую взаимосвязь с другими частями ОПОП, так как углубляет и закрепляет математические и естественнонаучные знания и навыки, сформированные в результате изучения дисциплин базовой части.

Изучение данной учебной дисциплины базируется на знании дисциплин: «Математический анализ», «Алгоритмы и алгоритмические языки» «Теория вероятностей», «Линейная алгебра и аналитическая геометрия».

Основные положения дисциплины должны быть использованы при изучении дисциплин: «Статистический анализ данных», «Искусственный интеллект», «Моделирование случайных процессов», «Интеллектуальный анализ данных», «Технологии больших данных».

3. Результаты освоения дисциплины «Статистические основы анализа больших данных»

Процесс изучения дисциплины направлен на формирование элементов следующих компетенций в соответствии с ФГОС ВО по данному направлению:

Коды компетенции	Наименование компетенции	Индикатор достижения компетенции (закрепленный за дисциплиной)	В результате освоения дисциплины обучающийся должен:
1	2	3	4
УК-1	Способен осуществлять поиск, критический анализ информации, применять системный подход для решения поставленных задач	УК-1.3. Осуществляет поиск информации для решения поставленной задачи по различным типам запросов	<p>Знать: современные средства анализа и обработки больших данных</p> <p>Уметь: проводить сравнительный анализ выбирать оптимальные средства для обработки большого объема информации.</p> <p>Владеть: использование современных информационных технологий, в том числе программных средств, для обработки и анализа больших неструктурированных объемов данных</p>

4. Структура и содержание дисциплины «Статистические основы анализа больших данных»

4.1. Структура дисциплины «Статистические основы анализа больших данных»

Общая трудоемкость дисциплины составляет 3 зачетные единицы, 108 часов.

№ п/п	Наименование разделов и тем дисциплины (модуля)	Семестр	Недели семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)							Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)		
				Контактная работа					Самостоятельная работа		Собеседование	Коллоквиум	Контроль выполнения домашнего задания
				Всего	Лекции	Практические занятия	Лабораторные занятия	Др. виды контакт. работы	Всего	Подготовка к аудиторным занятиям			
1.	Раздел 1. Обзор различных методов обработки статистической информации.	4	1-2	4	2	-	2		12	12			
1.1.	Тема 1.1. Виды статистических данных: количественные и категоризованные данные. Проблема и источники получения больших данных (БД).	4	1-2	4	2	-	2		12	12			
2.	Раздел 2. Методы анализа БД.	4	3-17	30	15	-	15						
2.1.	Тема 2.1. Data Mining: обучение ассоциативным правилам	4	3-4	4	2	-	2		12	12			
2.2.	Тема 2.2. Data Mining: классификация (методы категоризации новых данных)	4	5-6	4	2	-	2		12	12			

	на основе принципов, ранее применённых к уже наличествующим данным)													
3.	Тема 2.3. Data Mining: кластерный анализ	4	7-10	8	4	-	4		12	12				
3.2.	Тема 2.4. Data Mining: регрессионный анализ;	4	11-14	8	4	-	4		12	12				
3.4	Тема 2.5. многомерный статистический анализ	4	15-17	6	3	-	3		12.05	12.05				
	<i>Другие виды контактной работы</i>							1.95						
	Общая трудоемкость, в часах			35.95	17	-	17	1.95	72.05	72.05		Промежуточная аттестация		
											Форма	Семестр		
											Зачет	4		

4.2. Содержание дисциплины

№ п/п	Наименование раздела дисциплины	Содержание раздела
1.	Обзор различных методов обработки статистической информации.	Виды статистических данных: количественные и категоризованные данные. Проблема и источники получения больших данных (БД).
2.	Методы и технологии анализа БД.	-Data Mining: обучение ассоциативным правилам, классификация (методы категоризации новых данных на основе принципов, ранее применённых к уже наличествующим данным), кластерный анализ, регрессионный анализ; -многомерный статистический анализ; - R(язык программирования); -пакет Statistica.

5. Образовательные технологии

В процессе изучения дисциплины «Статистические основы анализа больших данных» предполагается использовать структурно-логические и интеграционные образовательные технологии, реализуемые посредством:

- лекций в виде вводных, текущих, обзорных и заключительно-обобщающих занятий;
- лабораторных работ и их защиты в виде собеседования;
- организации самостоятельной работы на основе лично-дифференцированного подхода планирования задания в виде воспроизводящей и частично-поисковой работ.
- организации текущего контроля знаний студентов методами: выполнения домашних заданий, оценки активности на практических занятиях и рейтинговой системы общей оценки знаний студентов.

Занятия, проводимые в интерактивных формах, с использованием интерактивных технологий составляют 30% занятий.

В целях реализации индивидуального подхода к обучению студентов, осуществляющих учебный процесс по собственной траектории в рамках индивидуального рабочего плана, изучение данной дисциплины базируется на следующих возможностях: обеспечение внеаудиторной работы со студентами в том числе в электронной образовательной среде с использованием соответствующего программного оборудования, дистанционных форм обучения, возможностей интернет-ресурсов, индивидуальных консультаций и т.д.

