

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
Факультет - экономический

УТВЕРЖДАЮ:  
Проректор по учебной работе,  
качеству образования – первый  
проректор  
Г.А. Хагуров  
(подпись)  
« 31 » мая 2024 г.



## РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

### Б1.О.07 ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ БИЗНЕС-АНАЛИТИКИ

Направление подготовки/специальность 38.04.05 Бизнес-информатика

Направленность (профиль) / специализация Инновации и бизнес в сфере  
информационных технологий

Форма обучения очная/заочная

Квалификация магистр

Краснодар 2024

Рабочая программа дисциплины «Информационные технологии бизнес-аналитики» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки 38.04.05 Бизнес-информатика

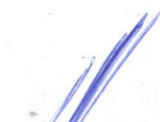
Программу составил(и):  
И.В. Ариничев, к.э.н., доцент  
И.О. Фамилия, должность, ученая степень, ученое звание

  
\_\_\_\_\_ подпись

Рабочая программа дисциплины утверждена на заседании кафедры теоретической экономики, протокол № 9 от «2» апреля 2024 г.

Заведующий кафедрой теоретической экономики

Сидоров В.А.  
фамилия, инициалы

  
\_\_\_\_\_ подпись

Утверждена на заседании учебно-методической комиссии экономического факультета, протокол № 9 от «14» мая 2024 г.

Председатель УМК факультета

Дробышевская Л.Н.  
фамилия, инициалы

  
\_\_\_\_\_ подпись

Рецензенты:

Соболев Э.В., директор Краснодарского филиала Финансового университета при Правительстве Российской Федерации, к.э.н., доцент.

Шевченко И.В., декан экономического факультета КубГУ, заведующий кафедрой мировой экономики и менеджмента, д.э.н., профессор.

## 1 Цели и задачи изучения дисциплины

### 1.1 Цель дисциплины

представить фундаментальную теоретическую основу для управления инновационными процессами и формировать научные основы принятия решения по управлению инновационной деятельностью. В процессе изучения дисциплины формируются знания в области законов, закономерностей и принципов инновационной деятельности, модели и методы управления инновационными процессами.

### 1.2 Задачи дисциплины

- изучение сущности и роли бизнес-аналитики в современных условиях растущей конкуренции и быстро изменяющегося рынка;
- ознакомление обучающихся с основными системами бизнес-аналитики, функциями и методами бизнес-аналитики;
- знакомство с основными технологиями бизнес-аналитики, такими как OLAP-технологии, DM-технологии, системы визуализации данных и решений и др.;
- определение основных проблем и перспектив развития систем бизнес-аналитики, оценка их эффективности;
- формирование прикладных умений и навыков в области поддержки принятия управленческих решений в организации с применением современных методов и средств.

### 1.3 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Информационные технологии бизнес-аналитики» относится к дисциплинам обязательной части цикла дисциплин учебного плана и имеет шифр Б1.О.07

Дисциплина базируется на знаниях, полученных по стандарту высшего образования, и является основой для решения исследовательских задач. Предусматривает использование знаний, полученных в ходе изучения следующих дисциплин: Архитектура предприятия, Системный анализ и методы принятия управленческих решений.

Знания, полученные в ходе изучения дисциплины, используются для изучения курсов «Диагностика бизнес-деятельности», «Информационный менеджмент», «Цифровой маркетинг» и в процессе выполнения магистерской диссертации.

### 1.4 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Процесс изучения дисциплины направлен на формирование следующих компетенций:

Код и наименование индикатора* достижения компетенции	Результаты обучения по дисциплине
<b>ОПК-3 Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта;</b>	
ИОПК-3.1. Применяет современные инструментальные методы и программный инструментарий сбора, обработки и анализа данных с использованием интеллектуального оборудования и систем искусственного интеллекта.	<i>Знает:</i> текущее состояние бизнес-процессов на предприятии и действующие ИС и ИКТ <i>Знает:</i> виды аналитики, современные методы и инструментальные средства анализа больших данных; <i>Знает:</i> возможности использования свободно распространяемого программного обеспечения бизнес-аналитики;

Код и наименование индикатора* достижения компетенции	Результаты обучения по дисциплине
	<p><i>Умеет:</i> проводить научные исследования для выработки стратегических решений в области ИКТ</p> <p><i>Умеет:</i> визуализировать результаты проведенного анализа;</p> <p><i>Трудовое действие:</i> выбор методов и инструментальных средств бизнес-аналитики;</p> <p><i>Трудовое действие:</i> Реализация интеллектуальных алгоритмов на структурированных и неструктурированных данных</p>

## 2. Структура и содержание дисциплины

### 2.1 Распределение трудоёмкости дисциплины по видам работ Общая

трудоёмкость дисциплины составляет 3 зач.ед. (108 часов), их распределение по видам работ представлено в таблице.

