

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ  
ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ**

Направление подготовки:  
09.03.03 Прикладная информатика

Направленность (профиль)  
Искусственный интеллект и анализ данных

Уровень высшего образования: бакалавриат

## 1. Цели и задачи освоения дисциплины

Цель - математические модели интеллектуального анализа данных типовых прикладных задач.

Для достижения поставленной цели при изучении дисциплины решаются следующие задачи:

управления социальными и экономическими системами,  
знать условия применимости и ограничения методов интеллектуального анализа данных,  
изучить подходы для интерпретации полученных результатов,  
проводить формализацию задачи,  
строить описательные и прогнозные модели с помощью современных программных аналитических средств,  
оценивать и интерпретировать полученные результаты,  
реализовывать алгоритмы предобработки и постобработки данных.

## 2. Место дисциплины в структуре образовательной программы

Дисциплина «Интеллектуальный анализ данных» относится к части, формируемой участниками образовательного процесса, учебного плана ОПОП ВО по направлению подготовки 09.03.03 Прикладная информатика.

## 3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

### 3.1. Компетенции обучающегося, формируемые в результате освоения дисциплины с текстом:

Процесс изучения дисциплины направлен на формирование у обучающихся следующих компетенций (результатов освоения образовательной программы):

Коды компетенций	Содержание компетенций
ПК-7	Способность выполнять интеллектуальный анализ больших данных
ПК-8	Способность разрабатывать методы извлечения, анализа и обработки информации
ПК-9	Способность применять математические методы моделирования процессов обработки информации с использованием средств интеллектуального анализа данных

### 3.2. Взаимосвязь планируемых результатов обучения по дисциплине с формируемыми компетенциями образовательной программы

Коды и формулировка компетенции	Индикаторы компетенции	Запланированные результаты обучения
ПК-7 Способность выполнять интеллектуальный анализ больших данных	ПК 7.1 Знать методы и инструментальные средства интеллектуального анализа больших данных ПК 7.2 Уметь выбирать средства представления результатов аналитики больших данных ПК 7.3 Владеть техническими, программными средствами для	ПК 7.1.1 знать методы и инструментальные средства для машинного обучения и анализа больших данных - предметную область анализа больших данных - теоретические и прикладные основы анализа данных ПК 7.1.2 уметь использовать инструментальные средства машинного обучения и анализа больших данных - проводить сравнительный анализ методов и инструментальных средств анализа больших данных ПК 7.1.3 владеть инструментальными средствами машинного обучения и анализа больших данных ПК 7.2.1 знать средства визуализации результатов решения задач машинного обучения и анализа больших данных методы интерпретации и

	разработки алгоритмов и программ в области интеллектуального анализа данных	визуализации решения задач машинного обучения и анализа больших данных ПК 7.2.2 уметь использовать стандартные программные библиотеки для визуализации решений задач машинного обучения и анализа данных ПК 7.2.3 владеть стандартными программными библиотеками для решения задач машинного обучения и анализа данных ПК 7.3.1 знать программные платформы и библиотеки для решения задач машинного обучения и анализа больших данных стандарты проведения анализа данных ПК 7.3.2 уметь использовать программные платформы и библиотеки для решения задач машинного обучения и анализа больших данных ПК 7.3.3 владеть программными средствами для разработки алгоритмов машинного обучения, алгоритмами построения искусственных нейронных сетей
ПК-8 Способность разрабатывать методы извлечения, анализа и обработки информации	ПК 8.1 Знать теоретические и прикладные основы анализа больших данных ПК 8.2 Уметь проводить анализ больших данных ПК 8.3 Владеть методами извлечения информации и знаний из гетерогенных, мультиструктурированных и неструктурированных источников	ПК 8.1.1 знать особенности методов обработки и анализа данных - современный опыт использования анализа больших данных ПК 8.1.2 уметь решать прикладные задачи с помощью методов интеллектуального анализа данных - разрабатывать и оценивать модели больших данных ПК 8.1.3 владеть теоретическими знаниями в области прикладной статистики и машинного обучения ПК 8.2.1 знать факторы, влияющие на эффективность работы методов анализа больших данных ПК 8.2.2 уметь формализовать задачу интеллектуального анализа данных в части методов поиска ассоциативных правил, кластеризации и прогнозирования ПК 8.2.3 владеть проводить очистку данных методами поиска ассоциативных правил, кластеризации и прогнозирования ПК 8.3.1 знать стандартные методы извлечения информации из разнородных источников - типы и виды источников данных ПК 8.3.2 уметь использовать алгоритмы предобработки и постобработки из неструктурированных источников данных - использовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников, определять требования и взаимодействовать с внутренними и внешними поставщиками данных из разнородных источников ПК 8.3.3 владеть методами извлечения информации из неструктурированных источников данных
ПК-9 Способность применять математические методы моделирования процессов обработки информации с использованием средств интеллектуального анализа данных	ПК 9.1 Знать принципы решения задач машинного обучения и интеллектуального анализа данных ПК 9.2 Уметь создавать алгоритмические и математические модели прикладных задач интеллектуального анализа данных ПК 9.3 Владеть навыками	ПК 9.1.1 знать основные принципы решения задач машинного обучения и анализа данных ПК 9.1.2 уметь проводить спецификацию задачи, реализовывать программы на алгоритмических языках высокого уровня, интерпретировать полученные результаты ПК 9.1.3 владеть методами и алгоритмами машинного обучения ПК 9.2.1 знать математические и алгоритмические модели интеллектуального анализа данных, методы оценки временных и стоимостных характеристик машинного обучения и анализа данных

	построения описательных и прогнозных аналитических моделей с использованием современных инструментов интеллектуального анализа данных	ПК 9.2.2 уметь создавать алгоритмические модели типовых прикладных задач обработки информации, реализовывать программы на алгоритмических языках высокого уровня ПК 9.2.3 владеть навыками использования прикладных библиотек для решения задач машинного обучения ПК 9.3.1 описательные и прогнозные аналитические модели для интеллектуального анализа данных - использовать современные программные инструменты интеллектуального анализа данных 9.3.2 разрабатывать и оценивать модели больших данных 9.3.3 навыками построения описательных и прогнозных аналитических моделей с использованием современных инструментов интеллектуального анализа данных, навыками выявления требований к результатам анализа больших данных в предметной области
--	---	---

#### 4. Структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 3 зачетных единицы, 108 часов.

##### 4.1. Объем дисциплины

Вид учебной работы	Всего часов		
	Очная форма обучения	Очно-заочная форма обучения	Заочная форма обучения
Общая трудоемкость дисциплины	108	не реализуется	не реализуется
Контактная работа обучающегося с преподавателем (по видам учебных занятий) (всего), в том числе:	32		
<i>Занятия лекционного типа</i>	16		
<i>Занятия семинарского типа (практич., семин., лаборат. и др.)</i>	16		
<i>Самостоятельная работа под руководством преподавателя</i>	76		
Самостоятельная работа обучающихся СРС/подготовка к экзамену (зачету) в соответствии с учебным планом			
Вид промежуточной аттестации (зачет/экзамен)	Дифференцированный зачет		

##### 4.2. Разделы и темы дисциплины, их трудоемкость по видам учебных занятий

###### 4.2.1. Темы дисциплины, их трудоемкость по видам учебных занятий для очной формы обучения:

Не реализуется

###### 4.2.2. Темы дисциплины, их трудоемкость по видам учебных занятий для очно-заочной формы обучения:

№ п/п	Наименование разделов, тем	Всего часов	Контактная работа обучающегося с	Процедура оценивания/

	учебных занятий		преподавателем (по видам учебных занятий) (всего), в том числе:				оцениваемые компетенции
			ЛЗ	СЗ	ЛР	СРС	
1	Введение в методы интеллектуального анализа данных	12	2			10	Участие в устном опросе, решение задач, выполнение заданий и упражнений для самостоятельной работы / ПК-7, ПК-8, ПК9
2	Выявление структур в данных. Поиск	12	2	6		4	Участие в устном опросе, решение задач, выполнение заданий и упражнений для самостоятельной работы / ПК-7, ПК-8, ПК9
3	Тематическое моделирование	12	2			10	Участие в устном опросе, решение задач, выполнение заданий и упражнений для самостоятельной работы / ПК-7, ПК-8, ПК9
4	Кластеризация	18	2	4		12	Тестирование, решение задач, выполнение заданий и упражнений для самостоятельной работы / ПК-7, ПК-8, ПК9
5	Задачи прогнозирования	12	2	2		8	Участие в устном опросе, решение задач, выполнение заданий и упражнений для самостоятельной работы / ПК-7, ПК-8, ПК9
6	Регрессионный анализ	12	2	2		8	Участие в устном опросе, решение задач, выполнение заданий и упражнений для самостоятельной работы / ПК-7, ПК-8, ПК9
7	Нейронные сети	12	2	2		8	Участие в устном опросе, решение задач, выполнение заданий и упражнений для самостоятельной работы / ПК-7, ПК-8, ПК9

8	Задача классификации и дерева решений	18	2			16	Тестирование, решение задач выполнение заданий и упражнений для самостоятельной работы / ПК-7, ПК-8, ПК9
	Дифференцированный зачет						Сдача диф.зачета
	Итого	108	16	16		76	

#### 4.2.3. Темы дисциплины, их трудоемкость по видам учебных занятий для заочной формы обучения:

Не реализуется

### 4.3. Содержание дисциплины

#### Тема 1 Введение в методы интеллектуального анализа данных

Актуальность и необходимость интеллектуального анализа данных. Проблема больших объемов («Data explosion»).

Интеллектуальный анализ данных (Data Mining): основные определения

Процесс ИАД: Анализ предметной области. Выбор (или разработка) алгоритма анализа. Проверка моделей и представление результатов анализа. Применение построенных моделей. Основные типы исходных данных: транзакционные, табличные, временные ряды и числовые данные большого объема, обработка результатов наблюдений, научных экспериментов, характеристик технологических процессов, электронные тексты на естественном языке о анализ содержимого документов, графовые данные. Основные типы данных для анализа.

#### Тема 2 Выявление структур в данных. Поиск

Типовая прикладная задача: анализ «корзины покупателя». Ассоциативный анализ. Правила с семантикой. Основной математический аппарат (дискретная математика, математическая логика, комбинаторная оптимизация). Тип моделей: «описательный» (descriptive) Data mining. Тип обучения: «без учителя» (unsupervised). Типы правил: булевы, числовые, иерархические, временные, пространственные. Прикладные задачи: «Экономические»: анализ корзины, маркетинг; «Безопасность» и Web usagemining: модели поведения пользователя; Text mining: поиск ключевых слов, характеристик и тематик; Биоинформатика, медицина. Задачи анализа: поиск самих правил, поиск исключений, выделение признаков, классификация и прогнозирование.

Булквы ассоциативные правила. Алгоритм Apriori. Использование метода ветвей и границ. Метод fp-tree. Объективные меры интересности. Использование ограничений: проблема итеративного анализа больших объемов данных; типы ограничений. Интерфейс SAS Enterprise Miner для работы с Ассоциативными правилами.

#### Тема 3 Тематическое моделирование

Тематическое моделирование. Статистическое обучение без учителя. Задача «самоорганизации». Задача поиска неизвестных зависимостей без эксперта «Истинный data mining». Основные задачи обучения без учителя. Поиск скрытых (латентных) признаков. Поиск скрытых структур (групп или зависимостей).

