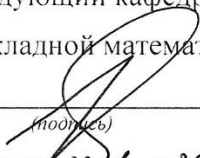


Государственное образовательное учреждение
«Приднестровский государственный университет им. Т.Г. Шевченко»

Физико-технический институт
Физико-математический факультет
Кафедра высшей и прикладной математики и информатики

УТВЕРЖДАЮ

Заведующий кафедрой высшей и
прикладной математики и информатики



(подпись) Коровай А.В.
(Ф.И.О.)

протокол № 1 «30» 08 2024 г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
по дисциплине
Б1.В.09. «Интеллектуальный анализ данных»

Направление

01.03.02 «Прикладная математика и информатика»

Профиль

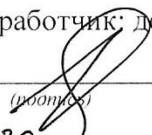
Системное программирование и компьютерные технологии

Квалификация

Бакалавр

ГОД НАБОРА 2022

Разработчик: доцент



(подпись) Коровай А.В.
(Ф.И.О.)

«30» 08 2024 г.

г. Тирасполь 2024 г.

Паспорт фонда оценочных средств по учебной дисциплине
«Интеллектуальный анализ данных»

1. В результате изучения дисциплины «Интеллектуальный анализ данных» у обучающихся должны быть сформированы следующие компетенции:

Категория (группа) компетенций	Код и наименование	Код и наименование индикатора достижения компетенции
Универсальные компетенции и индикаторы их достижения		
Учебным планом не предусмотрены		
Общепрофессиональные компетенции и индикаторы их достижения		
Учебным планом не предусмотрены		
Обязательные профессиональные компетенции и индикаторы их достижения		
	ПК-1 Способен демонстрировать общенаучные базовые знания естественных наук, математики и информатики, понимание основных фактов, концепций, принципов теорий, связанных с прикладной математикой и информатикой	ИД-1 _{ПК-1} Обладает базовыми знаниями, полученными в области математических и (или) естественных наук, программирования и информационных технологий.
		ИД-2 _{ПК-1} Умеет находить, формулировать и решать стандартные задачи в собственной научно-исследовательской деятельности в области математических и (или) естественных наук, программирования и информационных технологий.
		ИД-3 _{ПК-1} Имеет практический опыт научно-исследовательской деятельности в области математических и (или) естественных наук, программирования и информационных технологий.
	ПК-2 Способен понимать и применять в исследовательской и прикладной деятельности современный математический аппарат.	ИД-1 _{ПК-2} Знает современный математический аппарат.
		ИД-2 _{ПК-2} Умеет применять методы, алгоритмы и приёмы современного математического аппарата.
		ИД-3 _{ПК-2} Владеет практическими навыками применения современного математического аппарата в исследовательской и прикладной деятельности.
	ПК-5 Способен разрабатывать и применять алгоритмические и программные решения в области системного и прикладного программного обеспечения.	ИД-1 _{ПК-5} Знает разработку архитектуры, алгоритмических и программных решений системного и прикладного программного обеспечения.
		ИД-2 _{ПК-5} Умеет использовать языки программирования, алгоритмы, библиотеки и пакеты программ, продукты системного и прикладного программного обеспечения.
		ИД-3 _{ПК-5} Владеет навыками решения практических задач с применением языков программирования, алгоритмов, библиотек и пакетов программ.
Рекомендуемые профессиональные компетенции и индикаторы их достижения		
Учебным планом не предусмотрены		

2. Программа оценивания контролируемой компетенции:

Текущая аттестация	Контролируемые модули, разделы (темы) дисциплины и их наименование	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
1	Раздел 1. Введение в интеллектуальный анализ данных. Инструментальные средства интеллектуального анализа данных.	ПК-1, ПК-2, ПК-5	Комплект тестовых заданий

	Раздел 2. Предварительная обработка данных. Раздел 3. Техники Data Mining: классификация, кластеризация, прогнозирование и визуализация. Раздел 4. Нейросетевые алгоритмы.		
Промежуточная аттестация	Код контролируемой компетенции (или ее части)	Наименование оценочного средства	
1	ПК-1, ПК-2, ПК-5	Список вопросов к экзамену	

Комплект тестовых заданий
по дисциплине «Интеллектуальный анализ данных»