Виды работ	Всего часов	ОФО	ЗФО
	<b>ОФО/ЗФО</b>	2 семестр	1 курс
<b>Контактная работа, в том числе:</b>	<b>24,2/12,2</b>	<b>24,2</b>	<b>12,2</b>
<b>Аудиторные занятия (всего):</b>	<b>24/12</b>	<b>24</b>	<b>12</b>
занятия лекционного типа	6/4	6	4
лабораторные занятия	18/8	18	8
практические занятия			
семинарские занятия			
<b>Иная контактная работа:</b>	<b>0,2/0,2</b>	<b>0,2</b>	<b>0,2</b>
Контроль самостоятельной работы (КСР)			
Промежуточная аттестация (ИКР)	0,2/0,2	0,2	0,2
<b>Самостоятельная работа, в том числе:</b>	<b>83,8/92</b>	<b>83,8</b>	<b>92</b>
Самостоятельное изучение разделов, самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к лабораторным и практическим занятиям, коллоквиумам и т.д.)	83,8/92	83,8	92
<b>Контроль:</b>			<b>3,8</b>
<b>Общая трудоёмкость</b>	<b>час</b>	<b>108/108</b>	<b>108</b>
	<b>В т.ч. контактная</b>	<b>24,2/24,2</b>	<b>12,2</b>
	<b>Зач. ед.</b>	<b>4/4</b>	<b>4</b>

### 2.2 Структура дисциплины:

Распределение видов учебной работы и их трудоемкости по разделам дисциплины (ОФО).

№ раздела	Наименование разделов	Количество часов				
		Всего	Аудиторная работа			Самостоятельная работа
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1.	Системы поддержки управленческих решений (DSS/BI). Системы бизнес-аналитики (BA)	19	1		2	16
2.	Система Project Expert для бизнес-аналитики	21	1		4	16
3.	Управление эффективностью бизнеса (BPM): компоненты, стандарты	21	1		4	16
4.	Анализ данных и знаний	21	1		4	16
5.	Технологии интеллектуального анализа данных (Data Mining, DM)	25,8	2		4	19,8
6.	ИКР	0,2				
	<b>Итого по дисциплине:</b>	<b>108</b>	<b>6</b>		<b>18</b>	<b>83,8</b>

Распределение видов учебной работы и их трудоемкости по разделам дисциплины (ЗФО).

№ раздела	Наименование разделов	Количество часов				
		Всего	Аудиторная работа			Самостоятельная работа
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1.	Системы поддержки управленческих решений (DSS/BI). Системы бизнес-аналитики (BA)	23	1		2	20
2.	Система Project Expert для бизнес-аналитики	15	1		2	12
3.	Управление эффективностью бизнеса (BPM): компоненты, стандарты	23	1		2	20
4.	Анализ данных и знаний	21	1			20
5.	Технологии интеллектуального анализа данных (Data Mining, DM)	22			2	20
6.	ИКР	0,2				
7.	Контроль	3,8				
	<b>Итого по дисциплине:</b>	<b>108</b>	<b>4</b>		<b>8</b>	<b>92</b>

## 2.3 Содержание разделов дисциплины:

### 2.3.1 Занятия лекционного типа

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1.	Системы поддержки управленческих решений (DSS/BI). Системы бизнес-аналитики (BA)	Сущность бизнес-аналитики, ее роль на современном предприятии. Системы поддержки управленческих решений (DSS/BI). Появление термина «Business intelligence» (BI). Системы бизнес-аналитики (BA). BI как методы, технологии, средства извлечения и представления знаний; BI как знания о бизнесе и для бизнеса. Преимущества внедрения системы бизнес-аналитики на современном предприятии. Задачи, решаемые с помощью бизнес-аналитики. Классификация продуктов Business Intelligence. Функции бизнес-аналитики: идентификация, моделирование, прогнозирование, оптимизация решений, анализ чувствительности.	T, O
2.	Система Project Expert для бизнес-аналитики	Разработка бизнес-планов, оценка и реализация инвестиционных проектов. Построение финансовой модели проекта, компании, с учетом	T, O
3.	Управление эффективностью бизнеса (BPM): компоненты, стандарты	Введение в визуализацию. Визуализаторы общего назначения. Таблицы. Графики. Диаграммы. Круговые диаграммы. Гистограммы. Сложные визуализаторы общего назначения. OLAP-анализ. Географические карты. Тепловые карты. Плоское дерево. Диаграмма связей. Облако тегов. Пузырьковая диаграмма. Диаграмма рассеяния.	T, O
4.	Анализ данных и знаний	Введение. Постановки задач в машинном обучении. Обучение с учителем и без. Классификация, регрессия, ранжирование, кластеризация. Обучающая и тестовая выборки. Проблема переобучения. Кросс-валидация.	T, O
5.	Технологии интеллектуального анализа данных (Data Mining, DM)	Мультикласс и мультитейл классификация. Обучение классификаторов. Метод ближайших соседей. Решающие деревья.	T, O

### 2.3.2 Лабораторные работы

№	Наименование раздела	Тематика лабораторных работ	Форма текущего контроля
1	2	3	4
1.	Системы поддержки управленческих решений (DSS/BI). Системы бизнес-	Разработка и контроль календарного плана с использованием программного обеспечения MS Project и Project Expert. Диаграмма Ганта	Расчетно-графическое задание

	аналитики (BA)		
2.	Система Project Expert для бизнес-аналитики	Оценка финансовых результатов и эффективности инвестиций с использованием программного обеспечения MS Excel и Project Expert	Расчетно-графическое задание
3.	Управление эффективностью бизнеса (BPM): компоненты, стандарты	Разработка ресурсного плана и его оптимизация с использованием программного обеспечения MS Project	Расчетно-графическое задание
4.	Анализ данных и знаний	Регрессионный анализ в <i>Automated Neural Networks</i> пакета Statistica	Расчетно-графическое задание
5.	Технологии интеллектуального анализа данных (Data Mining, DM)	Задача классификации в <i>Automated Neural Networks</i> пакета Statistica. Нейросетевое прогнозирование временных рядов: с помощью нейромитатора Statistica Neural Networks или Neural Network Wizard или Matlab Neural Network Toolbox построить прогноз показателя по некоторому количеству его прошлых и текущих значений. Разработка интеллектуальных решений в среде Python	Расчетно-графическое задание