Тематическое моделирование текстовых данных. Тематическая модель. Скрытые признаки документа. Рекомендательные системы. Решение на основе ассоциативных данных. Свертка "транзакционной истории" в вектор признаков. Примеры использования метода главных компонент в SAS EM. Кластеризация переменных. Алгоритм группировки переменных. Неотрицательная матричная факторизация. Пример использования кластеризации переменных в SAS EM. Самоорганизующиеся отображения

#### Тема 4 Кластеризация

Кластеризация: иерархическая, метрическая, вероятностная. Определение кластера.

Кластерный анализ: разбиение множества объектов на группы (кластеры). Тип моделей: описательный, прогнозный, классификация. Этапы кластерного анализа: подготовка данных, применение алгоритма, визуализация и интерпретация результатов.

Качество кластеризации. Требования к методу кластеризации.

Подготовка данных для кластеризации. Отбор наблюдений. Отбор и трансформация переменных. Стандартизация переменных. Основные типы алгоритмов кластеризации: иерархические, на основе группировки (partitioning), на основе связности, модель-ориентированные (статистические).

### **Тема 5 Задачи прогнозирования**

Задача прогнозирования. Задача "Обучения с учителем". Обучающая выборка или тренировочный набор. Этап обучения и этап прогнозирования. Типы задач прогнозирования. Бинарная классификация (разделение). Регрессия. Классификация. Много-темная классификация (multi-label). Ранжирование. Проклятие размерности, переобучение, оценка и выбор моделей, валидация и кросс-валидация.

Метод k-ближайших соседей: общая схема работы. Метод KNN. Метод «взвешенных» k-ближайших соседей. Метод K ближайших соседей с адаптивным расстоянием. Свойства методов KNN: основные свойства, достоинства и недостатки

### **Тема 6 Регрессионный анализ**

Задача линейной регрессии. Уравнение линейной регрессии. Линеаризируемые регрессии: степенная, экспоненциальная, гиперболическая. Цель регрессионного анализа: определение наличия связи между переменными и характера этой связи, предсказание значения зависимой переменной с помощью независимой(-ых), определение вклада отдельных независимых переменных в вариацию зависимой переменной. Линейная модель с линейными эффектами. Линейная модель с нелинейными эффектами.

Метод наименьших квадратов и проблема мультиколлинеарности. Проблема недообучения и переобучения. Оценка и выбор моделей. Валидация и кросс-валидация и бутстреппинг. Выбор модели и регуляризация на примере линейных моделей регрессии. Оценка точности расчета коэффициентов. Проверка гипотез. Оценка общей точности модели. Интерпретация результатов регрессии.

Методы выбора важных переменных.

Нелинейные зависимости: полиномы, ступенчатые функции, сплайны, локальная регрессия, обобщенные аддитивные модели. Полиномиальная регрессия.

Ступенчатые функции. Кусочные полиномы. Линейные сплайны. Кубические сплайны. Естественные кубические сплайны. Размещение узлов. Сглаживание сплайнами.

Локальная регрессия. Обобщенные аддитивные модели. Регрессия Ridge и Lasso.

### **Тема 7 Нейронные сети**

Нейронные сети. Искусственный нейрон. Математическая модель для решения задач машинного обучения. Нейронная сеть (искусственная). Сети прямого распространения (Feed forward Neural Networks). Рекуррентные нейронные сети (Recurrent Neural Networks). Обучение нейронных сетей (с учителем). Правила обучения нейросетей. Правило Хэбба. Обобщенное дельта-правило. Универсальный аппроксиматор. Линейный персептрон. Функции активации.

Многослойный персептрон. Персептрон с прямыми соединениями. Два и более скрытых слоя. Форма сигмоида. Разложение по базису сигмоидальных функций. Типы решаемых задач. Радиально-базисные сети. Форма функции гаусса. RBF нейронная сеть. Проблема локального эффекта. RBF-нормализованная нейронная сеть. Проблема локальных минимумов. Обучение нейронных сетей: критерии сходимости, оценки отклонения, комбинации функций активации и распределения ошибок. Первые итерации стадии обучения. Оценки максимального правдоподобия.

Робастные оценки. Выбор распределения для обобщенной линейной модели. Задача оптимизации. Итерационные методы. Обратное распространение ошибки (градиентный метод). Быстрое обратное распространение ошибки.

Квазиньютоновские методы. Метод Сопряженных градиентов. Метод доверительных областей (trusted regions). Комбинированный (градиент+ньютон) Double-Dogleg. Интерфейс работы с нейронными сетями в SAS Miner. SVM Постановка и решение задачи. Метод множителей Лагранжа. Ранняя остановка – борьба с переобучением

#### **Тема 8 Задача классификации и деревья решений**

Задача классификации. Методы Байеса: вероятностная постановка задачи прогнозирования, возможность дообучения, вероятностный прогноз, де факто - эталон. Методы, основанные на деревьях решений. Применение деревьев решений для задач классификации: непрерывный отклик

Деревья решений в задачах классификации и регрессии. Дерево решений - граф. Процесс построения деревьев решений. «Отсечение» ветвей pruning - выявление и удаление ветвей (решений), приводящих к шуму или к выбросам. Проверка атрибутов – путь по ветви до листа. Минимизация целевого критерий – критерий разбиения. Жадный подход - рекурсивное разделение. Поиск разбиения по переменным. Множественные разбиения. Критерии разбиения. Индекс Джини, Энтропия, Logworth.

Переобучение и недообучение. Точность дерева. Оценка модели. Особенности популярных алгоритмов построения деревьев решений. Преимущества и недостатки деревьев решений

### **4.4. Темы семинарских занятий и лабораторных работ**

#### **Тема 1 Введение в методы интеллектуального анализа данных**

Решение задач по теме 1.

#### **Тема 2 Выявление структур в данных. Поиск**

Решение задач по теме 2.

#### **Тема 3 Тематическое моделирование**

Решение задач по теме 3.

#### **Тема 4 Кластеризация**

Решение задач по теме 4. Тестирование по темам 1,2,3 и 4.

#### **Тема 5 Задачи прогнозирования**

Решение задач по теме 5.

#### **Тема 6 Регрессионный анализ**

Решение задач по теме 6.

#### **Тема 7 Нейронные сети**

Решение задач по теме 7.

#### **Тема 8 Задача классификации и деревья решений**

Решение задач по теме 8. Тестирование по темам 1, 2, 3,4, 5, 6, 7 и 8

## **5. Учебно-методическое обеспечение самостоятельной работы**

### **5.1. Задания для самостоятельной подготовки к занятиям семинарского типа**

Семинарское занятие №1.

Тема занятия: Поиск ассоциативных правил

1. Запустите SAS Miner, создайте новый проект с именем “Prak”
2. В дереве проекта создайте новую диаграмму с именем Diag1 (правая кнопка мышина папке Diagrams, выберете “Create Diagrams”)
3. В дереве проекта добавьте набор данных из библиотеки SAMPSTO с именем Assocs (правая кнопка мыши на папке DataSources, выберете “Create Data Source”). Этот набор данных содержит информацию о покупках продуктов product клиентом customer в магазине.
4. В процессе подключения набора данных задайте метаданные для набора: роль набора Transaction; переменная Customer – роль ID, тип Nominal; Product – Рольtarget, тип Nominal; Time – роль Sequence, тип Interval.
5. На шаге определения метаданных для переменных выделите переменную Product и

нажмите кнопку Explore. Сколько различных значений принимает переменная?

6. Перетащите узел с добавленным набором данных на диаграмму, подключите посленего узел Associations

7. В разделе settings поставьте настройки support count 100 (не путать с supportpercent) и minimum confidence 30%, постройте модели, нажав в меню Run.

8. По узлу Associations (выбрав Results в меню или сразу после обучения) ответьте на вопросы (используйте Rules table из раздела View->Rules):

a. Найдите самое достоверное правило (если их несколько, то правило с самой большой поддержкой), содержащее продукт cracker (вариант I) или chicken (вариант II) в самом начале цепочки. Что это за правило? Что оно обозначает (дайте словесную интерпретацию)? Какая у него поддержка? Что означает это число?

b. Откройте граф взаимосвязей (View->Rules->Link graph), нажмите правой кнопкой на поле графика и выберите Graph properties, поставьте в разделе Links опцию Show Directed Links.

c. Посмотрите на график и ответьте на вопрос после какого продукта обычноприобретается soda (вариант I) или sardines (вариант II)

9. Ответы на вопросы включите в текстовый файл (в любом формате), запишите диаграмму в формате xml (нажав на диаграмму в дереве проекта и выбрав Save as). Перешлите для проверки текстовый файл с ответами и xml с диаграммой в качестве результата задания.

Семинарское занятие 2.

Тема занятия: Кластеризация

1. Запустите SAS Miner, откройте проект с именем "Prak"

2. В дереве проекта создайте новую диаграмму с именем Diag2

3. В дереве проекта добавьте набор данных из библиотеки SAMPSIO с именем Baseball. Этот набор данных содержит информацию о результатах игроков вбейсбол в 1986 году и за всю предыдущую карьеру.

4. В процессе подключения набора данных задайте метаданные для набора (полезно выбрать чекбокс Label при описании переменных): роль набора Raw; для переменной Name поставьте роль ID, переменной YrMajor (число лет в лиге), а также всем переменным, чье имя начинается с n (результаты 1986 года) или sr (результаты за всю карьеру) – роль Input, тип Interval. Остальным переменным поставьте роль Rejected.

5. Перетащите узел с добавленным набором данных на диаграмму, подключите посленего узел Cluster, а после Cluster узел Segment Profile. Обучите модель.

6. Сколько кластеров было сформировано узлом cluster? Используя информацию из графика CCC Plot (Summary Statistics) узла Cluster ответьте на вопрос, какое значение CCC было для выбранного числа кластеров, а также для числа кластеровна один больше и на один меньше?

7. Используя информацию из графиков узла Cluster (Segment Plot, Means Statistics, Input Mean Plot и другие) и из графиков узла Segment Profile ответьте на вопросы:

a) Вариант I. Какой кластер содержит самых опытных игроков (переменная YrMajor)?

Вариант II. Какой кластер содержит самых успешных по хитам в 1986 году игроков (переменная nHits)?

b) Сколько человек попало в этот кластер?

c) Какая переменная самая важная с точки зрения отделения этого кластера ото всей выборки (см. Segment Profile).

a. Выберите узел Cluster, в настройках нажмите кнопку Exported data, выберите Browse. Ответьте на вопрос, как зовут самого типичного игрока в кластере вашего варианта (у него должно быть минимальное расстояние до центра вашего кластера)?

b. Параллельно с существующей веткой после узла с набором данных подключите узел SOM, задав размер сетки 2 на 2 (раздел настроек Segment, Rows=2, Columns=2) и после узла SOM узел Segment Profile (после обучения SOM, откройте

Edit Variables на узле Segment Profile и поставьте Use = no для переменных скоординатами на решетке SOM\_Dimension).

А. Ответьте на все те же вопросы, что заданы в пункте 7, но после кластеризации SOM.

В. Ответы на вопросы включите в текстовый файл (в pdf формате), запишите диаграмму в формате xml (нажав на диаграмму в дереве проекта и выбрав Save as). Перешлите текстовый файл с ответами и xml с диаграммой в качестве результата задания.

Семинарское занятие 3.

Тема занятия: Выявление скрытых признаков

Задание 1.

1. После узла с подключенными данными Baseball подключите узел Variable Clustering, в настройках поставьте в Разделе Score->Variable Selection->Bestvariables. Обучите модель.

2. На сколько групп было разбито множество переменных? Какие представители каждой группы были выбраны? Какой процент вариации описывают отобранные переменные?