1. Что такое Big Data?
 - 1) данные, которые могут быть обработаны с помощью традиционных методов анализа
 - 2) большие объемы данных, которые невозможно эффективно обработать с помощью традиционных методов
 - 3) данные, которые имеют только текстовый формат
 - 4) малые объемы данных, содержащие высокую степень подробности
2. Что такое «обучение с учителем» (supervised learning)?
 - 1) процесс, при котором модель обучается на размеченных данных, с известными правильными ответами
 - 2) процесс, при котором модель самостоятельно находит структуры в данных
 - 3) процесс обучения без данных
 - 4) процесс, при котором модель обучается без анализа ошибок
3. Что из нижеперечисленного относится к задаче обучения с подкреплением?
 - 1) прогнозирование цены на жилье
 - 2) классификация изображений по категориям
 - 3) обучение агента, который взаимодействует с окружающей средой для максимизации награды
 - 4) кластеризация документов по темам
4. Какова цель применения метода кросс-валидации в машинном обучении?
 - 1) оценить эффективность модели на разных подмножествах данных, чтобы избежать переобучения
 - 2) увеличить размер данных
 - 3) уменьшить размер признаков для ускорения обучения
 - 4) использовать несколько моделей для прогнозирования
5. Что такое переобучение (overfitting) в контексте машинного обучения?
 - 1) когда модель слишком проста и не может точно предсказать результаты
 - 2) когда модель слишком сложна и слишком хорошо подходит к данным для обучения, теряя способность обобщать на новые данные
 - 3) когда модель не использует все доступные данные для обучения
 - 4) когда данные для обучения слишком шумные
6. Что такое «гиперпараметры» в машинном обучении и анализе данных?
 - 1) параметры, которые обучаются модели в процессе обучения
 - 2) параметры, определяющие структуру модели, устанавливаемые до начала обучения
 - 3) параметры, которые определяют способ обработки данных
 - 4) параметры, связанные с анализом ошибок модели
7. Как называется задача, в которой мы ищем группу схожих объектов в наборе данных?
 - 1) классификация
 - 2) кластеризация
 - 3) регрессия
 - 4) прогнозирование
8. Что из перечисленного является примером задачи классификации?
 - 1) прогнозирование температуры на основе исторических данных
 - 2) определение, принадлежит ли изображение кошке или собаке
 - 3) прогнозирование объема продаж на следующей неделе
 - 4) оценка качества работы сотрудника по числу выполненных задач
9. Какую задачу решает этап предварительной обработки данных?
 - 1) выбор алгоритма машинного обучения
 - 2) приведение данных в форму, удобную для анализа и обучения модели
 - 3) оценка точности модели
 - 4) определение структуры данных
10. Какие из следующих операций относятся к этапу очистки данных?
 - 1) преобразование категориальных признаков в числовые
 - 2) удаление пропущенных значений или замена их на среднее/медиану
 - 3) масштабирование признаков для улучшения сходимости алгоритмов
 - 4) создание новых признаков путем объединения существующих

11. Как называется процесс масштабирования признаков в диапазон от 0 до 1?
 - 1) стандартизация (Standardization)
 - 2) нормализация (Normalization)
 - 3) логарифмирование (Log Transformation)
 - 4) кодирование категориальных признаков
12. Какое преобразование применяется к данным при стандартизации?
 - 1) вычитание среднего и деление на стандартное отклонение
 - 2) преобразование в логарифмическую шкалу
 - 3) преобразование категориальных данных в числовые
 - 4) приведение всех значений к диапазону от 0 до 1
13. Какой метод используется для кодирования категориальных переменных в числовые значения?
 - 1) метод главных компонент (PCA)
 - 2) One-Hot кодирование
 - 3) логистическая регрессия
 - 4) кластеризация
14. Что такое выбросы в данных?
 - 1) пропущенные значения, которые не были заменены
 - 2) данные, которые находятся в пределах обычных значений
 - 3) данные, которые сильно отклоняются от остальных значений в наборе
 - 4) данные, которые повторяются несколько раз в наборе
15. Как можно выявить выбросы в данных?
 - 1) использовать метод скользящего среднего
 - 2) построить гистограмму или boxplot (диаграмму размаха)
 - 3) применить алгоритм линейной регрессии
 - 4) выполнить кластеризацию с методом k-средних
16. Какой из следующих методов относится к обучению с учителем?
 - 1) кластеризация
 - 2) регрессия
 - 3) алгоритм k-средних
 - 4) метод главных компонент
17. Какова цель задачи регрессии?
 - 1) классифицировать данные в предопределенные категории
 - 2) определить вероятностное распределение данных
 - 3) предсказать числовое значение на основе входных данных
 - 4) разделить данные на группы
18. Что такое функция потерь в задачах машинного обучения и анализа данных?
 - 1) метод оценки точности модели
 - 2) мера ошибки модели, которую необходимо минимизировать
 - 3) процесс очистки данных от шума
 - 4) алгоритм, используемый для оптимизации модели
19. Что такое «градиентный спуск» в машинном обучении?
 - 1) метод поиска локального минимума функции потерь для оптимизации параметров модели
 - 2) метод для оценки важности признаков
 - 3) метод для деления данных на тренировочную и тестовую выборки
 - 4) метод для обработки данных с пропущенными значениями
20. Что такое регуляризация в контексте линейной регрессии?
 - 1) уменьшение объема данных для ускорения обработки
 - 2) добавление штрафа за большие значения коэффициентов модели для предотвращения переобучения
 - 3) использование алгоритма кросс-валидации
 - 4) понижение числа параметров модели
21. В чем заключается основное отличие регрессии от классификации?
 - 1) регрессия используется для предсказания категориальных значений, а классификация – для предсказания числовых значений
 - 2) регрессия предсказывает числовые значения, а классификация – категориальные
 - 3) регрессия и классификация – это синонимы
 - 4) регрессия требует обучения без учителя, а классификация – с учителем

