

**Министерство науки и высшего образования Российской Федерации**  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
**«Владимирский государственный университет**  
**имени Александра Григорьевича и Николая Григорьевича Столетовых»**  
**(ВлГУ)**

**Институт экономики и туризма**

(Наименование института)

УТВЕРЖДАЮ:

Директор института



Козлов Д.А.

сентября 2023 года

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

**Инструментальные системы работы с данными**

(НАИМЕНОВАНИЕ ДИСЦИПЛИНЫ)

**направление подготовки / специальность**

01.03.05 Статистика

(код и наименование направления подготовки (специальности))

**направленность (профиль) подготовки**

«Бизнес-аналитика»

(направленность (профиль) подготовки)

г. Владимир

2023

## 1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целями освоения дисциплины «Инструментальные системы работы с данными» являются: изучение теоретических основ применения современных интеллектуальных информационно-аналитических систем обработки больших массивов данных в профессиональной деятельности; формирование комплекса практических навыков применения современного инструментария сбора, обработки и анализа больших массивов данных в профессиональной деятельности

Задачи:

- 1) формирование представлений о целях, способах реализации и инструментах сбора, обработки и анализа больших массивов данных;
- 2) изучение сфер применения, методов и средств анализа больших массивов данных;
- 3) формирование практических навыков сбора, обработки и анализа больших массивов данных в целях эффективного управления;
- 4) получение теоретических знаний и практических навыков при решении типовых профессиональных задач с использованием современных информационно-аналитических систем обработки больших массивов данных

## 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина «Инструментальные системы работы с данными» относится к относится к базовой части учебного плана по программе бакалавриата 01.03.05

## 3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения ОПОП (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине, в соответствии с индикатором достижения компетенции		Наименование оценочного средства
	Индикатор достижения компетенции (код, содержание индикатора)	Результаты обучения по дисциплине	
1	2	3	4
ОПК-3. Способен осознанно применять методы математической и дескриптивной статистики для анализа количественных данных, в том числе с применением необходимой вычислительной техники и стандартных компьютерных программ, содержательно интерпретировать полученные результаты, готовить статистические материалы для докладов, публикаций и других аналитических материалов	ОПК-3.1. Знает современный статистический и математический инструментарий для решения профессиональных задач ОПК-3.2. Умеет использовать информационно-коммуникационные технологии и программные средства для анализа количественных данных ОПК-3.3. Владеет навыками интерпретации полученных результатов анализа количественных данных и подготовки материалов для докладов, публикаций и других аналитических материалов	Знать инструментальные системы работы с данными, применяемые для математической и дескриптивной статистики в анализе количественных данных; Уметь использовать инструментальные системы работы с данными, применяемые для математической и дескриптивной статистики в анализе количественных данных; Владеть навыками применения инструментальных систем работы с данными для решения профессиональных задач для математической и дескриптивной статистики в анализе количественных данных	Тестовые и ситуационные задания. Практически задания
ОПК-4. Способен	ОПК-4.1. Знает принципы	Знать:	Тестовые и

понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	работы современных информационных технологий ОПК-4.2. Умеет выбирать информационные технологии ОПК-4.3. Владеет навыками использования современных информационных технологий при решении задач профессиональной деятельности	- основные инструментальные системы работы с данными, используемые в профессиональной деятельности; <i>Уметь:</i> - использовать основные инструментальные системы работы с данными в профессиональной деятельности; - осуществлять выбор инструментальных систем работы с данными для решения профессиональных задач <i>Владеть</i> - навыками применения инструментальных систем работы с данными для решения профессиональных задач	ситуационные задания. Практически задания
--	--	---	---

#### 4. ОБЪЕМ И СТРУКТУРА ДИСЦИПЛИНЫ

##### Тематический план форма обучения – очная

Трудоемкость дисциплины составляет 3 зачетные единицы, 108 часов.

№ п/п	Наименование тем и/или разделов/тем дисциплины	Семестр	Неделя семестра	Контактная работа обучающихся с педагогическим работником				Самостоятельная работа	Формы текущего контроля успеваемости, форма промежуточной аттестации (по семестрам)
				Лекции	Практические занятия	Лабораторные работы	в форме практической подготовки		
1	Понятие инструментальных систем работы с данными и его роль в решении профессиональных задач	3	1-6	6	6			24	Рейтинг-контроль №1
2	Сбор данных, описательный анализ и предварительная обработка данных	3	7-12	6	6			24	Рейтинг-контроль №2
3	Моделирование данных на основе методов машинного обучения	3	13-18	6	6			24	Рейтинг-контроль №3
Всего за 3_семестр:				18	18			72	Зачет
Наличие в дисциплине КП/КР									
Итого по дисциплине				18	18			72	Зачет

#### Содержание лекционных занятий по дисциплине

### **Тема 1. Понятие инструментальных систем работы с данными и его роль в решении профессиональных задач**

Информация и особенности ее хранения и обработки. Принципы и подходы к управлению данными. Содержание и задачи процесса управления данными.