Задание 2.

1. После узла с подключенными данными Baseball подключите узел Principal Components, в настройках поставьте в eigenvalue cutoff->cumulative=0.8 (80% вариации) . Обучите модель.

2. Сколько главных компонент получилось? Какие собственные значения у полученных главных компонент? Глядя на график Principal Component coefficient, как бы вы охарактеризовали одной фразой признак, соответствующий первой главной компоненте, с точки зрения качества игроков, глядя на коэффициенты основных характеристик в первой главной компоненте?

Семинарское занятие 4.

Тема занятия: Прогнозирование. Метод kNN

1) Запустите SAS Miner, откройте проект с именем "Prak".

2) В дереве проекта создайте новую диаграмму с именем Diag3

3) На диаграмму добавьте узел Import Data из раздела Sample и подключите csv набор данных white\_wine. В наборе указана целевая (target) переменная target\_quality типа Interval, которая описывает субъективную оценку качества сорта вина, заданную экспертом. Остальные переменные - входные (все Interval). **Задача**

– **спрогнозировать оценку эксперта** в зависимости от объективных химических показателей вина и выявить важные признаки.

4) После узла Import Data подключите узел Transform Variables и установите дискретизацию числового отклика на квантили: Interval Target = Quantile. Далее из раздела Utility подключите узел Metadata (позволяет менять метаданные внутри процесса анализа). Выберите раздел variables->train и для переменной с префиксом PCTL\_ поставьте новую роль Rejected (чтобы работать только с исходным откликом). Затем подключите узел Data partition, оставьте разбиение по умолчанию 40 – train и по 30 на тестовый и валидационный наборы. Откройте список переменных на этом узле и поставьте роль для переменной с префиксом PCTL\_Stratification. Зачем это делается?

5) Подключите узел Transform variables после Data Partition и сделайте настройку в разделе Interval Inputs – Standardize и подключите MBR после transform variables. Укажите в качестве числа соседей Number of Neighbors стандартную эвристику примерно  $\sqrt{N}$ , задайте 20. После MBR подключите узел Model

Comparison. Сделайте в нем настройку в разделе Model selection->Selection Statistics->Average Squared Error, Selection Table - > Test. Какие значения ASE получились на тренировочном, валидационном и тестовом наборе данных?

6) После узла Transform variables и перед узлом MBR вставьте узел PCA (**Вариант I**), Variable clustering (**Вариант II**). Сколько переменных теперь передается на вход MBR? Как изменились значения ASE в результате? Почему это могло произойти?

Семинарское занятие 5.

Тема занятия: Прогнозирование. Регрессия

*первые 6 шагов на базе предыдущего задания*

- 1) Запустите SAS Miner, откройте проект с именем “Prak”.
- 2) В дереве проекта создайте новую диаграмму с именем Diag3
- 3) На диаграмму добавьте узел Import Data из раздела Sample и подключите csv набор данных white\_wine. В наборе указана целевая (target) переменная target\_quality типа Interval, которая описывает субъективную оценку качества сорта вина, заданную экспертом. Остальные переменные - входные (все Interval). **Задача** – **спрогнозировать оценку эксперта** в зависимости от объективных химических показателей вина и выявить важные признаки.
- 4) После узла Import Data подключите узел Transform Variables и установите дискретизацию числового отклика на квантили: Interval Target = Quantile. Далее из раздела Utility подключите узел Metadata (позволяет менять метаданные внутри процесса анализа). Выберите раздел variables->train и для переменной с префиксом PCTL\_ поставьте новую роль Rejected (чтобы работать только с исходным откликом). Затем подключите узел Data partition, оставьте разбиение по умолчанию 40 – train и по 30 на тестовый и валидационный наборы. Откройте список переменных на этом узле и поставьте роль для переменной с префиксом PCTL\_Stratification. Зачем это делается?
- 5) Подключите узел Transform variables после Data Partition и сделайте настройку в разделе Interval Inputs – Standardize и подключите MBR после transform variables. Укажите в качестве числа соседей Number of Neighbors стандартную эвристику примерно  $\sqrt{N}$ , задайте 20. После MBR подключите узел Model Comparison. Сделайте в нем настройку в разделе Model selection->Selection Statistics->Average Squared Error, Selection Table -> Test. Какие значения ASE получились на тренировочном, валидационном и тестовом наборе данных?
- 6) После узла Transform variables и перед узлом MBR вставьте узел PCA (**Вариант I**), Variable clustering (**Вариант II**). Сколько переменных теперь передается на вход MBR? Как изменились значения ASE в результате? Почему это могло произойти?
- 7) Параллельно от узла Data Partition подключите узел Regression и сделайте следующие настройки. В разделе Model selection задайте метод Selection Model Backward (**Вариант I**) и Forward (**Вариант II**). Критерий для выбора модели Selection Criterion поставьте Validation Error. Выход узла регрессии соедините с узлом сравнения моделей. Какая модель показывает лучшее качество на тестовом наборе? Какие переменные не вошли в результирующую регрессионную модель? На каком шаге была выбрана лучшая модель (см. график Iteration plot (View->Model)? Посмотрите в журнале (раздел Output) какая из вошедших переменных наименее важна с точки зрения t-статистики?
- 8) Добавьте узел Transform variables перед узлом Regression и сделайте настройку Interval Inputs – Maximum Normal. Как изменились результаты регрессионной модели с точки зрения числа степеней свободы и значения ASE на тестовом наборе?
- 9) Отредактируйте уравнение регрессии, чтобы оно стало полным полиномом второй степени, в разделе Equation задайте Polynomial Terms = Yes, Two-factor interactions = yes. Normal. Как изменились результаты регрессионной модели с точки зрения числа степеней свободы и значения ASE на тестовом наборе?
- 10) После узла Data partition добавьте узел PLS и соедините его выход с узлом сравнения моделей. Задайте алгоритм Regression method = PLS (**I Вариант**), PCR (**II вариант**). Задайте настройки автоматического отбора числа факторов: CV Method = Test Set в Cross Validation. Сколько компонент было выбрано? Какие переменные были отобраны по критерию VIP (**I Вариант**), по абсолютным значениям коэффициентов (**II Вариант**). Какое значение ASE модель показывает на тестовом наборе?
- 11) После узла Data partition добавьте узел LARS и соедините его выход с узлом сравнения моделей. Задайте алгоритм отбора переменных LASSO и критерий для выбора моделей AIC (**I вариант**) или SBC (**II Вариант**). Какое значение ASE модель

показывает на тестовом наборе? Какая переменная была отобрана на первом шаге? Сколько всего переменных было отобрано? Обратите внимание на компоненту Iteration Plot в результатах. Если бы вместо критерия отбора модели вашего варианта использовался критерий ASE на валидационном наборе, то было бы отобрано больше или меньше переменных?

Ответы на вопросы включите в текстовый файл (в pdf формате), запишите диаграмму в формате xml (нажав на диаграмму в дереве проекта и выбрав Save as).

Семинарское занятие 6.

Тема занятия: Нейронные сети

1. Запустите SAS Enterprise Miner, откройте проект с именем “Prak”.
2. В дереве проекта откройте диаграмму с именем Diag3
3. Подключите узел «Neural Networks» после узла Data Partition, выберите Model Selection Criteria -> Average Error, выберите архитектуру Network->architecture, **Вариант I:**

“Multilayer perception”, **Вариант II:** “Normalized Radial Basis Network –unequal width and height“, обучите модель.

1. Сколько степеней свободы получилось в сети (см. Fit Statistics)? Почему?
2. Каково значение среднеквадратичной ошибки ASE на тренировочном, тестовом и валидационном наборах?

3. Обратите внимание на график Iteration plot: на каком шаге обучения выбрана оптимальная модель?

4. Переместите ваш узел «Neural Networks» после узла “LARS”. Обучите модель.

1. Сколько степеней свободы получилось в сети (см. Fit Statistics)? Почему?
2. Каково значение среднеквадратичной ошибки ASE на тренировочном, тестовом и валидационном наборах?

3. Обратите внимание на график Iteration plot: на каком шаге обучения выбрана оптимальная модель?

5. Выберите архитектуру Network->architecture и поставьте число нейронов скрытого слоя Number of hidden units = 7, обучите модель.

1. Сколько степеней свободы получилось в сети (см. Fit Statistics)? Почему?
2. Каково значение среднеквадратичной ошибки ASE на тренировочном, тестовом и валидационном наборах?

3. Обратите внимание на график Iteration plot: на каком шаге обучения выбрана оптимальная модель?

6. Ответы на вопросы включите в текстовый файл (в pdf формате), запишите диаграмму в формате xml (нажав на диаграмму в дереве проекта и выбрав Save as).

Результат: текстовый файл с ответами и xml с диаграммой.

## 5.2. Задания и упражнения для самостоятельной работы

№ п/п	Наименование темы дисциплины	Задания и упражнения для самостоятельной работы
1	Введение в методы интеллектуального анализа данных Выявление структур в данных. Поиск Тематическое моделирование Кластеризация Задачи прогнозирования Регрессионный анализ Нейронные сети Задача классификации	Дан набор данных, в котором описана история предложений клиентам банка застраховать свои вклады. Целевая бинарная переменная INS, содержит признак, согласился ли клиент приобрести такую услугу или нет. Каждый клиент имеет свой уникальный ID. Остальные переменные – входные. Информацию о них можно посмотреть, включив опции «label»  при добавлении переменных.

<p>и дерева решений Дифференцированный зачет</p>	<p>Определите метаданные для источника указано в таблице ниже: Variable Name Role Measurement Level Label</p> <ol style="list-style-type: none"> <li>1. ACCTAGE INPUT INTERVAL Age of Oldest Account</li> <li>2. AGE INPUT INTERVAL Age</li> <li>3. ATM INPUT BINARY ATM</li> <li>4. ATMAMT INPUT INTERVAL ATM Withdrawal Amount</li> <li>5. BRANCH INPUT NOMINAL Branch of Bank</li> <li>6. CASHBK INPUT INTERVAL Number Cash Back</li> <li>7. CC INPUT BINARY Credit Card</li> <li>8. CCBAL INPUT INTERVAL Credit Card Balance</li> <li>9. CCPURC INPUT INTERVAL Credit Card Purchases</li> <li>10. CD INPUT BINARY Certificate of Deposit</li> <li>11. CDBAL INPUT INTERVAL CD Balance</li> <li>12. CHECKS INPUT INTERVAL Number of Checks</li> <li>13. CRSCORE INPUT INTERVAL Credit Score</li> <li>14. DDA INPUT BINARY Checking Account</li> <li>15. DDABAL INPUT INTERVAL Checking Balance</li> <li>16. DEP INPUT INTERVAL Checking Deposits</li> <li>17. DEPAMT INPUT INTERVAL Amount Deposited</li> <li>18. DIRDEP INPUT BINARY Direct Deposit</li> <li>19. HMOWN INPUT BINARY Owns Home</li> <li>20. HMVAL INPUT INTERVAL Home Value</li> <li>21. id ID NOMINAL</li> <li>22. ILS INPUT BINARY Installment Loan</li> <li>23. ILSBAL INPUT INTERVAL Loan Balance</li> <li>24. INAREA INPUT BINARY Local Address</li> <li>25. INCOME INPUT INTERVAL Income</li> <li>26. INS TARGET BINARY Insurance Product</li> <li>27. INV INPUT BINARY Investment</li> <li>28. INVBAL INPUT INTERVAL Investment Balance</li> <li>29. IRA INPUT BINARY Retirement Account</li> <li>30. IRABAL INPUT INTERVAL IRA Balance</li> <li>31. LOC INPUT BINARY Line of Credit</li> <li>32. LOCBAL INPUT INTERVAL Line of Credit Balance</li> <li>33. LORES INPUT INTERVAL Length of Residence</li> <li>34. MM INPUT BINARY Money Market</li> <li>35. MMBAL INPUT INTERVAL Money Market Balance</li> <li>36. MMCRED INPUT INTERVAL Money Market Credits</li> <li>37. MOVED INPUT BINARY Recent Address Change</li> <li>38. MTG INPUT BINARY Mortgage</li> <li>39. MTGBAL INPUT INTERVAL Mortgage Balance</li> <li>40. NSF INPUT BINARY Number Insufficient Fund</li> <li>41. NSFAMT INPUT INTERVAL Amount NSF</li> <li>42. PHONE INPUT NOMINAL Number Telephone Banking</li> <li>43. POS INPUT INTERVAL Number Point of Sale</li> <li>44. POSAMT INPUT INTERVAL Amount Point of Sale</li> <li>45. RES INPUT NOMINAL Area Classification</li> <li>46. SAV INPUT BINARY Saving Account</li> <li>47. SAVBAL INPUT INTERVAL Saving Balance</li> <li>48. SDB INPUT BINARY Safety Deposit Box</li> <li>49. TELLER INPUT INTERVAL Teller Visits</li> </ol>
--	---