Другие виды контактной работы: проведение консультаций, прием зачетов.

**6. Учебно-методическое обеспечение самостоятельной работы студентов.
Оценочные средства для текущего контроля успеваемости,
промежуточной аттестации по итогам освоения дисциплины.**

6.1. План самостоятельной работы студентов

№ нед.	Тема	Вид самостоятельной работы	Задание	Рекомендуемая литература	Количество часов
1	Тема 1.1. Виды статистических данных: количественные и категоризованные данные. Проблема и источники получения больших данных (БД).	Подготовка к лабораторным занятиям, изучение литературы, выполнение лабораторной работы №1	Изучение языка программирования R, изучение работы среды R-Studio.	П.7 а 1-5	12
2	Тема 2.1. Data Mining: обучение ассоциативным правилам	Подготовка к аудиторным занятиям, изучение литературы, решение задач	Подробно изучить тему 1.2. Решить задание №2.	П.7 а 1-5	12
3	Тема 2.2. Data Mining: классификация (методы категоризации новых данных на основе принципов, ранее применённых к уже наличествующим данным)	Подготовка к аудиторным занятиям, изучение литературы, решение задач	Подробно изучить тему 1.3. Решить задание №3.	П.7 а 1-5	12
4	Тема 2.3. Data Mining: кластерный анализ	Подготовка к аудиторным занятиям, изучение литературы, решение задач	Подробно изучить тему 1.4. Решить задание №4.	П.7 а 1-5	12
5	Тема 2.4. Data Mining: регрессионный анализ	Подготовка к аудиторным занятиям, изучение литературы, решение задач	Подробно изучить тему 1.5. Решить задание №5.	П.7 а 1-5	12
6	Тема 2.5. многомерный статистический анализ	Подготовка к аудиторным занятиям, изучение	Подробно изучить тему 1.6. Решить	П.7 а) 1-5	12.5

		литературы, решение задач	задание №6.		
--	--	------------------------------	-------------	--	--

Типовое задание к лабораторной работе №1

Расчеты коэффициента корреляции и уравнения регрессии в пакете Statistica

1. Построить две группы данных.
2. Вычислить коэффициент корреляции между величинами X и Y. Сделать вывод о зависимости случайных величин.
3. Определить коэффициенты a, b уравнения регрессии $y = ax + b$.
4. Построить график функции. Сделать выводы.
5. Сделайте прогноз о поведении функции Y.

Типовое задание к лабораторной работе №2

Проведения многофакторного регрессионного анализа.

1. Получить описательные статистики по каждому признаку.
2. Составить уравнение множественной регрессии, оценить его параметры пояснить их экономический смысл.
3. Проанализировать линейные коэффициенты парной и частной корреляции.
4. Оценить значения линейных коэффициентов множественной корреляции.
5. С помощью F-критерия Фишера оценить статистическую надежность уравнения регрессии в целом.

Типовое задание к лабораторной работе №3

Кластерный анализ для определения и получения однородных групп (кластеров) по методу Уорда.

1. Сформировать структуру данных объекта исследования
2. Провести анализ Уорда для выбранного объекта

Варианты заданий к лабораторным работам

1. Кластеризация рынка автомобилей
2. Кластеризация рынка ценных бумаг
3. Кластеризация рынка инвестиционных фондов
4. Кластеризация пользователей ВК
5. Кластеризация языков программирования
6. Своя тема

6.2. Методические указания по организации самостоятельной работы студентов

- Подготовка к аудиторным занятиям проводится посредством изучения курса лекций, литературы, и поставленных задач в указанных пакетах
- Подготовка к зачету – изучение курса лекций, упражнения в решении типовых задач, изучение дополнительной литературы.

6.3. Материалы для проведения текущего и промежуточного контроля знаний студентов

Контроль освоения компетенций

№ п\п	Вид контроля	Контролируемые темы (разделы)	Компетенции, компоненты которых контролируются
1	Собеседование/защита лабораторной работы	Разделы 1,2.	УК-1

Материалы для проведения текущего контроля знаний и промежуточной аттестации составляют отдельный документ – Фонд оценочных средств по дисциплине «Статистические основы анализа больших данных».

Демонстрационные варианты оценочных средств для каждого вида контроля можно посмотреть <http://moodle.pnzgu.ru> в разделе дисциплины

7. Учебно-методическое и материально-техническое обеспечение дисциплины «Статистические основы анализа больших данных»