Понятие инструментальных систем работы с данными и его роль в решении профессиональных задач

Сущность интеллектуального анализа данных. Интеллектуальный анализ данных в бизнесе (описание стандартных задач). Технологии интеллектуального анализа. Обзор программных решений для интеллектуального анализа данных. Основные этапы проведения интеллектуального анализа данных и их краткая характеристика. Основные методы анализа данных.

### **Тема 2. Сбор данных, описательный анализ и предварительная обработка данных**

Внешние и внутренние источники данных: системы управления основными и вспомогательными процессами предприятия, специализированные базы данных. Импорт, экспорт, интеграция данных из внешних и внутренних источников. Обзор ключевых операций с наборами однородных и неоднородных данных. Методы описательного анализа данных.

Понятие качества данных. Критерии качества данных: доступность, точность, взаимосвязанность, полнота, непротиворечивость, однозначность, релевантность, надежность и своевременность.

Сущность очистки данных. Методы очистки данных при обработке пропущенных значений, дубликатов, противоречий, аномальных значений и выбросов, шумов, фиктивных значений и ошибок ввода данных.

Сущность оптимизация данных. Методы оптимизации данных: снижение размерности, выявление и исключение незначущих признаков.

### **Тема 3. Моделирование данных на основе методов машинного обучения**

Сущность моделирования данных как этапа проведения интеллектуального анализа данных.

Сущность машинного обучения. Способы машинного обучения: обучение с учителем, обучение без учителя, обучение с подкреплением. Модели машинного обучения для решения прикладных задач: регрессии, классификации, кластеризации, понижения размерности данных, выявления аномалий.

Модели регрессии. Сущность регрессионного анализа. Линейные модели регрессии. Нелинейные модели регрессии

Модели классификации. Классификация как задача машинного обучения. Линейные модели классификации. Модели классификации с нелинейными разделяющими поверхностями.

Модели кластеризации. Кластерный анализ как задача машинного обучения. Модели итеративной кластеризации. Модели иерархической кластеризации. Модели плотностной кластеризации.

Введение в искусственные нейронные сети. Сверточные и рекуррентные нейронные сети. Обучение с подкреплением.

## **Тема 1. Понятие инструментальных систем работы с данными и его роль в решении профессиональных задач**

### **Занятие 1**

Информация и особенности ее хранения и обработки.

Принципы и подходы к управлению данными.

Содержание и задачи процесса управления данными.

Понятие инструментальных систем работы с данными и его роль в решении профессиональных задач

### **Занятие 2**

Области человеческой деятельности наиболее и наименее подходят для их анализа методами интеллектуального анализа

Проблемы организации процессов операционной аналитики

Практические аспекты постановки и решения задачи машинного обучения с учителем

Практические аспекты постановки и решения задачи машинного обучения без учителя

### **Занятие 3**

Практические аспекты постановки и решения задачи машинного обучения с подкреплением

Основные этапы развития теории и инструментария машинного обучения

Правовые проблемы применения инструментов машинного обучения

## **Тема 2. Сбор данных, описательный анализ и предварительная обработка данных**

### **Занятие 4**

Преимущества и недостатки наиболее востребованных инструментов машинного обучения.

Платформы для реализации аналитических процессов с применением программных языков высокого уровня: библиотеки языка R, библиотеки языка Python, библиотеки ML.NET и др.

Платформы для реализации всех аналитических процессов при работе с большими данными: Mahout, MLib, MADLib и др.

Low-code платформы для реализации аналитических процессов: Loginom, Qlik, Microsoft Power BI и др.