		<p>50. _dataobs_ REJECTED INTERVAL Observation Number</p> <p>Необходимо построить модель прогнозирования для бинарного отклика, которая будет наилучшим образом его предсказывать. Можно использовать любые методы, которые реализованы в SAS Enterprise Miner.</p>
--	--	---

#### 5.4. Перечень тем (задания) для курсовой работы

Не предусмотрено

### 6.Оценочные средства для проведения текущего контроля и промежуточной аттестации обучающихся

6.1 Примерные оценочные средства, включая тестовые оценочные задания для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине (модулю) приведены в Приложении 1 к рабочей программе дисциплины.

6.2 Оценочные средства для проведения промежуточной аттестации обучающихся по дисциплине (модулю) включают следующие разделы:

- перечень компетенций, формируемых в процессе освоения учебной дисциплины;
- описание показателей и критериев оценивания компетенций, описание шкал оценивания;
- типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения по учебной дисциплине, обеспечивающих достижение планируемых результатов освоения образовательной программы;
- методические материалы, определяющие процедуры оценивания результатов обучения по учебной дисциплине, обеспечивающих достижение планируемых результатов освоения образовательной программы.

### 7.Учебно-методическое и информационное обеспечение дисциплины

#### а) основная литература:

1. Орлов, А. И. Искусственный интеллект: статистические методы анализа данных : учебник / А. И. Орлов. — Москва : Ай Пи Ар Медиа, 2022. — 843 с. — ISBN 978-5-4497-1470-1. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/117029.html> (дата обращения: 03.02.2022). — Режим доступа: для авторизир. пользователей

2. Макрусев, В. В. Основы системного анализа : учебник / В. В. Макрусев. — 2-е изд. — Санкт-Петербург : Троицкий мост, 2022. — 250 с. — ISBN 978-5-4377-0138-6. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/111173.html> (дата обращения: 03.02.2022). — Режим доступа: для авторизир. пользователей

#### б) дополнительная литература:

1. Операционные системы : учебное пособие для бакалавров / составители И. В. Винокуров. — Москва : Ай Пи Ар Медиа, 2022. — 133 с. — ISBN 978-5-4497-1406-0. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/115696.html> (дата обращения: 03.02.2022). — Режим доступа: для авторизир. пользователей

#### в) интернет-ресурсы:

1. ЭБС IPR Books <http://iprbookshop.ru> (учебники и учебные пособия, монографии, сборники научных трудов, научная периодика, профильные журналы, справочники, энциклопедии);
2. ООО «ИВИС» <https://dlib.eastview.com> (электронные версии периодических изданий ООО «ИВИС»);
3. Web of Science <http://webofknowledge.com/> (обширная международная универсальная реферативная база данных);

4. Scopus <https://www.scopus.com> (международная универсальная реферативная база данных, индексирующая более 21 тыс. наименований научно-технических, гуманитарных и медицинских журналов, материалов конференций примерно 5000 международных издательств);
5. Научная электронная библиотека eLIBRARY.RU <https://elibrary.ru> (крупнейший российский информационный портал в области науки, технологии, медицины и образования);
6. Единая межведомственная информационно-статистическая система (ЕМИСС) <https://fedstat.ru/indicators/>
7. Федеральная служба государственной статистики <http://www.gks.ru/>
8. Министерство науки и высшего образования Российской Федерации <https://minobrnauki.gov.ru/>
9. Министерство экономического развития Российской Федерации <http://economy.gov.ru/minec/about/structure>
10. Федеральное агентство по печати и массовым коммуникациям <http://www.fapmc.ru/rospechat.html>
11. Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации <https://digital.gov.ru/ru>
12. <http://www.mathnet.ru/> Общероссийский математический портал (информационная система)

## 8. Материально-техническое обеспечение дисциплины

№	Наименование учебных аудиторий (лабораторий) и помещений для самостоятельной работы	Перечень оборудования и технических средств обучения	Программное обеспечение, в т.ч. отечественного производства
1	Учебные аудитории для проведения занятий лекционного типа	<p>Преподавательский стол;  столы обучающихся;  стулья;  классная доска;  мультимедийный комплекс;  наглядные пособия (плакаты)  <i>Место, оборудованное для лиц с ограниченными возможностями и.</i>  Лицензионное программное обеспечение, подключенное к Интернет</p>	<ol style="list-style-type: none"> <li>1. 7-Zip (Бесплатное ПО);</li> <li>2. 10-Strike Network Inventory ПО РФ (ПО)</li> <li>3. Ductor Academic ПО РФ (Бесплатное ПО); <a href="https://basegroup.ru/deductor/manual/licence-deductor-academic">https://basegroup.ru/deductor/manual/licence-deductor-academic</a></li> <li>4. Eset Endpoint security (Платное ПО) <a href="https://help.eset.com/eula/GIMP">https://help.eset.com/eula/GIMP</a> (Бесплатное ПО); <a href="https://docs.gimp.org/2.10/ru/">https://docs.gimp.org/2.10/ru/</a></li> <li>5. microsoft office профессиональный плюс 2016 (ПО) <a href="https://www.microsoft.com/en-us/Useterms/Retail/Office/2016Professional/Useterms_Retail_Office_2016Professional_RUS.htm">https://www.microsoft.com/en-us/Useterms/Retail/Office/2016Professional/Useterms_Retail_Office_2016Professional_RUS.htm</a></li> <li>6. Microsoft power Bi (Бесплатное ПО); <a href="https://powerbi.microsoft.com/ru-ru/windows-license-terms/">https://powerbi.microsoft.com/ru-ru/windows-license-terms/</a></li> <li>7. Notepad ++ (Бесплатное ПО); <a href="https://visualstudio.microsoft.com/ru/license-terms/mt171584/">https://visualstudio.microsoft.com/ru/license-terms/mt171584/</a></li> <li>8. Zoom (Бесплатное ПО); <a href="https://www.gnu.org/licenses/old-licenses/gpl-2.0.html">https://www.gnu.org/licenses/old-licenses/gpl-2.0.html</a></li> <li>9. Anaconda3 2019 (Бесплатное ПО); <a href="https://www.gnu.org/licenses/old-licenses/gpl-2.0.html">https://www.gnu.org/licenses/old-licenses/gpl-2.0.html</a></li> <li>10. <a href="https://www.anaconda.com/eula-anaconda-individual-edition">https://www.anaconda.com/eula-anaconda-individual-edition</a></li> <li>11. Android studio (Бесплатное ПО); <a href="https://developer.android.com/studio/terms">https://developer.android.com/studio/terms</a></li> <li>12. Brackets (Бесплатное ПО); <a href="https://github.com/brackets-cont/brackets/blob/master/LICENSE">https://github.com/brackets-cont/brackets/blob/master/LICENSE</a></li> <li>13. CodeBlocks (Бесплатное ПО); <a href="https://www.codeblocks.org/license/">https://www.codeblocks.org/license/</a></li> <li>14. Firebird (Бесплатное ПО); <a href="https://firebirdsql.org/en/licensing/">https://firebirdsql.org/en/licensing/</a></li> <li>15. KNIME analytics platform (Бесплатное ПО); <a href="https://www.knime.com/downloads/full-license">https://www.knime.com/downloads/full-license</a></li> <li>16. Loginom community РФ ПО (Бесплатное ПО); <a href="https://loginom.ru/legal">https://loginom.ru/legal</a></li> <li>17. Monogame SDK (Бесплатное ПО); <a href="https://github.com/MonoGame/MonoGame/blob/develop/LICENSE.txt">https://github.com/MonoGame/MonoGame/blob/develop/LICENSE.txt</a></li> <li>18. Openproj (Бесплатное ПО); <a href="https://opensource.org/licenses/CPAL-1.0">https://opensource.org/licenses/CPAL-1.0</a></li> </ol>