22. Какая из следующих метрик используется для оценки качества классификации?
- 1) медиана
 - 2) F1-мера
 - 3) среднеквадратическая ошибка
 - 4) дисперсия
23. Что такое логистическая регрессия?
- 1) алгоритм для линейной регрессии, который предсказывает непрерывные значения
 - 2) алгоритм классификации, который предсказывает вероятность принадлежности к классу
 - 3) алгоритм кластеризации для разделения объектов на группы
 - 4) алгоритм для уменьшения размерности данных
24. В чем состоит основное отличие логистической регрессии от линейной регрессии?
- 1) логистическая регрессия предсказывает дискретные значения, а линейная – непрерывные
 - 2) логистическая регрессия использует экспоненциальную функцию для преобразования выходных данных, а линейная – только линейную
 - 3) линейная регрессия всегда приводит к бинарной классификации, а логистическая – к многоклассовой
 - 4) логистическая регрессия использует только категориальные признаки, а линейная – только числовые.
25. Какую функцию активации использует логистическая регрессия?
- 1) сигмоидную функцию
 - 2) линейную функцию
 - 3) функцию ReLU
 - 4) тангенс гиперболический
26. Какое значение принимает сигмоидная функция в логистической регрессии?
- 1) значения от $-\infty$ до $+\infty$
 - 2) значения от 0 до 1
 - 3) значения от -1 до 1
 - 4) значения от 0 до 10
27. Как вычисляется функция потерь в логистической регрессии?
- 1) через среднеквадратичную ошибку (MSE)
 - 2) через кросс-энтропию
 - 3) через сумму квадратов ошибок
 - 4) через сумму абсолютных ошибок
28. Как можно интерпретировать коэффициенты в логистической регрессии?
- 1) они показывают, насколько сильно изменение каждого признака влияет на вероятность того, что объект будет отнесен к положительному классу
 - 2) они показывают, насколько точно предсказывает модель значения в задачах регрессии
 - 3) они показывают точность классификации модели
 - 4) они показывают количество ошибок в классификации
29. Какова роль порога вероятности в логистической регрессии?
- 1) порог помогает определять, когда классифицировать объект как положительный класс, а когда как отрицательный
 - 2) порог используется для регулировки скорости обучения
 - 3) порог помогает определять, какие признаки включать в модель
 - 4) порог не используется в логистической регрессии
30. Какой алгоритм машинного обучения использует концепцию гиперплоскости для разделения данных?
- 1) дерево решений
 - 2) логистическая регрессия
 - 3) метод опорных векторов (SVM)
 - 4) метод k-ближайших соседей
31. Какое из утверждений о деревьях решений является правильным?
- 1) деревья решений всегда не склонны к переобучению
 - 2) деревья решений могут быть использованы только для задач классификации
 - 3) для построения дерева решений можно использовать как классификацию, так и регрессию
 - 4) деревья решений не могут работать с категориальными данными
32. Какой из критериев наиболее часто используется для оценки разделения на каждом шаге при построении дерева решений для задач классификации?
- 1) сумма квадратов ошибок