## 2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	Занятия лекционного и семинарского типа	Методические указания для подготовки к занятиям лекционного и семинарского типа. Утверждены на заседании Совета экономического факультета ФГБОУ ВО «КубГУ». Протокол № 1 от 30 августа 2018 года.. Режим доступа: <a href="https://www.kubsu.ru/ru/econ/metodicheskie-ukazaniya">https://www.kubsu.ru/ru/econ/metodicheskie-ukazaniya</a>
2	Выполнение самостоятельной работы обучающихся	Методические указания по выполнению самостоятельной работы обучающихся. Утверждены на заседании Совета экономического факультета ФГБОУ ВО «КубГУ». Протокол № 1 от 30 августа 2018 года.. Режим доступа: <a href="https://www.kubsu.ru/ru/econ/metodicheskie-ukazaniya">https://www.kubsu.ru/ru/econ/metodicheskie-ukazaniya</a>
3	Выполнение лабораторных работ	Методические указания по выполнению лабораторных работ. Утверждены на заседании Совета экономического факультета ФГБОУ ВО «КубГУ». Протокол № 1 от 30 августа 2018 года.. Режим доступа: <a href="https://www.kubsu.ru/ru/econ/metodicheskie-ukazaniya">https://www.kubsu.ru/ru/econ/metodicheskie-ukazaniya</a>
4	Интерактивные методы обучения	Методические указания по интерактивным методам обучения. Утверждены на заседании Совета экономического факультета ФГБОУ ВО «КубГУ». Протокол № 1 от 30 августа 2018 года. Режим доступа: <a href="https://www.kubsu.ru/ru/econ/metodicheskie-ukazaniya">https://www.kubsu.ru/ru/econ/metodicheskie-ukazaniya</a>

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,
- в форме аудиофайла,
- в печатной форме на языке Брайля.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,
- в форме аудиофайла.

Для освоения дисциплины инвалидами и лицами с ограниченными возможностями здоровья имеются издания в электронном виде <https://www.kubsu.ru/ru/node/14538> в электронно-библиотечных системах:

ЭБС «Университетская библиотека онлайн» <https://biblioclub.ru/> ЭБС  
издательства «Лань» <https://e.lanbook.com> ЭБС «Юрайт» <https://biblio-online.ru>  
ЭБС «ZNANIUM.COM» <http://znanium.com> ЭБС  
«Book.ru» <https://www.book.ru>

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

### 3. Образовательные технологии

В соответствии с требованиями ФГОС ВО по направлению подготовки 38.04.05 «Бизнес-информатика» реализация компетентностного подхода должна предусматривать



широкое использование в учебном процессе активных и интерактивных форм проведения занятий в сочетании с внеаудиторной работой с целью формирования и развития требуемых компетенций обучающихся:

1. Лабораторная работа в компьютерном классе, компьютерная технология обучения.
2. Метод проектов.
3. Поисковый, эвристический метод.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

#### **4. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации**

##### **4.1 Фонд оценочных средств для проведения текущей аттестации**

В ходе текущей аттестации оцениваются промежуточные результаты освоения студентами дисциплины «ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ БИЗНЕС-АНАЛИТИКИ». Текущий контроль осуществляется с использованием традиционной технологий оценивания качества знаний студентов и включает оценку самостоятельной (внеаудиторной) и аудиторной работы (в том числе рубежный контроль). В качестве оценочных средств используются:

- различные виды устного и письменного контроля (выступление на семинаре, реферат);
- отчет по лабораторной работе.

##### **Тематика рефератов**

1. Проектирование корпоративной информационно-аналитической системы (Business Intelligence) (на примере конкретного предприятия).
2. Подходы к созданию информационно-аналитических систем (на примере конкретного предприятия).
3. Проектирование системы сбалансированных показателей (BSC) (на примере конкретного предприятия).
4. Управление эффективностью бизнеса (на примере конкретного предприятия).
5. Многомерное проектирование данных в Business Intelligence – приложении (для конкретного предприятия).
6. Выбор и проектирование архитектуры OLAP-приложения (на примере конкретного предприятия).
7. Аналитические функции для прогнозирования показателей в BI (на примере конкретного предприятия).
8. Обзор продуктов Business Intelligence (на примере конкретного предприятия).
9. Формирование информационно-аналитической системы (Business Intelligence) (на примере конкретного предприятия).
10. Адаптация КИС к потребностям конкретного предприятия
11. Проектирование системы визуализации данных и решений (на примере конкретного предприятия).
12. Проектирование инфраструктуры BRP (на примере конкретного предприятия)
13. Проектирование прикладных сервисов BRP (на примере конкретного предприятия).
14. Проектирование инфраструктуры и архитектуры TQM (системы всеобщего управления качеством) (на примере конкретного предприятия).
15. Проектирование инфраструктуры и архитектуры TQM (системы всеобщего управления качеством) (на примере конкретного предприятия).
16. Проектирование инфраструктуры и архитектуры ERP (комплексной системы планирования и управления ресурсами организации) на примере конкретного предприятия.
17. Проектирование прикладных сервисов BRP (на примере конкретного предприятия).
18. Инжиниринг и аналитика бизнес-архитектуры на примере конкретного предприятия.
19. Business Intelligence на этапе бизнес-планирования.