### **Занятие 5**

Организация различных способов сбора данных и их реализация в платформе Loginom

Средства визуализации наборов данных в платформе Loginom

Проблемы оценки качества собранных данных

Целесообразность применения и выбор методов очистки и оптимизации данных при решении различных категорий задач машинного обучения

Общие подходы к организации вычислений с использованием платформы Loginom

Способы обработки отсутствующих данных с применением платформы Loginom

### **Занятие 6**

Способы обнаружения и фильтрации выбросов с применением платформы Loginom

Дискретизация непрерывных данных с применением платформы Loginom

Иерархическая индексация с применением платформы Loginom

Работа с временными рядами с применением платформы Loginom

Увеличение производительности платформы Loginom

Разбор веб-страниц с помощью платформы Loginom

### **Тема 3. Моделирование данных на основе методов машинного обучения и искусственных нейронных сетей**

#### **Занятие 7**

Проблемы практического применения регрессионных моделей

Методы отбора признаков в задачах линейной регрессии

Значимость коэффициентов линейной регрессии

Методы борьбы с переобучением в различных методах линейной регрессии

Вероятностная постановка задач регрессии

Взаимосвязь между сложностью регрессионной модели и размером набора данных

Вычисление оптимальных параметров логистической регрессии

Проблема определения оптимальных параметров линейной регрессии

Метрики качества регрессии

Регрессионный анализ с помощью платформы Logitom

Проблемы практического применения классификационных моделей

#### **Занятие 8**

Сравнение методов классификации

Метрики качества разбиения в задачах классификации и регрессии.

Композиции алгоритмов, применяемые в задачах классификации

Вычисление оптимальных параметров моделей классификации

Методы борьбы с переобучением в различных методах классификации

Классификация с помощью платформы Logitom

Проблема определения оптимальных параметров классификации

Метрики качества классификации: доля правильных ответов, матрица ошибок, точность и полнота, AUC, индекс Джини, чувствительность к размеру классов.

Чувствительность и специфичность метода классификации

#### **Занятие 9**

Проблемы практического применения моделей кластеризации

Сравнение методов кластеризации. Подходы к классификации методов кластеризации

Методы итеративной кластеризации: преимущества и недостатки, область применения

Методы иерархической кластеризации: преимущества и недостатки, область применения

Методы плотностной кластеризации: преимущества и недостатки, область применения

Методы борьбы с переобучением в различных методах кластеризации

Проблема определения оптимальных параметров кластеризации

Метрики качества классификации

Кластеризация с помощью платформы Logitom

## **5. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ**

**5.1. Текущий контроль успеваемости** (*рейтинг-контроль 1, рейтинг-контроль 2, рейтинг-контроль 3*).

### **Рейтинг-контроль №1**

Задание №1. Выполните следующие задания:

1. Сформулируйте задачи, возникающие в заданном по вариантам виде деятельности, которые можно было бы решить с использованием машинного обучения

(необходимо выделить как минимум по одной задачи регрессии, классификации и кластеризации).

2. Опишите каждую задачу по следующей схеме: сущность задачи, класс задачи, состав признаков (наименование, тип данных, ограничения на значения), состав меток (наименование, тип данных, ограничения на значения).

Варианты видов деятельности: 1) банковская деятельность; 2) электронная коммерция; 3) риэлтерская деятельность; 4) информационная безопасность; 5) розничная торговля; 6) сельское хозяйство; 7) транспортные услуги; 8) туристические услуги; 9) услуги страхования.

### **Рейтинг-контроль №2**

Задание №1. Выполните следующие задания:

1. На основании вариантов видов деятельности, найдите во внешних открытых репозиториях несколько наборов данных им соответствующих и сделайте их описание по следующей схеме:

- а) наименование репозитория с указанием его интернет-адреса,
- б) краткое и полное наименование набора данных,
- в) идентификатор набора данных,
- г) краткая характеристика (решаемые задачи, первоисточник данных и т.д.),
- д) количественные параметры набора данных (размер выборки),
- е) описание признаков и меток.

2. Проведите анализ данных в наборе на доступность, точность, взаимосвязанность, полноту, непротиворечивость, однозначность, релевантность, надежность и своевременность.

Варианты видов деятельности: 1) банковская деятельность; 2) электронная коммерция; 3) риэлтерская деятельность; 4) информационная безопасность; 5) розничная торговля; 6) сельское хозяйство; 7) транспортные услуги; 8) туристические услуги; 9) услуги страхования.