			<p>16. tableau 2019 (Бесплатное ПО); <a href="https://www.tableau.com/legal">https://www.tableau.com/legal</a></p> <p>17. Visual studio community 2017 (Бесплатное ПО); <a href="https://www.google.com/url?sa=t&amp;rct=j&amp;q=&amp;esrc=s&amp;source=web&amp;cd=&amp;ved=2ahUKEwi-qZeRxv7zAhXhsYsKHZoRBAsQFnoECBgQAQ&amp;url=https%3A%2F%2Fvisualstudio.microsoft.com%2Fwp-content%2Fuploads%2F2017%2F01%2FVS2017_COMMUNITY_RC_RUS_Eula.1049-1.docx&amp;usg=AOvVaw0tLxIQ4E2McNypfRn9tTo">https://www.google.com/url?sa=t&amp;rct=j&amp;q=&amp;esrc=s&amp;source=web&amp;cd=&amp;ved=2ahUKEwi-qZeRxv7zAhXhsYsKHZoRBAsQFnoECBgQAQ&amp;url=https%3A%2F%2Fvisualstudio.microsoft.com%2Fwp-content%2Fuploads%2F2017%2F01%2FVS2017_COMMUNITY_RC_RUS_Eula.1049-1.docx&amp;usg=AOvVaw0tLxIQ4E2McNypfRn9tTo</a></p> <p>18. Visual studio community 2019 (Бесплатное ПО); <a href="https://visualstudio.microsoft.com/ru/license-terms/mlt110718/">https://visualstudio.microsoft.com/ru/license-terms/mlt110718/</a></p> <p>19. Консультант плюс</p>
2	Компьютерный класс	<p>Преподавательский стол; столы обучающихся; стулья; классная доска; мультимедийный комплекс; ПК преподавателя; ПК обучающихся; наглядные пособия (плакаты) <i>Место, оборудованное для лиц с ограниченными возможностями и.</i> Лицензионное программное обеспечение, подключенное к Интернет</p>	<p>1. 7-Zip (Бесплатное ПО);</p> <p>2. 10-Strike Network Inventory ПО РФ (ПО)</p> <p>3. Ductor Academic ПО РФ (Бесплатное ПО); <a href="https://basegroup.ru/deductor/manual/licence-deductor-academic">https://basegroup.ru/deductor/manual/licence-deductor-academic</a></p> <p>4. Eset Endpoint security (Платное ПО) <a href="https://help.eset.com/eula/GIMP">https://help.eset.com/eula/GIMP</a> (Бесплатное ПО); <a href="https://docs.gimp.org/2.10/ru/">https://docs.gimp.org/2.10/ru/</a></p> <p>5. microsoft office профессиональный плюс 2016 (ПО) <a href="https://www.microsoft.com/en-us/Useterms/Retail/Office/2016Professional/Useterms_Retail_Office_2016Professional_RUS.htm">https://www.microsoft.com/en-us/Useterms/Retail/Office/2016Professional/Useterms_Retail_Office_2016Professional_RUS.htm</a></p> <p>6. Microsoft power Bi (Бесплатное ПО); <a href="https://powerbi.microsoft.com/ru-ru/windows-license-terms/microsoft-visual-studio">https://powerbi.microsoft.com/ru-ru/windows-license-terms/microsoft-visual-studio</a> (Бесплатное ПО); <a href="https://visualstudio.microsoft.com/ru/license-terms/mt171584/">https://visualstudio.microsoft.com/ru/license-terms/mt171584/</a></p> <p>7. Notepad ++ (Бесплатное ПО); <a href="https://www.gnu.org/licenses/old-licenses/gpl-2.0.html">https://www.gnu.org/licenses/old-licenses/gpl-2.0.html</a></p> <p>8. Zoom (Бесплатное ПО); <a href="https://explore.zoom.us/ru/terms/">https://explore.zoom.us/ru/terms/</a></p> <p>9. Anaconda3 2019 (Бесплатное ПО);</p> <p>10. <a href="https://www.anaconda.com/eula-anaconda-individual-edition">https://www.anaconda.com/eula-anaconda-individual-edition</a> Android studio (Бесплатное ПО); <a href="https://developer.android.com/studio/terms">https://developer.android.com/studio/terms</a></p> <p>11. Brackets (Бесплатное ПО); <a href="https://github.com/brackets-cont/brackets/blob/master/LICENSE">https://github.com/brackets-cont/brackets/blob/master/LICENSE</a></p> <p>12. CodeBlocks (Бесплатное ПО); <a href="https://www.codeblocks.org/license/Firebird">https://www.codeblocks.org/license/Firebird</a> (Бесплатное ПО); <a href="https://firebirdsql.org/en/licensing/">https://firebirdsql.org/en/licensing/</a></p> <p>13. KNIME analytics platform (Бесплатное ПО); <a href="https://www.knime.com/downloads/full-license">https://www.knime.com/downloads/full-license</a></p> <p>14. Loginom community РФ ПО (Бесплатное ПО); <a href="https://loginom.ru/legal">https://loginom.ru/legal</a></p> <p>15. Monogame SDK (Бесплатное ПО); <a href="https://github.com/MonoGame/MonoGame/blob/develop/LICENSE.txt">https://github.com/MonoGame/MonoGame/blob/develop/LICENSE.txt</a> Openproj (Бесплатное ПО); <a href="https://opensource.org/licenses/CPAL-1.0">https://opensource.org/licenses/CPAL-1.0</a></p> <p>16. tableau 2019 (Бесплатное ПО); <a href="https://www.tableau.com/legal">https://www.tableau.com/legal</a></p> <p>17. Visual studio community 2017 (Бесплатное ПО); <a href="https://www.google.com/url?sa=t&amp;rct=j&amp;q=&amp;esrc=s&amp;source=web&amp;cd=&amp;ved=2ahUKEwi-qZeRxv7zAhXhsYsKHZoRBAsQFnoECBgQAQ&amp;url=https%3A%2F%2Fvisualstudio.microsoft.com%2Fwp-content%2Fuploads%2F2017%2F01%2FVS2017_COMMUNITY_RC_RUS_Eula.1049-1.docx&amp;usg=AOvVaw0tLxIQ4E2McNypfRn9tTo">https://www.google.com/url?sa=t&amp;rct=j&amp;q=&amp;esrc=s&amp;source=web&amp;cd=&amp;ved=2ahUKEwi-qZeRxv7zAhXhsYsKHZoRBAsQFnoECBgQAQ&amp;url=https%3A%2F%2Fvisualstudio.microsoft.com%2Fwp-content%2Fuploads%2F2017%2F01%2FVS2017_COMMUNITY_RC_RUS_Eula.1049-1.docx&amp;usg=AOvVaw0tLxIQ4E2McNypfRn9tTo</a></p> <p>18. Visual studio community 2019 (Бесплатное ПО); <a href="https://visualstudio.microsoft.com/ru/license-terms/mlt110718/">https://visualstudio.microsoft.com/ru/license-terms/mlt110718/</a></p> <p>19. Консультант плюс</p>
3	Учебные аудитории для проведения занятий семинарского типа, курсового проектирования (выполнения)	<p>Преподавательский стол; столы обучающихся; стулья; классная доска; мультимедийный комплекс; ПК</p>	<p>1. 7-Zip (Бесплатное ПО);</p> <p>2. 10-Strike Network Inventory ПО РФ (ПО)</p> <p>3. Ductor Academic ПО РФ (Бесплатное ПО); <a href="https://basegroup.ru/deductor/manual/licence-deductor-academic">https://basegroup.ru/deductor/manual/licence-deductor-academic</a></p> <p>4. Eset Endpoint security (Платное ПО) <a href="https://help.eset.com/eula/GIMP">https://help.eset.com/eula/GIMP</a> (Бесплатное ПО); <a href="https://docs.gimp.org/2.10/ru/">https://docs.gimp.org/2.10/ru/</a></p> <p>5. microsoft office профессиональный плюс 2016 (ПО) <a href="https://www.microsoft.com/en-us/Useterms/Retail/Office/2016Professional/Useterms_Retail_Office_2016Professional_RUS.htm">https://www.microsoft.com/en-us/Useterms/Retail/Office/2016Professional/Useterms_Retail_Office_2016Professional_RUS.htm</a></p>

	курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также самостоятельной работы обучающихся	преподавателя; ПК обучающихся; наглядные пособия (плакаты) Место, оборудованное для лиц с ограниченными возможностями и. Лицензионное программное обеспечение, подключенное к Интернет	<a href="https://www.microsoft.com/ru-ru/office/2016professional/useterms_retail_office_2016professional_rus.htm">us/Useterms/Retail/Office/2016Professional/Useterms_Retail_Office_2016Professional_RUS.htm</a> 6. Microsoft power Bi (Бесплатное ПО); <a href="https://powerbi.microsoft.com/ru-ru/windows-license-terms/">https://powerbi.microsoft.com/ru-ru/windows-license-terms/</a> icrosoft Visual Studio (Бесплатное ПО); <a href="https://visualstudio.microsoft.com/ru/license-terms/mt171584/">https://visualstudio.microsoft.com/ru/license-terms/mt171584/</a> 7. Notepad ++ (Бесплатное ПО); <a href="https://www.gnu.org/licenses/old-licenses/gpl-2.0.html">https://www.gnu.org/licenses/old-licenses/gpl-2.0.html</a> 8. Zoom (Бесплатное ПО); <a href="https://explore.zoom.us/ru/terms/">https://explore.zoom.us/ru/terms/</a> 9. Anaconda3 2019 (Бесплатное ПО); <a href="https://www.anaconda.com/eula-anaconda-individual-edition">https://www.anaconda.com/eula-anaconda-individual-edition</a> Android studio (Бесплатное ПО); <a href="https://developer.android.com/studio/terms">https://developer.android.com/studio/terms</a> 11. Brackets (Бесплатное ПО); <a href="https://github.com/brackets-cont/brackets/blob/master/LICENSE">https://github.com/brackets-cont/brackets/blob/master/LICENSE</a> 12. CodeBlocks (Бесплатное ПО); <a href="https://www.codeblocks.org/license/">https://www.codeblocks.org/license/</a> Firebird (Бесплатное ПО); <a href="https://firebirdsql.org/en/licensing/">https://firebirdsql.org/en/licensing/</a> 13. KNIME analytics platform (Бесплатное ПО); <a href="https://www.knime.com/downloads/full-license">https://www.knime.com/downloads/full-license</a> 14. Loginom community РФ ПО (Бесплатное ПО); <a href="https://loginom.ru/legal">https://loginom.ru/legal</a> 15. Monogame SDK (Бесплатное ПО); <a href="https://github.com/MonoGame/MonoGame/blob/develop/LICENSE.txt">https://github.com/MonoGame/MonoGame/blob/develop/LICENSE.txt</a> Openproj (Бесплатное ПО); <a href="https://opensource.org/licenses/CPAL-1.0">https://opensource.org/licenses/CPAL-1.0</a> 16. tableau 2019 (Бесплатное ПО); <a href="https://www.tableau.com/legal">https://www.tableau.com/legal</a> 17. Visual studio community 2017 (Бесплатное ПО); <a href="https://www.google.com/url?sa=t&amp;rct=j&amp;q=&amp;esrc=s&amp;source=web&amp;cd=&amp;ved=2ahUKEwi-qZeRxxv7zAhXhsYsKHZoRBAsQFnoECBgQAQ&amp;url=https%3A%2F%2Fvisualstudio.microsoft.com%2Fwp-content%2Fuploads%2F2017%2F01%2FVS2017_COMMUNITY_RC_RUS_Eula.1049-1.docx&amp;usq=AOvVaw0tLx1QA4E2McNypfRn9tTo">https://www.google.com/url?sa=t&amp;rct=j&amp;q=&amp;esrc=s&amp;source=web&amp;cd=&amp;ved=2ahUKEwi-qZeRxxv7zAhXhsYsKHZoRBAsQFnoECBgQAQ&amp;url=https%3A%2F%2Fvisualstudio.microsoft.com%2Fwp-content%2Fuploads%2F2017%2F01%2FVS2017_COMMUNITY_RC_RUS_Eula.1049-1.docx&amp;usq=AOvVaw0tLx1QA4E2McNypfRn9tTo</a> 18. Visual studio community 2019 (Бесплатное ПО); <a href="https://visualstudio.microsoft.com/ru/license-terms/mlt110718/">https://visualstudio.microsoft.com/ru/license-terms/mlt110718/</a> 19. Консультант плюс
4	Библиотека с читальным залом	Стол обучающегося, стулья, ПК обучающегося, принтер Электронная библиотечная система и библиотечное абонентное обслуживание (учебная литература на бумажных носителях) Лицензионное программное обеспечение, подключенное к Интернет	1. 7-Zip (Бесплатное ПО); 2. microsoft office профессиональный плюс 2016 (ПО) <a href="https://www.microsoft.com/en-us/Useterms/Retail/Office/2016Professional/Useterms_Retail_Office_2016Professional_RUS.htm">https://www.microsoft.com/en-us/Useterms/Retail/Office/2016Professional/Useterms_Retail_Office_2016Professional_RUS.htm</a> 3. Microsoft power Bi (Бесплатное ПО); <a href="https://powerbi.microsoft.com/ru-ru/windows-license-terms/">https://powerbi.microsoft.com/ru-ru/windows-license-terms/</a> icrosoft Visual Studio (Бесплатное ПО); <a href="https://visualstudio.microsoft.com/ru/license-terms/mt171584/">https://visualstudio.microsoft.com/ru/license-terms/mt171584/</a> 4. Антиплагиат 5. Консультант плюс

## 9. Перечень информационных технологий

Образовательный процесс по дисциплине поддерживается средствами электронной информационно-образовательной среды Университета, которая обеспечивает:

- доступ к учебным планам, рабочим программам дисциплин (модулей), практик, к изданиям электронных библиотечных систем и электронным образовательным ресурсам, указанным в рабочей программе, через личный кабинет студента и преподавателя;

- фиксацию хода образовательного процесса, результатов промежуточной аттестации и результатов освоения основной образовательной программы;
- проведение всех видов занятий, процедур оценки результатов обучения, реализация которых предусмотрена с применением дистанционных образовательных технологий;
- формирование электронного портфолио обучающегося, в том числе сохранение работ обучающегося, рецензий и оценок на эти работы со стороны любых участников образовательного процесса;
- взаимодействие между участниками образовательного процесса, в том числе синхронное и (или) асинхронное взаимодействие посредством сети Интернет.