20. Методики обнаружения нового знания в хранилищах данных (KDD) (пример реализации).

#### **Критерии оценки реферата**

Оценка «отлично» ставится, если выполнены все требования к написанию и защите реферата: обозначена проблема и обоснована её актуальность, сделан краткий анализ различных точек зрения на рассматриваемую проблему и логично изложена собственная позиция, сформулированы выводы, тема раскрыта полностью, соблюдены требования к внешнему оформлению, даны правильные ответы на дополнительные вопросы. Оценка «хорошо» – основные требования к реферату и его защите выполнены, но при этом допущены недочёты. В частности, имеются неточности в изложении материала; отсутствует логическая последовательность в суждениях; имеются упущения в оформлении; на дополнительные вопросы при защите даны неполные ответы. Оценка «удовлетворительно» – имеются существенные отступления от требований к реферированию. В частности, тема освещена лишь частично; допущены фактические ошибки в содержании реферата или при ответе на дополнительные вопросы; во время защиты отсутствует вывод. Оценка «неудовлетворительно» – тема реферата не раскрыта, обнаруживается существенное непонимание проблемы.

**Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы**

#### *Тест на тему «Анализ данных и знаний»*

Для работы вам понадобятся преобработанные данные на kaggle: «Прогноз популярности статьи на Хабре». Скачайте данные <https://drive.google.com/file/d/1nV2qV9otN3LnVSDqy95hvpJdb6aWtATk/view>, загрузите первые 1500 наблюдений в Pandas и ответьте на следующие вопросы:

1. В каком месяце (и какого года) было больше всего публикаций?
  2. март 2016
  3. март 2015
  4. апрель 2015
  5. апрель 2016
2. Проанализируйте публикации в месяце из предыдущего вопроса. Выберите один или несколько вариантов:
  1. Один или несколько дней сильно выделяются из общей картины
  2. На хабре всегда больше статей, чем на гиктаймсе
  3. По субботам на гиктаймсе и на хабрахабр публикуют примерно одинаковое число статей
  4. Подсказки: постройте график зависимости числа публикаций от дня; используйте параметр hue; не заморачивайтесь сильно с ответами и не ищите скрытого смысла :)
3. Когда лучше всего публиковать статью?
  1. Больше всего просмотров набирают статьи, опубликованные в 12 часов дня
  2. У опубликованных в 10 утра постов больше всего комментариев
  3. Больше всего просмотров набирают статьи, опубликованные в 6 часов утра
  4. Максимальное число комментариев на гиктаймсе набрала статья, опубликованная в 9 часов вечера
  5. На хабре дневные статьи комментируют чаще, чем вечерние
4. Кого из топ-20 авторов чаще всего минусуют?

1. @Mordatyj
2. @Mithgol
3. @alizar
4. @ilya42

5. Сравните субботы и понедельники¶

Правда ли, что по субботам авторы пишут в основном днём, а по понедельникам — в основном вечером?

1. Правда
2. Неправда

Для продолжения скачайте данные из репозитория UCI [Adult](#) и ответьте на следующие вопросы

6. Сколько мужчин и женщин (признак *sex*) представлено в этом наборе данных?

# Ваш код и ответ здесь

7. Каков средний возраст (признак *age*) женщин?

# Ваш код и ответ здесь

8. Какова доля граждан Германии (признак *native-country*)?

# Ваш код и ответ здесь

9. Каковы средние значения и среднеквадратичные отклонения возраста тех, кто получает более 50К в год (признак *salary*) и тех, кто получает менее 50К в год?

# Ваш код и ответ здесь

10. Правда ли, что люди, которые получают больше 50к, имеют как минимум высшее образование? (признак *education* – *Bachelors*, *Prof-school*, *Assoc-acdm*, *Assoc-voc*, *Masters* или *Doctorate*)

# Ваш код и ответ здесь

11. Выведите статистику возраста для каждой расы (признак *race*) и каждого пола. Используйте *groupby* и *describe*. Найдите таким образом максимальный возраст мужчин расы *Amer-Indian-Eskimo*.

# Ваш код и ответ здесь

12. Среди кого больше доля зарабатывающих много (>50К): среди женатых или холостых мужчин (признак *marital-status*)? Женатыми считаем тех, у кого *marital-status* начинается с *Married* (*Married-civ-spouse*, *Married-spouse-absent* или *Married-AF-spouse*), остальных считаем холостыми.

# Ваш код и ответ здесь

13. Какое максимальное число часов человек работает в неделю (признак *hours-per-week*)? Сколько людей работают такое количество часов и каков среди них процент зарабатывающих много?

# Ваш код и ответ здесь

14. Посчитайте среднее время работы (*hours-per-week*) зарабатывающих мало и много (*salary*) для каждой страны (*native-country*).

# Ваш код и ответ здесь

**Расчетно-графическое задание на тему : Технологии интеллектуального анализа данных (Data Mining, DM)**

Мы будем работать с набором данных [Samsung Human Activity Recognition](#). Скачайте данные [отсюда](#). Данные поступают с акселерометров и гироскопов мобильных телефонов Samsung Galaxy S3 (подробнее про признаки – по ссылке на UCI выше), также известен вид активности человека с телефоном в кармане – ходил ли он, стоял, лежал, сидел или шел вверх/вниз по лестнице.