а) Литература

1. Гмурман В.Е. Теория вероятностей и математическая статистика: учебное пособие для ВУЗов - М.: Высш. обр., 2009 – 479 с. Библ. ПГУ 20 экз
http://kleopatra.pnzgu.ru/cgi-bin/irbis64r_91/cgiirbis_64.exe?P21DBN=KATL&I21DBN=KATL_PRINT&S21FMT=fullw_print&C21COM=F&Z21MFN=5269
- М.: Высш. обр., 2008 – 479 с. Библ. ПГУ 30 экз.
http://kleopatra.pnzgu.ru/cgi-bin/irbis64r_91/cgiirbis_64.exe?P21DBN=KATL&I21DBN=KATL_PRINT&S21FMT=fullw_print&C21COM=F&Z21MFN=9638
– М.: Высш. обр., 2007 – 479 с. Библ. ПГУ 48 экз.
http://kleopatra.pnzgu.ru/cgi-bin/irbis64r_91/cgiirbis_64.exe?P21DBN=KATL&I21DBN=KATL_PRINT&S21FMT=fullw_print&C21COM=F&Z21MFN=8138
– М.: Высш. обр., 2005 – 479 с. Библ. ПГУ 95 экз.
http://kleopatra.pnzgu.ru/cgi-bin/irbis64r_91/cgiirbis_64.exe?P21DBN=KATL&I21DBN=KATL_PRINT&S21FMT=fullw_print&C21COM=F&Z21MFN=5279
2. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учебное пособие. –
- М., Высшее образование, 2009– 404 с. Библ. ПГУ 19 экз.
http://kleopatra.pnzgu.ru/cgi-bin/irbis64r_91/cgiirbis_64.exe?P21DBN=KATL&I21DBN=KATL_PRINT&S21FMT=fullw_print&C21COM=F&Z21MFN=11392
- М., Высшее образование, 2007– 404 с. Библ. ПГУ 25 экз.
http://kleopatra.pnzgu.ru/cgi-bin/irbis64r_91/cgiirbis_64.exe?P21DBN=KATL&I21DBN=KATL_PRINT&S21FMT=fullw_print&C21COM=F&Z21MFN=7601
- М., Высшее образование, 2006– 404 с. Библ. ПГУ 14 экз.
http://kleopatra.pnzgu.ru/cgi-bin/irbis64r_91/cgiirbis_64.exe?P21DBN=KATL&I21DBN=KATL_PRINT&S21FMT=fullw_print&C21COM=F&Z21MFN=7374
- М., Высшее образование, 2005– 404 с. Библ. ПГУ 92 экз.
http://kleopatra.pnzgu.ru/cgi-bin/irbis64r_91/cgiirbis_64.exe?P21DBN=KATL&I21DBN=KATL_PRINT&S21FMT=fullw_print&C21COM=F&Z21MFN=5269
http://kleopatra.pnzgu.ru/cgi-bin/irbis64r_91/cgiirbis_64.exe?P21DBN=KATL&I21DBN=KATL_PRINT&S21FMT=fullw_print&C21COM=F&Z21MFN=13285
3. **Горелова, Галина Викторовна.**
Теория вероятностей и математическая статистика в примерах и задачах с применением

Excel : учебное пособие / Г. В. Горелова. - 4-е изд. - Ростов н/Д : Феникс, 2006. - 475 с. - (Высшее образование). Библ. ПГУ 51 экз.

http://kleopatra.pnzgu.ru/cgi-bin/irbis64r_91/cgiirbis_64.exe?P21DBN=KATL&I21DBN=KATL_PRINT&S21FMT=fullw_print&C21COM=F&Z21MFN=7375

4. Свешников А.А. Прикладные методы теории вероятностей. - М: Лань, 2013, 480с. Электронный ресурс

https://e.lanbook.com/book/3184#book_name

5. Свешников А.А. Сборник задач по теории вероятностей, математической статистике и теории случайных функции. М: Лань, 2013. Электронный ресурс

https://e.lanbook.com/book/5711#book_name

б) Интернет-ресурсы

в) Программное обеспечение


1. ПО «Microsoft Windows» (подписка DreamSpark/Microsoft Imagine Standard); регистрационный номер 00037FFEBACF8FD7 Договор № СД-130712001 от 12.07.2013 (подписка с 1 сентября 2013 г. до 31 августа 2017 г.) Продление Microsoft Imagine Standard KDF-00031 (подписка с 1 сентября 2017 г. до 31 августа 2020 г.)

г) Другое материально-техническое обеспечение

1. Персональные компьютеры

Рабочая программа дисциплины «Статистические основы анализа больших данных» составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 01.03.02 — «Прикладная математика и информатика» утвержденного приказом Министерства образования и науки Российской Федерации от « 10» января 2018 № 9.

Программу составили:

1. Бойкова А.И.  _____ доцент кафедры ВиПМ

Настоящая программа не может быть воспроизведена ни в какой форме без предварительного письменного разрешения кафедры-разработчика программы.

Программа одобрена на заседании кафедры «Высшая и прикладная математика»

Протокол № 11 от « 01 » июня 2019 года

Зав. кафедрой ВиПМ  _____ д.ф.м.н., проф. Бойков И.В.

Программа согласована с заведующим выпускающей кафедрой «КТ»

Зав. кафедрой КТ  _____ д.т.н., проф. В.И.Горбаченко

Программа одобрена методической комиссией ФВТ

Протокол № 10 от « 3 » июня 2019 года

Председатель методической комиссии ФВТ

к.т.н., доцент  _____ Глотова Т.В.

**Сведения о переутверждении программы на очередной учебный год и
регистрации изменений**

Учебный год	Решение кафедры (№ протокола, дата)	Внесенные изменения	Подпись зав. кафедрой