### **Рейтинг-контроль №3**

Задание №1. Выполните следующие задания:

1. На основании набора данных о покупках в розничном магазине, занимающемся продажей подарков и сувениров через интернет-сайт, осуществите кластеризацию заказов и покупателей. Источник данных: <https://archive.ics.uci.edu/ml/datasets/online+retail>

2. На основании данных о транзакциях по кредитным картам необходимо выявить те из них, которые являются мошенническими. Источник данных: <https://www.kaggle.com/mlg-ulb/creditcardfraud>

3. На основании данных о производстве конфет в США с января 1972 по настоящий момент необходимо спрогнозировать индустриальный продуктовый индекс (IPG3113N – универсальный индекс уровня производства, который измеряется как % от уровня производства 2012 года). Горизонт прогнозирования: временной интервал в 24 месяца. Источник данных: <https://fred.stlouisfed.org/series/IPG3113N>

Задание №2. Подготовьте Развернутый ответ на следующие вопросы:

1. Обзор задач, решаемых алгоритмами машинного обучения.
2. Классификация алгоритмов машинного обучения
3. Алгоритм линейной регрессии
4. Алгоритм логистической регрессии.
5. Алгоритм дерева решений.
6. Алгоритм AdaBoost.
7. Алгоритм k-means.

**1.2. Промежуточная аттестация** по итогам освоения дисциплины производится в виде зачета, который включает в себя ответы на теоретические вопросы.

*Вопросы к зачету:*

1. Сущность интеллектуального анализа данных.
2. Интеллектуальный анализ данных в бизнесе (описание стандартных задач).
3. Технологии интеллектуального анализа.
4. Обзор программных решений для интеллектуального анализа данных.
5. Основные этапы проведения интеллектуального анализа данных и их краткая характеристика.
6. Основные методы анализа данных.
7. Внешние и внутренние источники данных: системы управления основными и вспомогательными процессами предприятия, специализированные базы данных.
8. Импорт, экспорт, интеграция данных из внешних и внутренних источников.
9. Обзор ключевых операций с наборами однородных и неоднородных данных.
10. Методы описательного анализа данных.
11. Понятие качества данных.
12. Критерии качества данных: доступность, точность, взаимосвязанность, полнота, непротиворечивость, однозначность, релевантность, надежность и своевременность.
13. Сущность очистки данных.
14. Методы очистки данных при обработке пропущенных значений, дубликатов, противоречий, аномальных значений и выбросов, шумов, фиктивных значений и ошибок ввода данных.
15. Сущность оптимизация данных.
16. Методы оптимизации данных: снижение размерности, выявление и исключение незначимых признаков.
17. Сущность моделирования данных как этапа проведения интеллектуального анализа данных.
18. Сущность машинного обучения.
19. Способы машинного обучения: обучение с учителем, обучение без учителя, обучение с подкреплением.
20. Модели машинного обучения для решения прикладных задач: регрессии, классификации, кластеризации, понижения размерности данных, выявления аномалий.
21. Модели регрессии.
22. Сущность регрессионного анализа.
23. Линейные модели регрессии.
24. Нелинейные модели регрессии
25. Модели классификации.
26. Классификация как задача машинного обучения.
27. Линейные модели классификации.
28. Модели классификации с нелинейными разделяющими поверхностями.
29. Модели кластеризации.
30. Кластерный анализ как задача машинного обучения.
31. Сущность понятия «Big Data».
32. Принципы и подходы к управлению Big Data.
33. Содержание и задачи процесса управления большими данными.
34. Проблемы использования Big Data.
35. Системы хранения больших данных.

**5.3. Самостоятельная работа обучающегося** производится в виде докладов с презентацией.



*Подготовка докладов по следующим темам:*

1. Критерии качества данных: доступность, точность, взаимосвязанность, полнота, непротиворечивость, однозначность, релевантность, надежность и своевременность.
2. Методы очистки данных при обработке пропущенных значений, дубликатов, противоречий, аномальных значений и выбросов, шумов, фиктивных значений и ошибок ввода данных.
3. Методы оптимизации данных: снижение размерности, выявление и исключение незначимых признаков.
4. Линейные модели регрессии.
5. Нелинейные модели регрессии
6. Линейные модели классификации.
7. Модели классификации с нелинейными разделяющими поверхностями.
8. Модели итеративной кластеризации.
9. Модели иерархической кластеризации.
10. Модели плотностной кластеризации.

Фонд оценочных материалов (ФОМ) для проведения аттестации уровня сформированности компетенций обучающихся по дисциплине оформляется отдельным документом