Вначале мы представим, что вид активности нам неизвестен, и попробуем кластеризовать людей просто на основе имеющихся признаков. Затем решим задачу определения вида физической активности именно как задачу классификации.

Заполните код в клетках (где написано "Ваш код здесь") и ответьте на вопросы.

Подготовительный этап (загрузка библиотек и стилей)

```
import numpy as np
import pandas as pd
import seaborn as sns
from tqdm import tqdm_notebook

%matplotlib inline
from matplotlib import pyplot as plt

plt.style.use(['seaborn-darkgrid'])
plt.rcParams['figure.figsize'] = (12, 9)
plt.rcParams['font.family'] = 'DejaVu Sans'

from sklearn import metrics
from sklearn.cluster import AgglomerativeClustering, KMeans, SpectralClustering
from sklearn.decomposition import PCA
from sklearn.model_selection import GridSearchCV
from sklearn.preprocessing import StandardScaler
from sklearn.svm import LinearSVC

RANDOM_STATE = 17
X_train = np.loadtxt("../data/samsung_HAR/samsung_train.txt")
y_train = np.loadtxt("../data/samsung_HAR/samsung_train_labels.txt").astype(int)

X_test = np.loadtxt("../data/samsung_HAR/samsung_test.txt")
y_test = np.loadtxt("../data/samsung_HAR/samsung_test_labels.txt").astype(int)
# Проверим размерности
assert(X_train.shape == (7352, 561) and y_train.shape == (7352,))
assert(X_test.shape == (2947, 561) and y_test.shape == (2947,))
```

Для кластеризации нам не нужен вектор ответов, поэтому будем работать с объединением обучающей и тестовой выборок. Объедините  $X_{train}$  с  $X_{test}$ , а  $y_{train}$  – с  $y_{test}$ .

*# Ваш код здесь*

X =

y =

Определим число уникальных значений меток целевого класса.

```
np.unique(y)
array([1, 2, 3, 4, 5, 6])
n_classes = np.unique(y).size
```

Отмасштабируйте выборку с помощью `StandardScaler` с параметрами по умолчанию.

*# Ваш код здесь*

scaler =

X\_scaled =

Понижаем размерность с помощью PCA, оставляя столько компонент, сколько нужно для того, чтобы объяснить как минимум 90% дисперсии исходных (отмасштабированных) данных. Используйте отмасштабированную выборку и зафиксируйте random\_state (константа RANDOM\_STATE).

*# Ваш код здесь*

pca =

X\_pca =

### **Вопрос 1:**

Какое минимальное число главных компонент нужно выделить, чтобы объяснить 90% дисперсии исходных (отмасштабированных) данных?

*# Ваш код здесь*

#### **Варианты:**

- 56
- 65
- 66
- 193

### **Вопрос 2:**

Сколько процентов дисперсии приходится на первую главную компоненту? Округлите до целых процентов.

#### **Варианты:**

- 45
- 51
- 56
- 61

*# Ваш код здесь*

Визуализируйте данные в проекции на первые две главные компоненты.

*# Ваш код здесь*

plt.scatter( , c=y, s=20, cmap='viridis');

### **Вопрос 3:**

Если все получилось правильно, Вы увидите сколько-то кластеров, почти идеально отделенных друг от друга. Какие виды активности входят в эти кластеры?

#### **Ответ:**

- 1 кластер: все 6 активностей
- 2 кластера: (ходьба, подъем вверх по лестнице, спуск по лестнице) и (сидение, стояние, лежание)
- 3 кластера: (ходьба), (подъем вверх по лестнице, спуск по лестнице) и (сидение, стояние, лежание)
- 6 кластеров

Сделайте кластеризацию данных методом KMeans, обучив модель на данных со сниженной за счет PCA размерностью. В данном случае мы подскажем, что нужно искать именно 6 кластеров, но в общем случае мы не будем знать, сколько кластеров надо искать.

Параметры:

- `n_clusters = n_classes` (число уникальных меток целевого класса)
- `n_init = 100`
- `random_state = RANDOM_STATE` (для воспроизводимости результата)

Остальные параметры со значениями по умолчанию.

*# Ваш код здесь*

Визуализируйте данные в проекции на первые две главные компоненты. Раскрасьте точки в соответствии с полученными метками кластеров.

*# Ваш код здесь*

```
plt.scatter(, c=cluster_labels, s=20, cmap='viridis');
```

Посмотрите на соответствие между метками кластеров и исходными метками классов и на то, какие виды активностей алгоритм KMeans путает.

```
tab = pd.crosstab(y, cluster_labels, margins=True)
tab.index = ['ходьба', 'подъем вверх по лестнице',
            'спуск по лестнице', 'сидение', 'стояние', 'лежание', 'все']
tab.columns = ['cluster' + str(i + 1) for i in range(6)] + ['все']
tab
```

Видим, что каждому классу (т.е. каждой активности) соответствуют несколько кластеров. Давайте посмотрим на максимальную долю объектов в классе, отнесенных к какому-то одному кластеру. Это будет простой метрикой, характеризующей, насколько легко класс отделяется от других при кластеризации.