- 2) кросс-валидация
 - 3) индекс Джини или прирост информации (Information Gain)
 - 4) логистическая функция потерь
33. Какой из методов машинного обучения часто используется для классификации текста?
- 1) метод опорных векторов (SVM)
 - 2) логистическая регрессия
 - 3) алгоритм наивного Байеса
 - 4) алгоритм k-средних
34. Какую задачу решает алгоритм наивного Байеса?
- 1) классификация данных с использованием вероятностной модели
 - 2) прогнозирование временных рядов с учетом сезонности
 - 3) выявление закономерностей в данных
 - 4) разбиение данных на группы
35. Какой метод обучения используется для обучения с учителем, когда данные содержат как числовые, так и категориальные признаки?
- 1) метод ближайших соседей (KNN)
 - 2) логистическая регрессия
 - 3) случайный лес (Random Forest)
 - 4) метод главных компонент (PCA)
36. Какой из методов является примером ансамблевого обучения?
- 1) метод k-средних
 - 2) случайный лес (Random Forest)
 - 3) логистическая регрессия
 - 4) байесовский классификатор
37. Какой из методов является примером обучения без учителя?
- 1) логистическая регрессия
 - 2) метод опорных векторов
 - 3) кластеризация методом k-средних
 - 4) нейронные сети
38. Что такое «кластеризация»?
- 1) разделение данных на заранее определенные группы с одинаковыми признаками
 - 2) процесс обучения модели на размеченных данных
 - 3) разбиение данных на несколько групп, так чтобы объекты в каждой группе были похожи друг на друга
 - 4) процесс, при котором модель оценивает вероятность принадлежности объекта к классу
39. В чем состоит основное отличие иерархической кластеризации от неиерархической?
- 1) иерархическая кластеризация не требует определения количества кластеров заранее
 - 2) неиерархическая кластеризация основывается на деревьях решений
 - 3) иерархическая кластеризация имеет фиксированное количество кластеров
 - 4) иерархическая кластеризация делит данные на заранее заданное количество кластеров
40. Какие из методов являются примером неиерархической кластеризации?
- 1) алгоритм DBSCAN
 - 2) агломеративная кластеризация
 - 3) метод k-средних
 - 4) метод ближайших соседей
41. В чем заключается суть алгоритма DBSCAN для кластеризации?
- 1) разделение данных на k заранее определенных кластеров
 - 2) кластеризация на основе плотности точек, где выбросы отбрасываются как «шум»
 - 3) иерархическое разделение данных на множество уровней
 - 4) использование метрики «ближайшего соседа» для формирования кластеров
42. К целям анализа временных рядов относятся
- 1) прогнозирование будущих значений
 - 2) выявление сезонности и трендов
 - 3) классификация объектов в категорию на основе признаков
 - 4) выявление аномалий и выбросов
 - 5) сегментация данных на основе расстояния между точками

43. К основным компонентам временного ряда относятся
- 1) циклическая
 - 2) временная
 - 3) трендовая
 - 4) сезонная
 - 5) пространственная
44. К методам анализа временных рядов относятся
- 1) метод экспоненциального сглаживания
 - 2) метод k-средних
 - 3) метод скользящего среднего
 - 4) метод дискриминантного анализа
 - 5) метод полной связи
45. Что такое «скользящее среднее» в анализе временных рядов?
- 1) метод предсказания будущих значений на основе автокорреляции
 - 2) статистический метод, сглаживающий данные путем усреднения значений по определенному интервалу времени
 - 3) модель для выявления сезонных колебаний во временных рядах
 - 4) алгоритм для обработки выбросов в данных
46. Что такое автокорреляция во временном ряду?
- 1) связь между текущим значением ряда и его предыдущими значениями
 - 2) мера тренда в данных
 - 3) измерение сезонных колебаний данных
 - 4) мера точности прогноза
47. Что такое функция активации в нейронных сетях?
- 1) функция, которая генерирует данные для обучения модели
 - 2) функция, которая помогает сети принимать решения о выходных значениях
 - 3) функция, которая используется для оптимизации алгоритма обучения
 - 4) функция, которая контролирует скорость обучения сети
48. Как называется процесс, в котором нейронные сети обучаются на основе ошибок, которые они делают в ходе предсказания?
- 1) обратное распространение ошибки (Backpropagation)
 - 2) градиентный спуск (Gradient Descent)
 - 3) линейная регрессия
 - 4) метод кросс-валидации
49. Что такое скрытые слои в нейронных сетях?
- 1) слои, которые соединяют входные и выходные нейроны напрямую
 - 2) слои, которые преобразуют данные, полученные на входе, и передают их на выходные слои
 - 3) слои, которые не используются для обработки данных, но фиксируют ошибки
 - 4) слои, которые только вычисляют потери
50. Что является основной особенностью сверточных нейронных сетей (CNN)?
- 1) использование рекуррентных связей для обработки последовательностей данных
 - 2) применение свертки для выделения признаков из данных, что позволяет эффективно обрабатывать изображения и видео
 - 3) применение метода градиентного спуска для обучения с учителем
 - 4) применение слоев с прямым распространением, без использования активационных функций