## 6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### 6.1. Книгообеспеченность

Наименование литературы: автор, название, вид издания, издательство	Год издания	КНИГООБЕСПЕЧЕННОСТЬ
		Наличие в электронном каталоге ЭБС
Основная литература		
Горбаченко, В. И. Машинное обучение: учебное пособие / В. И. Горбаченко, К. Е. Савенков, М. А. Малахов. — Москва : Ай Пи Ар Медиа, 2023. — 217 с. — ISBN 978-5-4497-1860-0. — Текст : электронный // Цифровой образовательный ресурс IPR SMART. Режим доступа: для авторизир. пользователей	2023	<a href="https://www.iprbookshop.ru/125886.html">https://www.iprbookshop.ru/125886.html</a>
Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — Москва : Издательство Юрайт, 2023. — 85 с. — (Высшее образование). — ISBN 978-5-534-15561-7. — Текст : электронный // Образовательная платформа Юрайт [сайт].	2023	<a href="https://urait.ru/bcode/520544">https://urait.ru/bcode/520544</a>
Бессмертный, И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — Москва : Издательство Юрайт, 2023. — 243 с. — (Высшее образование). — ISBN 978-5-534-01042-8. — Текст : электронный // Образовательная платформа Юрайт [сайт].	2023	<a href="https://urait.ru/bcode/511999">https://urait.ru/bcode/511999</a>
Дополнительная литература		
Воронова Л.И. Machine Learning: регрессионные методы интеллектуального анализа данных : учебное пособие / Воронова Л.И., Воронов В.И.. — Москва :	2018	<a href="http://znanium.com/bookread2.php?book=517558">http://znanium.com/bookread2.php?book=517558</a>

Московский технический университет связи и информатики — 82 с. — ISBN 2227-8397.		
Афанасьева Т.В. Введение в проектирование систем интеллектуального анализа данных : учебное пособие / Афанасьева Т.В., Афанасьев А.Н.. — Ульяновск : Ульяновский государственный технический университет. — 64 с. — ISBN 978-5-9795-1686-8.	2017	<a href="http://www.iprbookshop.ru/106086.html">http://www.iprbookshop.ru/106086.html</a>
Билл Фрэнкс Революция в аналитике: Как в эпоху Big Data улучшить ваш бизнес с помощью операционной аналитики / Билл Фрэнкс. — Москва : Альпина Паблишер — 320 с. — ISBN 978-5-9614-5302-7.	2020	<a href="http://www.iprbookshop.ru/93032.html">http://www.iprbookshop.ru/93032.html</a>

## 6.2. Периодические издания

1. Журнал «Инновации».
2. Журнал «Наука, инновации, технологии».
3. Журнал «Информационное общество».
4. Издание о высоких технологиях.
5. The Journal of Machine Learning Research (JMLR).
6. Machine Learning.
7. Data Mining And Knowledge Discovery.
8. Intelligent Data Analysis.
9. Pattern Recognition And Image Analysis: Advances In Mathematical Theory And Applications.

## 6.3. Интернет-ресурсы

1. <https://loginom.ru/>
2. <http://www.government.gov.ru>
3. <http://www.innovbusiness.ru>
4. <http://www.rsci.ru>
5. <http://www.technopark.al.ru>
6. <http://economics.edu.ru>
7. <http://e.lib.vlsu.ru>

## 7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Учебная аудитория, компьютерный класс с выходом в Internet для проведения занятий лекционного и семинарского типа, текущего и промежуточного контроля, групповых и индивидуальных консультаций.

Демонстрационное оборудование: 10 компьютеров kraftwey с мышками и клавиатурой, маркерная доска, диапроектор Epson, маркерная настенная доска (4x1,2 м), кондиционер Hitachi.

Количество посадочных мест: 22.

Расположена по адресу: 600005, Российская Федерация, Владимирская область, г.о. город Владимир, г. Владимир, ул. Горького, д. 79, 3 этаж учебного корпуса № 6, 30,0 м<sup>2</sup>, № 24.

Перечень используемого лицензионного программного обеспечения: операционная

система семейства Microsoft Windows, пакет офисных программ Microsoft Office, справочная правовая система «Консультант Плюс» (инсталлированный ресурс ВлГУ).

Примечание

В соответствии с нормативно-правовыми актами для инвалидов и лиц с ограниченными возможностями здоровья при необходимости тестирование может быть проведено только в письменной или устной форме, а также могут быть использованы другие материалы контроля качества знаний, предусмотренные рабочей программой дисциплины.

Рабочую программу составил: ст. преподаватель Виноградов Д.В.

Рецензент (представитель работодателя):

Начальник отдела информатизации арбитражного суда Владимирской области Дигилин Д.Е.

Программа рассмотрена и одобрена на заседании кафедры БИиЭ  
протокол № 1 от «30» августа 2023 года.

Заведующий кафедрой: д.э.н., профессор Тесленко И.Б.

Рабочая программа рассмотрена и одобрена на заседании учебно-методической комиссии  
направления 01.03.05 Статистика  
протокол № 1 от «05» сентября 2023 года.

Председатель комиссии: к.э.н., доцент Ярьес О.Б.