Пример: если для класса "спуск по лестнице", в котором 1406 объектов, распределение кластеров такое:

- кластер 1 – 900
- кластер 3 – 500
- кластер 6 – 6,

то такая доля будет  $900 / 1406 \approx 0.64$ .

#### Вопрос 4:

Какой вид активности отделился от остальных лучше всего в терминах простой метрики, описанной выше?

Ответ:

- ходьба
- стояние
- спуск по лестнице
- перечисленные варианты не подходят

Видно, что kMeans не очень хорошо отличает только активности друг от друга. Используйте метод локтя, чтобы выбрать оптимальное количество кластеров. Параметры алгоритма и данные используем те же, что раньше, меняем только `n_clusters`.

*# Ваш код здесь*

```
inertia = []
for k in tqdm_notebook(range(1, n_classes + 1)):
    #
    #
```

### Вопрос 5:

Какое количество кластеров оптимально выбрать, согласно методу локтя?

### Ответ:

- 1
- 2
- 3
- 4

Попробуем еще один метод кластеризации, который описывался в статье – агломеративную кластеризацию.

```
ag = AgglomerativeClustering(n_clusters=n_classes,
                             linkage='ward').fit(X_pca)
```

Посчитайте Adjusted Rand Index (sklearn.metrics) для получившегося разбиения на кластеры и для KMeans с параметрами из задания к 4 вопросу.

*# Ваш код здесь*

### Вопрос 6:

Отметьте все верные утверждения.

### Варианты:

- Согласно ARI, KMeans справился с кластеризацией хуже, чем Agglomerative Clustering
- Для ARI не имеет значения какие именно метки присвоены кластерам, имеет значение только разбиение объектов на кластеры
- В случае случайного разбиения на кластеры ARI будет близок к нулю

Можно заметить, что задача не очень хорошо решается именно как задача кластеризации, если выделять несколько кластеров (> 2). Давайте теперь решим задачу классификации, вспомнив, что данные у нас размечены.

Для классификации используйте метод опорных векторов – класс sklearn.svm.LinearSVC.

Настройте для LinearSVC гиперпараметр C с помощью GridSearchCV.

- Обучите новый StandardScaler на обучающей выборке (со всеми исходными признаками), примените масштабирование к тестовой выборке
- В GridSearchCV укажите cv=3.

*# Ваш код здесь*

*#*

```
X_train_scaled =
X_test_scaled =
svc = LinearSVC(random_state=RANDOM_STATE)
svc_params = {'C': [0.001, 0.01, 0.1, 1, 10]}
```

*# Ваш код здесь*

*best\_svc =*

*# Ваш код здесь*



### Вопрос 7

Какое значение гиперпараметра C было выбрано лучшим по итогам кросс-валидации?

Ответ:

- 0.001
- 0.01
- 0.1
- 1
- 10

```
y_predicted = best_svc.predict(X_test_scaled)
tab = pd.crosstab(y_test, y_predicted, margins=True)
tab.index = ['ходьба', 'подъем вверх по лестнице', 'спуск по лестнице',
            'сидение', 'стояние', 'лежание', 'все']
tab.columns = tab.index
tab
```

### Вопрос 8:

Какой вид активности SVM определяет хуже всего в терминах точности? Полноты?

Ответ:

- по точности – подъем вверх по лестнице, по полноте – лежание
- по точности – лежание, по полноте – сидение
- по точности – ходьба, по полноте – ходьба
- по точности – стояние, по полноте – сидение

Наконец, проделайте то же самое, что в 7 вопросе, только добавив PCA.

- Используйте выборки `X_train_scaled` и `X_test_scaled`
- Обучите тот же PCA, что раньше, на отмасштабированной обучающей выборке, примените преобразование к тестовой
- Настройте гиперпараметр C на кросс-валидации по обучающей выборке с PCA-преобразованием. Вы заметите, насколько это проходит быстрее, чем раньше.

### Вопрос 9:

Какова разность между лучшим качеством (долей верных ответов) на кросс-валидации в случае всех 561 исходных признаков и во втором случае, когда применялся метод главных компонент? Округлите до целых процентов.

Варианты:

- Качество одинаковое
- 2%
- 4%
- 10%
- 20%

### Вопрос 10:

Выберите все верные утверждения:

Варианты:

- Метод главных компонент в данном случае позволил уменьшить время обучения модели, при этом качество (доля верных ответов на кросс-валидации) очень пострадало, более чем на 10%
- PCA можно использовать для визуализации данных, однако для этой задачи есть и лучше подходящие методы, например, tSNE. Зато PCA имеет меньшую вычислительную сложность
- PCA строит линейные комбинации исходных признаков, и в некоторых задачах они могут плохо интерпретироваться человеком

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине (модулю) предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

#### **4.2 Фонд оценочных средств для проведения промежуточной аттестации**

##### **Вопросы к зачету**

1. Сущность бизнес-анализа.
2. Особенности профессии бизнес-аналитиков.
3. Возможные подходы к организации обследования предприятия.
4. Методы и организация сбора данных о бизнес-процессах предприятия.
5. Описание этапов обследования предприятия.
6. Сущность системного подхода к управлению.
7. Основные понятия системного подхода к управлению.
8. Понятие и особенности системного анализа.
9. Основные принципы системного анализа.
10. Теоретическая основа системного анализа.
11. Этапы процесса системного анализа и их характеристика.
12. Сущность структурного анализа организации.
13. Этапы процесса структурного анализа организации.
14. Содержание, цели и объекты стратегического анализа.
15. Этапы стратегического процесса.
16. Отличия стратегического анализа от традиционного анализа.
17. Основные принципы стратегического системного анализа.