Каждый обучающийся обеспечен индивидуальным неограниченным доступом к электронно-библиотечной системе (ЭБС университета), содержащей издания учебной, учебно-методической и иной литературы по основным изучаемым дисциплинам и сформированной на основании прямых договоров с правообладателями.

Перечень программного обеспечения:

1. Adobe flash player 31;
2. Adobe reader 10;
3. Java 6.0;
4. K-Lite Codec Pack;
5. Win rar;
6. Microsoft Office 10;
7. Microsoft Visio 10;
8. Microsoft Visual studio;

Профессиональные базы данных и информационно-справочные системы:

1. Kaspersky Endpoint Security для бизнеса <http://inion.ru/resources/bazy-dannykh-inion-ran/> - библиографические базы данных ИНИОН РАН по социальным и гуманитарным наукам;
2. <http://inion.ru/resources/bazy-dannykh-inion-ran/> - библиографические базы данных ИНИОН РАН по социальным и гуманитарным наукам;
3. <http://www.scopus.com/> - реферативная база данных Scopus – международная универсальная реферативная база данных;
4. <http://elibrary.ru/defaultx.asp> - крупнейший российский информационный портал электронных журналов и баз данных по всем отраслям наук.
5. <https://www.cfin.ru/finanalysis/math/>, База данных «Финансовая математика – Библиотека управления» - Корпоративный менеджмент

## **10. Методические указания для обучающихся**

### **10.1. Преподавание дисциплины осуществляется в соответствии с Федеральным государственным образовательным стандартом высшего образования**

Основными формами получения и закрепления знаний по данной дисциплине являются занятия лекционного и семинарского типа, самостоятельная работа обучающегося, в том числе под руководством преподавателя, прохождение рубежного контроля.

Основной объем часов по изучению дисциплины согласно учебным планам приходится на самостоятельную работу обучающихся. Самостоятельная работа включает в себя изучение учебной, учебно-методической и специальной литературы, её конспектирование, подготовку к занятиям семинарского типа, текущему контролю и промежуточной аттестации (зачету или (и) экзамену).

Текущий контроль успеваемости по учебной дисциплине и промежуточная аттестация осуществляются в соответствии с Положением о текущем контроле успеваемости и промежуточной аттестации обучающихся по образовательным программам высшего образования: программам бакалавриата, программам специалитета, программам магистратуры.

Наличие в Университете электронной информационно-образовательной среды, а также электронных образовательных ресурсов позволяет осваивать курс инвалидам и лицам с ОВЗ.

## **10.2. Особенности освоения учебной дисциплины инвалидами и лицами с ограниченными возможностями здоровья.**

Особенности освоения учебной дисциплины инвалидами и лицами с ОВЗ определены в Положении об организации обучения студентов-инвалидов и студентов с ограниченными возможностями здоровья, утвержденным приказом ректора.

Обучение инвалидов и лиц с ОВЗ может осуществляться индивидуально, а также с применением электронного обучения, дистанционных образовательных технологий.

Выбор методов и средств обучения, образовательных технологий и учебно-методического обеспечения реализации образовательной программы осуществляется Университетом самостоятельно, исходя из необходимости достижения обучающимися планируемых результатов освоения образовательной программы, а также с учетом индивидуальных возможностей обучающихся из числа инвалидов и лиц с ОВЗ.

Форма проведения промежуточной аттестации для студентов-инвалидов и лиц с ОВЗ устанавливается с учетом индивидуальных психофизических особенностей (устно, письменно на бумаге, письменно на компьютере, в форме тестирования и т.п.). При необходимости инвалидам и лицам с ОВЗ предоставляется дополнительное время для подготовки ответа на зачете или экзамене.

В группах, в состав которых входят студенты с ОВЗ, с целью реализации индивидуального подхода, а также принципа индивидуализации и дифференциации, рекомендуется использовать технологию нелинейной конструкции учебных занятий, предусматривающую одновременное сочетание фронтальных, групповых и индивидуальных форм работы с различными категориями студентов, в т.ч. имеющих ОВЗ.

В случае наличия обучающихся с нарушением функций опорно-двигательного аппарата, зрения и слуха, они обеспечиваются необходимым оборудованием, имеющимся в Университете, а также предоставляемым в рамках Соглашения с РУМЦ РГСУ от 14 ноября 2019 года.

## **11. Методические рекомендации преподавателю по организации учебного процесса по дисциплине**

При изучении дисциплины рекомендуется использовать следующий набор средств и способов обучения:

- рекомендуемую основную и дополнительную литературу;
- задания для подготовки к занятиям семинарского типа (вопросы для обсуждения, кейс задания, расчетные задачи и др.);
- задания для текущего контроля успеваемости (задания для самостоятельной работы обучающихся, тестовые задания в рамках электронной системы тестирования);
- вопросы и задания для подготовки к промежуточной аттестации по итогам освоения дисциплины, позволяющие оценить знания, умения и уровень приобретенных компетенций.

При проведении занятий лекционного и семинарского типа, в том числе в форме вебинаров и on-line курсов необходимо строго придерживаться тематического плана дисциплины, приведенного в РПД. Необходимо уделить внимание рассмотрению вопросов и заданий, включенных в тестовые оценочные задания, при необходимости, решить аналогичные задачи с объяснением алгоритма решения.

Следует обратить внимание обучающихся на то, что для успешной подготовки к текущему контролю (выполнению ТОЗ) и промежуточной аттестации (зачету или экзамену) недостаточно прочитать рабочий учебник, размещенный в личном кабинете. Нужно изучить материалы основной и дополнительной литературы, список которой приведен в РПД, законодательные и нормативные акты, а также материалы, рекомендованные в разделе «Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых

для освоения дисциплины».

Текущий контроль успеваемости по учебной дисциплине и промежуточная аттестация осуществляются в соответствии с Положением о текущем контроле успеваемости и промежуточной аттестации обучающихся по образовательным программам высшего образования: программам бакалавриата, программам специалитета, программам магистратуры.

Программа разработана Замега Э.Н.

Рабочая программа дисциплины рассмотрена и принята на заседании кафедры от 27.01.2022 г., протокол №6.

**Лист регистрации изменений и дополнений  
в рабочую учебную программу**

Составителем внесены следующие изменения:

<b>Содержание изменений</b>	<b>Номер протокола и дата заседания кафедры, по утверждению изменений</b>
Рабочая программа дисциплины дополнена и утверждена	№ 1 от 28.08.2023
Рабочая программа дисциплины дополнена и утверждена	№ 1 от 29.08.2024

ОЦЕНОЧНЫЕ СРЕДСТВА ПО ДИСЦИПЛИНЫ  
**ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ**

Направление подготовки:

09.03.03 Прикладная информатика

Направленность (профиль)

Искусственный интеллект и анализ данных

Уровень высшего образования: бакалавриат

## СОДЕРЖАНИЕ

1. Перечень компетенций, формируемых в процессе освоения учебной дисциплины
2. Описание показателей и критериев оценивания компетенций, описание шкал оценивания
3. Типовые контрольные задания или иные материалы.
4. Методические материалы, определяющие процедуры оценивания результатов обучения по учебной дисциплине

## 1. Перечень компетенций, формируемых в процессе освоения учебной дисциплины

### 1.1. Планируемые результаты освоения образовательной программы

В результате изучения дисциплины обучающийся должен обладать следующими компетенциями (результатами освоения образовательной программы):

Коды компетенций	Содержание компетенций
ПК-7	Способность выполнять интеллектуальный анализ больших данных
ПК-8	Способность разрабатывать методы извлечения, анализа и обработки информации
ПК-9	Способность применять математические методы моделирования процессов обработки информации с использованием средств интеллектуального анализа данных

### 1.2. Взаимосвязь планируемых результатов обучения по дисциплине с планируемыми результатами освоения образовательной программы

Коды и формулировка компетенции	Индикаторы компетенции	Запланированные результаты обучения
ПК-7 Способность выполнять интеллектуальный анализ больших данных	ПК 7.1 Знать методы и инструментальные средства интеллектуального анализа больших данных ПК 7.2 Уметь выбирать средства представления результатов аналитики больших данных ПК 7.3 Владеть техническими, программными средствами для разработки алгоритмов и программ в области интеллектуального анализа данных	ПК 7.1.1 знать методы и инструментальные средства для машинного обучения и анализа больших данных - предметную область анализа больших данных - теоретические и прикладные основы анализа данных ПК 7.1.2 уметь использовать инструментальные средства машинного обучения и анализа больших данных - проводить сравнительный анализ методов и инструментальных средств анализа больших данных ПК 7.1.3 владеть инструментальными средствами машинного обучения и анализа больших данных ПК 7.2.1 знать средства визуализации результатов решения задач машинного обучения и анализа больших данных методы интерпретации и визуализации решения задач машинного обучения и анализа больших данных ПК 7.2.2 уметь использовать стандартные программные библиотеки для визуализации решений задач машинного обучения и анализа данных ПК 7.2.3 владеть стандартными программными библиотеками для решения задач машинного обучения и анализа данных ПК 7.3.1 знать программные платформы и библиотеки для решения задач машинного обучения и анализа больших данных стандарты проведения анализа данных ПК 7.3.2 уметь использовать программные платформы и библиотеки для решения задач машинного обучения и анализа больших данных ПК 7.3.3 владеть программными средствами для разработки алгоритмов машинного обучения, алгоритмами построения искусственных нейронных сетей
ПК-8 Способность разрабатывать методы извлечения,	ПК 8.1 Знать теоретические и прикладные основы анализа больших данных	ПК 8.1.1 знать особенности методов обработки и анализа данных - современный опыт использования анализа больших данных ПК 8.1.2 уметь решать прикладные задачи с помощью