18. Основные методологические подходы к стратегическому анализу и конкретные объекты такого анализа.
19. Характеристика внешней среды компании. Методы стратегического анализа внешней среды.
20. Характеристика внутренней среды организации предприятия. Методы стратегического анализа внутренней среды.
21. Понятие, критерии и методы анализа сегментации.
22. Сущность PEST-анализа.
23. Цель, объект и этапы отраслевого анализа.
24. Основные факторы для оценки общей ситуации в отрасли.
25. Сущность и значение модели конкурентных сил М.Портера.
26. Элементов декомпозиции конкурентного давления со стороны существующих конкурентов.
27. Элементы декомпозиции для оценки давления на компанию со стороны поставщиков.
28. Элементы декомпозиции для оценки угрозы появления на рынке товаров-заменителей (субститутов).
29. Понятие «Business Intelligence». Основной функциональный набор и программные средства BI-систем.
30. Возможности специализированных информационно-аналитических систем по поддержке решения задач бизнес-аналитики.
31. Определение понятия Data Mining.
32. Определение понятия KDD.
33. Программные средства анализа больших данных.
34. Первичный анализ данных с помощью инструментов визуализации. Очистка данных.
35. Основные понятия машинного обучения. Основные постановки задач. Примеры прикладных задач.
36. Метрические методы классификации. Метрики качества алгоритмов классификации и регрессии.
37. Линейные методы классификации и регрессии: функционалы качества, методы настройки, особенности применения.
38. Оценивание качества алгоритмов. Отложенная выборка, ее недостатки. Оценка полного скользящего контроля. Кросс-валидация. Leave-one-out.
39. Деревья решений. Методы построения деревьев.
40. Случайный лес, его особенности.
41. Градиентный бустинг, его особенности при использовании деревьев в качестве базовых алгоритмов.
42. Нейронные сети. Метод обратного распространения ошибок.
43. Обучение без учителя. Методы понижения размерности. PCA.
44. Обучение без учителя. Кластеризация.
45. Вопросы безопасности Big Data

### **Критерии оценки ответов студентов на зачёте.**

«Зачёт» ставится, если студент строит свой ответ в соответствии с планом. В ответе представлены различные подходы к проблеме. Устанавливает содержательные межпредметные связи. Развернуто аргументирует выдвигаемые положения, приводит убедительные примеры, обнаруживает последовательность анализа. Выводы правильны. Речь грамотна, используется профессиональная лексика. Демонстрирует знание специальной литературы в рамках учебного методического комплекса и дополнительных источников информации.

«Незачёт» ставится, если ответ недостаточно логически выстроен, план ответа соблюдается непоследовательно. Студент обнаруживает слабость в развернутом раскрытии профессиональных понятий. Выдвигаемые положения декларируются, но недостаточно

аргументируются. Ответ носит преимущественно теоретический характер, примеры отсутствуют.

## **5. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)**

### **5.1 Основная литература:**

1. Чернышева, Ю. Г. Бизнес-анализ : учебник / Ю.Г. Чернышева. — Москва : ИНФРА-М, 2023. — 648 с. + Доп. материалы [Электронный ресурс]. — (Высшее образование: Бакалавриат). — DOI 10.12737/1858243. - ISBN 978-5-16-017488-4. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1893972>

2. Бизнес-анализ деятельности организации : учебник / Л.Н. Усенко, Ю.Г., Чернышева, Л.В. Гончарова [и др.] ; под ред. проф. Л.Н. Усенко. — Москва : Альфа-М : ИНФРА-М, 2021. — 560 с. : ил. + Доп. материалы [Электронный ресурс; Режим доступа: <http://www.znanium.com>]. — (Магистратура). - ISBN 978-5-98281-358-9. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1245073>

### **5.2 Дополнительная литература:**

1 Тюрин Ю.Н. Анализ данных на компьютере: Учебное пособие / Тюрин Юрий Николаевич, Макаров Алексей Алексеевич; Рец. С.А.Айвазян, В.Н.Тутубалин; Науч.ред. В.Э.Фигурнов. - 4-е изд., перераб. - М.: ФОРУМ, 2010. -368с.: ил. - (Высшее образование). - Прил.:табл.мат.статистики:с.324.- Лит.:с.355. - ISBN 978-58199-0356-8.

2 Елиферов В.Г. Бизнес-процессы: Регламентация и управление: Учебник / Елиферов Виталий Геннадьевич, Репин Владимир Владимирович; Науч.ред.совет В.И.Видяпин и др. - М.: ИНФРА-М, 2008. - 319с. - (Учебники для программы МВА). - Лит. - ISBN 978-5-16-001825-6.

3 Реинжиниринг бизнес-процессов:[Электронный ресурс] : Учебное пособие / Блинов Андрей Олегович, Захаров И.В., Рудакова О.С., Захаров В.Я.; Под ред. А.О.Блинова; Рец. А.Е.Суглобов. - М.: ЮНИТИ-ДАНА, 2012. - 342с. - Библиогр.список:с.334-338. - ISBN 978-5-238-01823-2.