анализа и обработки информации	<p>ПК 8.2 Уметь проводить анализ больших данных</p> <p>ПК 8.3 Владеть методами извлечения информации и знаний из гетерогенных, мультиструктурированных и неструктурированных источников</p>	<p>методов интеллектуального анализа данных - разрабатывать и оценивать модели больших данных</p> <p>ПК 8.1.3 владеть теоретическими знаниями в области прикладной статистики и машинного обучения</p> <p>ПК 8.2.1 знать факторы, влияющие на эффективность работы методов анализа больших данных</p> <p>ПК 8.2.2 уметь формализовать задачу интеллектуального анализа данных в части методов поиска ассоциативных правил, кластеризации и прогнозирования</p> <p>ПК 8.2.3 владеть проводить очистку данных методами поиска ассоциативных правил, кластеризации и прогнозирования</p> <p>ПК 8.3.1 знать стандартные методы извлечения информации из разнородных источников - типы и виды источников данных</p> <p>ПК 8.3.2 уметь использовать алгоритмы предобработки и постобработки из неструктурированных источников данных - использовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников, определять требования и взаимодействовать с внутренними и внешними поставщиками данных из разнородных источников</p> <p>ПК 8.3.3 владеть методами извлечения информации из неструктурированных источников данных</p>
ПК-9 Способность применять математические методы моделирования процессов обработки информации с использованием средств интеллектуального анализа данных	<p>ПК 9.1 Знать принципы решения задач машинного обучения и интеллектуального анализа данных</p> <p>ПК 9.2 Уметь создавать алгоритмические и математические модели прикладных задач интеллектуального анализа данных</p> <p>ПК 9.3 Владеть навыками построения описательных и прогнозных аналитических моделей с использованием современных инструментов интеллектуального анализа данных</p>	<p>ПК 9.1.1 знать основные принципы решения задач машинного обучения и анализа данных</p> <p>ПК 9.1.2 уметь проводить спецификацию задачи, реализовывать программы на алгоритмических языках высокого уровня, интерпретировать полученные результаты</p> <p>ПК 9.1.3 владеть методами и алгоритмами машинного обучения</p> <p>ПК 9.2.1 знать математические и алгоритмические модели интеллектуального анализа данных, методы оценки временных и стоимостных характеристик машинного обучения и анализа данных</p> <p>ПК 9.2.2 уметь создавать алгоритмические модели типовых прикладных задач обработки информации, реализовывать программы на алгоритмических языках высокого уровня</p> <p>ПК 9.2.3 владеть навыками использования прикладных библиотек для решения задач машинного обучения</p> <p>ПК 9.3.1 описательные и прогнозные аналитические модели для интеллектуального анализа данных - использовать современные программные инструменты интеллектуального анализа данных</p> <p>9.3.2 разрабатывать и оценивать модели больших данных</p> <p>9.3.3 навыками построения описательных и прогнозных аналитических моделей с использованием современных инструментов интеллектуального анализа данных, навыками выявления требований к результатам анализа больших данных в предметной области</p>

## 2. Описание показателей и критериев оценивания компетенций, описание шкал оценивания

2.1. Текущий контроль успеваемости по учебной дисциплине и промежуточная аттестация осуществляются в соответствии с Положением о текущем контроле успеваемости и промежуточной аттестации обучающихся по образовательным программам высшего образования: программам бакалавриата, программам специалитета, программам магистратуры.

2.2. В семестре степень освоения компетенций оценивается по 100-балльной шкале в форме тестирования два раза в семестр. В зачетно-экзаменационный период баллы приводят к среднеарифметическому значению и переводятся в традиционную четырехбалльную систему. Данная оценка может повлиять на итоговую.

Этапы формирования компетенций и критериев оценивания компетенций на различных этапах формирования:

Коды и формулировка компетенции	Индикаторы достижения компетенции	Уровень выраженности и критерии оценивания	Этапы формирования
ПК-7 Способность выполнять интеллектуальный анализ больших данных	ПК 7.1 Знать методы и инструментальные средства интеллектуального анализа больших данных	<b>Низкий (пороговый):</b> Демонстрирует поверхностные знания методов и инструментальных средств для машинного обучения и анализа больших данных - предметную область анализа больших данных - теоретические и прикладные основы анализа данных <b>Средний:</b> Уверенно показывает умение использовать инструментальные средства машинного обучения и анализа больших данных - проводить сравнительный анализ методов и инструментальных средств анализа больших данных показывает <b>Высокий:</b> На отличном уровне показывает владение инструментальными средствами машинного обучения и анализа больших данных	<b>Первый этап:</b> Проведение текущего контроля успеваемости по дисциплине <b>Второй этап:</b> Проведение промежуточной аттестации по учебной дисциплине.
	ПК 7.2 Уметь выбирать средства представления результатов аналитики больших данных	<b>Низкий (пороговый):</b> Демонстрирует поверхностные знания средств визуализации результатов решения задач машинного обучения и анализа больших данных методы интерпретации и визуализации решения задач машинного обучения и анализа больших данных <b>Средний:</b> Уверенно показывает умение использовать стандартные программные библиотеки для визуализации решений задач машинного обучения и анализа данных <b>Высокий:</b> На отличном уровне показывает владение стандартными программными библиотеками для решения задач машинного обучения и анализа данных	<b>Первый этап:</b> Проведение текущего контроля успеваемости по дисциплине <b>Второй этап:</b> Проведение промежуточной аттестации по учебной дисциплине.
	ПК 7.3 Владеть техническими, программными	<b>Низкий (пороговый):</b> Демонстрирует поверхностные знания программных платформ и	<b>Первый этап:</b> Проведение текущего

	<p>средствами для разработки алгоритмов и программ в области интеллектуального анализа данных</p>	<p>библиотеки для решения задач машинного обучения и анализа больших данных стандарты проведения анализа данных  <b>Средний:</b> Уверенно показывает умение использовать программные платформы и библиотеки для решения задач машинного обучения и анализа больших данных  <b>Высокий:</b> На отличном уровне показывает владение программными средствами для разработки алгоритмов машинного обучения, алгоритмами построения искусственных нейронных сетей</p>	<p>контроля успеваемости по дисциплине  <b>Второй этап:</b> Проведение промежуточной аттестации по учебной дисциплине.</p>
<p>ПК-8 Способность разрабатывать методы извлечения, анализа и обработки информации</p>	<p>ПК 8.1 Знать теоретические и прикладные основы анализа больших данных</p>	<p><b>Низкий (пороговый):</b> Демонстрирует поверхностные знания особенностей методов обработки и анализа данных - современный опыт использования анализа больших данных  <b>Средний: Уверенно</b> умеет решать прикладные задачи с помощью методов интеллектуального анализа данных - разрабатывать и оценивать модели больших данных.  <b>Высокий:</b> На отличном уровне владеет аналитическими навыками и теоретическими знаниями в области прикладной статистики и машинного обучения</p>	<p><b>Первый этап:</b> Проведение текущего контроля успеваемости по дисциплине  <b>Второй этап:</b> Проведение промежуточной аттестации по учебной дисциплине.</p>
	<p>ПК 8.2 Уметь проводить анализ больших данных</p>	<p><b>Низкий (пороговый):</b> Демонстрирует поверхностные знания факторов, влияющих на эффективность работы методов анализа больших данных  <b>Средний:</b> Уверенно формирует задачу интеллектуального анализа данных в части методов поиска ассоциативных правил, кластеризации и прогнозирования.  <b>Высокий:</b> Навыки очистки данных методами поиска ассоциативных правил, кластеризации и прогнозирования на отличном уровне</p>	<p><b>Первый этап:</b> Проведение текущего контроля успеваемости по дисциплине  <b>Второй этап:</b> Проведение промежуточной аттестации по учебной дисциплине.</p>
	<p>ПК 8.3 Владеть методами извлечения информации и знаний из гетерогенных, мультиструктурированных и неструктурированных источников</p>	<p><b>Низкий (пороговый):</b> Демонстрирует поверхностные знания стандартных методов извлечения информации из разнородных источников - типы и виды источников данных  <b>Средний:</b> Уверенно использует алгоритмы предобработки и постобработки из неструктурированных источников данных - использовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников,</p>	<p><b>Первый этап:</b> Проведение текущего контроля успеваемости по дисциплине  <b>Второй этап:</b> Проведение промежуточной аттестации по учебной дисциплине.</p>

		определять требования и взаимодействовать с внутренними и внешними поставщиками данных из разнородных источников <b>Высокий:</b> Отличные навыки владения методами извлечения информации из неструктурированных источников данных	
ПК-9 Способность применять математические методы моделирования процессов обработки информации с использованием средств интеллектуального анализа данных	ПК 9.1 Знать принципы решения задач машинного обучения и интеллектуального анализа данных	<b>Низкий (пороговый):</b> Демонстрирует поверхностные знания основных принципов решения задач машинного обучения и анализа данных <b>Средний:</b> Умеет проводить спецификацию задачи, реализовывать программы на алгоритмических языках высокого уровня, интерпретировать полученные результаты <b>Высокий:</b> На отличном уровне показывает владение методами и алгоритмами машинного обучения	<b>Первый этап:</b> Проведение текущего контроля успеваемости по дисциплине <b>Второй этап:</b> Проведение промежуточной аттестации по учебной дисциплине.
	ПК 9.2 Уметь создавать алгоритмические и математические модели прикладных задач интеллектуального анализа данных	<b>Низкий (пороговый):</b> Демонстрирует поверхностные знания математических и алгоритмических моделей интеллектуального анализа данных, методы оценки временных и стоимостных характеристик машинного обучения и анализа данных <b>Средний:</b> Умеет создавать алгоритмические модели типовых прикладных задач обработки информации, реализовывать программы на алгоритмических языках высокого уровня <b>Высокий:</b> На отлично уровне владеет навыками использования прикладных библиотек для решения задач машинного обучения	
	ПК 9.3 Владеть навыками построения описательных и прогнозных аналитических моделей с использованием современных инструментов интеллектуального анализа данных	<b>Низкий (пороговый):</b> Демонстрирует поверхностные знания в области описательных и прогнозных аналитических моделей для интеллектуального анализа данных <b>Средний:</b> Умеет разрабатывать и оценивать модели больших данных, использовать современные программные инструменты интеллектуального анализа данных <b>Высокий:</b> На отличном уровне владеет навыками построения описательных и прогнозных аналитических моделей с использованием современных инструментов интеллектуального анализа данных, навыками выявления требований к	<b>Первый этап:</b> Проведение текущего контроля успеваемости по дисциплине <b>Второй этап:</b> Проведение промежуточной аттестации по учебной дисциплине.

		результатам анализа больших данных в предметной области	
--	--	---	--

В ходе текущего контроля успеваемости при ответах на семинарских и практических занятиях, промежуточной аттестации в форме экзамена (зачет с оценкой) обучающиеся оцениваются по четырёхбалльной шкале оценивания: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно»:

шкала оценки	описание
оценка "отлично"	выставляется обучающимся, показавшим всестороннее, систематическое и глубокое знание учебно-программного материала, умение свободно выполнять задания, предусмотренные программой, усвоивших основную и дополнительную литературу, рекомендованную программой. Оценка "отлично" выставляется студентам, усвоившим взаимосвязь основных понятий дисциплины в их значении для приобретаемой профессии, проявившим творческие способности в понимании, изложении и использовании учебно-программного материала.
оценка "хорошо"	выставляется обучающимся, показавшим полное знание учебно-программного материала, успешно выполняющим предусмотренные в программе задания, усвоившим основную литературу, рекомендованную в программе. Оценка "хорошо" выставляется студентам, продемонстрировавшим систематический характер знаний по дисциплине и способным к их самостоятельному пополнению и обновлению в ходе дальнейшей учебной работы и профессиональной деятельности.
оценка "удовлетворительно"	выставляется обучающимся, показавшим знания основного учебно-программного материала в объеме, необходимом для дальнейшей учебы и предстоящей работы по специальности, справившимся с выполнением заданий, предусмотренных программой, ориентирующимся в основной литературе, рекомендованной программой. Оценка "удовлетворительно" выставляется студентам, допустившим погрешности в ответе на экзамене и при выполнении экзаменационных заданий, но обладающим необходимыми знаниями для их устранения под руководством преподавателя.
оценка "неудовлетворительно"	выставляется обучающимся, имеющим пробелы в знаниях основного учебно-программного материала, допустившим принципиальные ошибки в выполнении предусмотренных программой заданий. Оценка "неудовлетворительно" ставится студентам, которые не могут продолжить обучение или приступить к профессиональной деятельности по окончании вуза без дополнительных занятий по соответствующей дисциплине.

В ходе промежуточной аттестации в форме зачёта обучающиеся оцениваются «зачтено» или «не зачтено»:

шкала оценки	описание
оценка "зачтено"	выставляется обучающимся, показавшим знания основного учебно-программного материала, справившимся с выполнением заданий, предусмотренных программой, ориентирующимся в основной и дополнительной литературе, рекомендованной программой.

оценка "не зачтено"	выставляется обучающимся, имеющим пробелы в знаниях основного учебно-программного материала, допустившим принципиальные ошибки в выполнении предусмотренных программой заданий
---------------------	--

### 3. Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения по учебной дисциплине.

#### 3.1. Примерные варианты оценочных заданий (ТОЗ) для контрольного рубежа в рамках текущего контроля

Задание	Оцениваемая компетенция
Для чего можно использовать ООВ-предсказания? +: для оценки качества модели +: для реализации стекинга -: для регуляризации -: для кодирования целевого вектора +: для вычисления рейтинга (важности) признаков	ПК-7
Отметьте верные утверждения: +: Обрезку (post-pruning) используют крайне редко +: Деревья особенно эффективны в ансамбле +: Деревья – нестабильный (неустойчивый) алгоритм -: Деревья часто используют для экстраполяции	ПК-8
При минимизации функции $x^2$ методом градиентного спуска с темпом 1.0 и начальной точкой 1.0, какая будет оценка $\text{argmin}$ после 4й итерации? +: 1 -: 2 -: -4 -: 0	ПК-7
Что лучше использовать для определения монотонной зависимости между переменными? -: Корреляционный коэффициент Пирсона +: Коэффициент корреляции Спирмена -: оценку ММП (MLE)	ПК-7
Чем экстремальные леса (Extreme Random Trees) отличаются от случайных (Random Forest)? -: не используем критерии расщепления (типа gini и энтропийного) +: быстрее построение ансамбля -: нужен градиент функции ошибки +: качество, как правило, чуть хуже	ПК-8
Пусть случайная величина равна сумме двух равномерно распределённых величин на отрезке $[0,1]$ . Как выглядит её плотность распределения? +: колокольчик -: треугольник -: трапеция -: прямоугольник	ПК-9
В какой модели разнообразие базовых алгоритмов повышается за счёт варьирования обучающей выборки? +: метод случайных подпространств (Random Subspaces)	ПК-8

+: бэггинг (Bagging) -: нейросети +: случайные леса (Random Forests)	
В каких моделях увеличение числа базовых алгоритмов не приводит к переобучению? -: стекинг -: бустинг +: случайные леса	ПК-9
Какие из перечисленных ниже моделей являются последовательными ансамблями (Sequential ensembles)? +: бэггинг (Bagging) +: Adaboost -: градиентный бустинг -: нейронные сети -: случайные леса (RF)	ПК-9
Выберите верные фразы: +: Для селекции признаков можно использовать LASSO +: Устойчивая регрессия (Robust Regression) хороша в задаче с выбросами -: В логистической регрессии минимизируют среднее квадратичное отклонение ответов модели от истинных меток -: решение линейной регрессии робастно (устойчиво к выбросам)	ПК-8
Задачу регрессии можно решать с помощью: +: kNN +: метода Надарая-Ватсона -: метода логистической регрессии	ПК-9
Пусть случайные величины одинаково распределены (среднее равно 1, дисперсия – 2), корреляция между любой парой величин равна 0.1. К чему стремится среднее арифметическое этих величин при увеличении числа наблюдений (т.е. увеличении числа этих величин)? +: 0,2 +: 0.2	ПК-9
Что происходит при увеличении глубины деревьев (считаем, что в ансамблях достаточное число деревьев)? +: как правило, увеличивается качество случайного леса на тесте +: как правило, увеличивается качество случайного леса на обучении -: как правило, увеличивается качество бустинга над деревьями на тесте +: как правило, увеличивается качество бустинга на обучении	ПК-7
Какие расстояния численно наибольшие для пары точек (1,1) и (2,2) -: Евклидово -: Чебышева +: Манхэттенское -: нет правильного ответа	ПК-8

### 3.2. Вопросы для подготовки к промежуточной аттестации (к зачету)

1. Проблема больших объемов («Data explosion»)
2. Интеллектуальный анализ данных (Data Mining): основные определения
3. Актуальность и необходимость интеллектуального анализа данных

4. Анализ предметной области
5. Проверка моделей и представление результатов анализа
6. Применение построенных моделей
7. Основные типы исходных данных
8. Основные типы данных для анализа
9. Ассоциативный анализ
10. Метод «ветвей и границ»
11. Правила с семантикой
12. Тип моделей
13. Задачи анализа
14. Алгоритмы argioi и fp-tree
15. Объективные меры интересности
16. Использование ограничений
17. Проблема итеративного анализа больших объемов данных
18. Статистическое обучение без учителя
19. Задача «самоорганизации»
20. Задача поиска неизвестных зависимостей без эксперта «Истинный data mining»
21. Поиск скрытых (латентных) признаков
22. Поиск скрытых структур (групп или зависимостей)
23. Тематическое моделирование текстовых данных. Скрытые признаки документа
24. Примеры использования метода главных компонент в SAS EM
25. Кластеризация переменных
26. Алгоритм группировки переменных
27. Пример использования кластеризации переменных в SAS EM
28. Самоорганизующиеся отображения
29. Кластеризация: иерархическая
30. Кластеризация: метрическая
31. Кластеризация: вероятностная
32. Этапы кластерного анализа
33. Качество кластеризации
34. Требования к методу кластеризации
35. Определение кластера
36. Подготовка данных для кластеризации
37. Отбор наблюдений
38. Основные типы алгоритмов кластеризации
39. Отбор и трансформация переменных
40. Задача "Обучения с учителем"
41. Обучающая выборка или тренировочный набор
42. Этап обучения и этап прогнозирования
43. Типы задач прогнозирования: бинарная классификация (разделение)
44. Типы задач прогнозирования: регрессия
45. Типы задач прогнозирования: классификация
46. Типы задач прогнозирования: много-темная классификация (mutli-label)
47. Метод k-ближайших соседей
48. Метод KNN
49. Метод K ближайших соседей с адаптивным расстоянием

50. Свойства методов KNN
51. Задача линейной регрессии
52. Уравнение линейной регрессии
53. Линеаризируемые регрессии: степенная
54. Линеаризируемые регрессии: экспоненциальная
55. Линеаризируемые регрессии: гиперболическая
56. Определение наличия связи между переменными и характера этой связи, предсказание значения зависимой переменной с помощью независимой(-ых)
57. Определение вклада отдельных независимых переменных в вариацию зависимой переменной
58. Линейная модель с нелинейными эффектами
59. Метод наименьших квадратов и проблема мультиколлинеарности
60. Проблема недообучения и переобучения
61. Выбор модели и регуляризация на примере линейных моделей регрессии
62. Полиномиальная регрессия
63. Локальная регрессия
64. Регрессия Ridge
65. Регрессия Lasso
66. Обобщенные аддитивные модели
67. Сглаживание сплайнами
68. Типовые архитектуры RBF
69. Типовые архитектуры MLP
70. Ранняя остановка обучения
71. Алгоритмы оптимизации для обучения нейронных сетей
72. Обучение нейронных сетей (с учителем)
73. Правила обучения нейросетей
74. Линейный перцептрон
75. Функции активации
76. Искусственный нейрон
77. Перцептрон с прямыми соединениями
78. Многослойный перцептрон
79. Обучение нейронных сетей: критерии сходимости
80. Обучение нейронных сетей: оценки отклонения
81. Обучение нейронных сетей: комбинации функций активации и распределения ошибок
82. Метод множителей Лагранжа
83. RBF-нормализованная нейронная сеть
84. Обратное распространение ошибки (градиентный метод)
85. Метод доверительных областей (trusted regions)
86. Комбинированный (градиент+ньютон) Double-Dogleg
87. Задача классификации
88. Методы Байеса
89. Методы, основанные на деревьях решений
90. Применение деревьев решений для задач классификации
91. Деревья решений в задачах классификации и регрессии
92. Процесс построения деревьев решений
93. «Отсечение» ветвей дерева
94. Проверка атрибутов дерева

95. Жадный подход - рекурсивное разделение
96. Множественные разбиения
97. Поиск разбиения по переменным
98. Индекс Джини
99. Точность дерева

#### **4. Методические материалы, определяющие процедуры оценивания результатов обучения по учебной дисциплине.**

Процедура оценивания результатов обучения по учебной дисциплине осуществляется на основе Положением о текущем контроле и промежуточной аттестации обучающихся, утвержденными приказом ректора.

##### *4.1 Первый этап: Проведение текущего контроля успеваемости по дисциплине*

Проведение текущего контроля успеваемости по дисциплине осуществляется в ходе контактной работы с преподавателем в рамках аудиторных занятий и в ходе самостоятельной работы студента.

Текущий контроль в ходе контактной работы осуществляется по следующим видам:

**1) Вид контроля:** проверка сформированности компетенций в ходе самостоятельной работы обучающихся; текущий опрос, проводимый во время аудиторных (семинарских/практических/лабораторных) занятий; оценивание подготовленных докладов, сообщений, презентаций, домашних заданий.

**Порядок проведения:** в ходе подготовки к занятиям оценивается выполнение задания, рекомендованного к самостоятельной работе обучающихся, путем выборочной проверки.

Фиксируются результаты работы студентов в ходе проведения семинарских и практических занятий (активность, полнота ответов, способность поддерживать дискуссию, профессиональный язык и др.).

В ходе отдельных занятий обеспечивается проведение письменных опросов по тематике прошедших занятий. В ходе выполнения заданий обучающийся должен в меру имеющихся знаний, умений, навыков, сформированности компетенции дать развернутые ответы на поставленные в задании открытые вопросы и ответить на вопросы закрытого типа в установленное преподавателем время. Продолжительность проведения процедуры определяется преподавателем самостоятельно, исходя из сложности индивидуальных заданий, количества вопросов, объема оцениваемого учебного материала.

Задания по подготовке докладов, сообщений, презентаций, домашних заданий выдаются заранее при подготовке к семинарским и практическим занятиям; подготовленные работы оцениваются с фиксацией в журнале учета посещаемости и успеваемости обучающихся.

**2) Вид контроля:** Контроль с использованием тестовых оценочных заданий по итогам освоения тем дисциплины (текущий (рубежный) контроль).

**Порядок проведения:** До начала проведения процедуры преподавателем подготавливаются необходимые оценочные материалы для оценки знаний, умений, навыков.

Оценка знаний, умений и навыков, характеризующих сформированность компетенций, осуществляется с помощью тестовых оценочных заданий (ТОЗ), сформированные в соответствии с Требованиями по подготовке тестовых оценочных заданий

Внеаудиторная контактная работа преподавателя с обучающимся осуществляется в ходе выполнения рейтинговой работы и контроля со стороны преподавателя за самостоятельной работой студента. Текущий контроль в ходе самостоятельной работы осуществляется в следующем виде:

3) Вид контроля: Подготовка курсовой работы (при наличии в учебном плане).

**Технология проведения:** За каждым обучающимся, принимающим участие в процедуре преподавателем закрепляется тема курсовой работы. После получения задания и в процессе его подготовки обучающийся должен в меру имеющихся знаний, умений, навыков, сформированности компетенции дать развернутое раскрытие темы, выполнить расчетное или иное задание.

*4.2 Второй этап: Проведение промежуточной аттестации по учебной дисциплине.*

В соответствии с учебным планом по учебной дисциплине предусмотрена подготовка и сдача зачета.

Порядок проведения промежуточной аттестации регламентируется Положением о текущем контроле и промежуточной аттестации, утвержденным приказом ректора Университета.