4 Ильин В.В. Реинжиниринг бизнес-процессов с помощью ARIS: Пособие по подготовке к внедрению корпоративной информационной системы управления компанией / Ильин Владислав Владимирович. - 2-е изд. - М.: Вильямс, 2008. - 256с.: ил. - (Практика реального бизнеса). - Прил.-Лит.:с.245- 246.Предм.указ.:с.247-249. - ISBN 978-5-8459-1338-8.

### **5.3. Периодические издания, онлайн-курсы, медиатеки, образовательные курсы нового поколения (МООС - Massive Open Online Course):**

1. PC Magazine/Russian Edition / учредитель: ЗАО «СК Пресс»; гл. ред. О. Лебедев. - М.: СК Пресс.

2. Информационные ресурсы России: [Электронный ресурс]: научно-практический журнал / учредители: ФГБУ «Российское энергетическое агентство» (РЭА) Минэнерго России; гл. ред. Ю.Ю. Ухин. - М.: РЭА.

3. Информационные технологии: теоретический и прикладной научно-технический журнал / учредитель: Изд-во «Новые технологии»; гл. ред. И.П. Норенков. - М.: Новые технологии.

4. Лекториум — академический образовательный проект, "Обзорный курс по анализу данных", Юлия Киселева <https://www.lektorium.tv/course/22822>

5. Школа анализа данных Яндекс, видеолекции курса «Машинное обучение», К.В.Воронцов <https://yandexdataschool.ru/edu-process/courses/machine-learning#item-14>

6. Coursera Inc., курс "Анализ данных: финальный проект", <https://www.coursera.org/learn/data-analysis-project>

7. Применение Data mining в продажах <https://habrahabr.ru/post/158047/>

**Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля)**

1. Научная электронная библиотека Российского фонда фундаментальных исследований (РФФИ) <http://www.elibrary.ru/>
2. КиберЛенинка — научная электронная библиотека, <https://cyberleninka.ru/>
1. Доступ к базам данных компании EBSCO Publishing, <http://search.ebscohost.com/>
2. Электронный доступ к авторефератам <http://vak.ed.gov.ru/search/>  
<http://vak.ed.gov.ru/announcements/techn/581/>
3. Электронная библиотека диссертаций Российской Государственной Библиотеки (РГБ) <http://diss.rsl.ru/>
4. Информационная система «Единое окно доступа к образовательным ресурсам» <http://window.edu.ru/window>
5. Библиотека электронных учебников <http://www.book-ua.org/>
6. <http://www.statsoft.ru> STATISTICA Neural Networks <http://www.statsoft.ru/home/products/version6/snn.htm#tackling>
7. Приступаем к работе в Project Expert  
[http://www.masterplans.ru/rukovodstvo\\_project\\_expert.html](http://www.masterplans.ru/rukovodstvo_project_expert.html)

## 6. Методические указания для обучающихся по освоению дисциплины (модуля)

Самостоятельная работа студентов является неотъемлемой частью процесса подготовки. Дисциплину рекомендуется изучать путем систематической проработки лекционного материала, самостоятельной проработки рекомендуемой литературы, руководств и методических указаний к выполнению практических занятий. Цель самостоятельной работы – расширение кругозора и углубление знаний в области финансового инструментария.

Контроль за выполнением самостоятельной работы проводится при изучении каждой темы дисциплины на семинарских занятиях. Это текущий опрос, тестовые задания, контрольная работа.

В часы, отведенные для самостоятельной работы, студенты под руководством преподавателя обязаны выполнять индивидуальные практические задания, полученные на практических занятиях. При выполнении этих заданий необходимо использовать теоретический материал, делать ссылки на соответствующие формулы, проверять выполнимость предпосылок, необходимых для применения того или иного метода.

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная учебная работа (консультации) – дополнительное разъяснение учебного материала.

Индивидуальные консультации по предмету являются важным фактором, способствующим индивидуализации обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или лицом с ограниченными возможностями здоровья.

## 7. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине (модулю).

№	Вид работ	Материально-техническое обеспечение дисциплины (модуля) и оснащенность
1.	Лекционные занятия	Аудитори и, укомплектованные презентационной техникой (проектор, экран, ноутбук) и

		<p>прикладным программным обеспечением (Microsoft Office).</p> <p>Ауд. 520А, 207Н, 208Н, 209Н, 212Н, 214Н, 201А, 205А, 4033Л, 4038Л, 4039Л, 5040Л, 5041Л, 5042Л, 5045Л, 5046Л</p>
2.	Лабораторные занятия	<p>Аудитории А208Н, 202А, 210Н, 216Н, 513А, 514А, 515А, 516А, а также аудитории, укомплектованные презентационной техникой (проектор, экран, ноутбук) и прикладным программным обеспечением (Microsoft Office).</p> <p>Ауд., 2026Л, 2027Л, 4034Л, 4035Л, 4036Л, 5043Л, 201Н, 202Н, 203Н, А203Н</p>
3.	Групповые (индивидуальные) консультации	Кафедра Теоретической экономики (ауд. 223, 224, 230, 236, 206А, 205Н, 218Н), ауд. А208Н
4.	Помещения для самостоятельной работы, с рабочими местами, оснащенными компьютерной техникой с подключением к сети «Интернет» и обеспечением неограниченного доступа в электронную информационно-образовательную среду организации для каждого обучающегося, в соответствии с объемом изучаемых дисциплин	Ауд. 213А, 218А