Критерии оценки:

За каждое правильно выполненное задание назначается 1 балл. В случае задания с выбором нескольких правильных ответов 1 балл дается за полностью выполненное задание, 0 баллов – за хотя бы один неверный ответ. Затем первичные баллы пропорционально переводятся в тестовые баллы согласно технологической карте дисциплины.

Вопросы к экзамену
по дисциплине «Интеллектуальный анализ данных»

1. Введение в интеллектуальный анализ данных. Основные понятия и определения. Big Data: основные понятия, свойства.
2. Сферы применения интеллектуального анализа данных. Примеры.
3. Этапы интеллектуального анализа данных. Общие типы закономерностей при анализе данных.
4. Основные задачи анализа данных.
5. Классификация методов интеллектуального анализа данных.
6. Понятие машинного обучения. Основные виды машинного обучения.
7. Разделение данных на обучающую и тестовую выборки. Валидационная выборка. Кросс-валидация.
8. Жизненный цикл модели машинного обучения.
9. Основные инструменты Data Mining. Библиотеки Python для анализа данных.
10. Цели разведочного анализа данных (EDA). Инструменты и методы разведочного анализа данных.
11. Предварительная обработка данных. Этапы очистки данных в машинном обучении.
12. Предварительная обработка данных. Нормализация и стандартизация данных.
13. Предварительная обработка данных. Методы кодирования категориальных признаков.
14. Постановка задачи линейной регрессии. Модель одномерной и многомерной линейной регрессии.
15. Метод наименьших квадратов и функция потерь линейной регрессии.
16. Метод градиентного спуска для обучения модели линейной регрессии.
17. Метрики качества линейных регрессионных моделей. Преимущества и недостатки линейной регрессии.
18. Модель полиномиальной регрессии. Переобучение и недообучение.
19. Регуляризация линейной регрессии (Ridge, Lasso, ElasticNet).
20. Задача классификации. Этапы процесса классификации. Основные проблемы классификации.
21. Бинарная классификация и метрики качества ее моделей.
22. Многоклассовая классификация. Методы one-vs-all и one-vs-one.
23. Метрики качества при многоклассовой классификации.
24. Метод k ближайших соседей: общее описание метода, принцип работы. Преимущества и недостатки метода.
25. Бинарная логистическая регрессия. Функции потерь для логистической регрессии. Преимущества и недостатки метода.
26. Наивный байесовский классификатор: общее описание метода, принцип работы. Преимущества и недостатки метода.
27. Деревья решений: общее описание метода, основной принцип работы. Преимущества и недостатки метода.
28. Бутстрап. Бэггинг. Случайный лес. Преимущества и недостатки случайного леса.
29. Бустинг. Алгоритм AdaBoost.
30. Постановка задачи кластерного анализа. Области применения кластеризации. Цели кластерного анализа.
31. Классификация методов кластеризации. Метрики расстояния между объектами. Метрики расстояния между кластерами.
32. Агломеративная иерархическая кластеризация. Дендрограмма.
33. Алгоритм кластеризации k-средних: общее описание метода, принцип работы. Преимущества и недостатки метода.
34. Алгоритм кластеризации DBSCAN: общее описание метода, принцип работы. Преимущества и недостатки метода.
35. Оценка качества кластеризации.
36. Понятие временного ряда. Примеры. Цели, этапы и методы анализа временных рядов.
37. Детерминированная и случайная составляющая временного ряда. Тренд, сезонная и циклическая компонента. Модели тренда.
38. Стационарность. Тест Дики-Фуллера. Автокорреляция.
39. Метод скользящего среднего.
40. Метод экспоненциального сглаживания.
41. Модель ARMA.
42. Искусственные нейронные сети: основные понятия, сферы применения. Классификация нейронных сетей. Архитектуры нейронных сетей.
43. Искусственный нейрон. Функции активации нейронных элементов.
44. Однослойная нейронная сеть. Обучение нейронной сети.
45. Многослойная нейронная сеть.
46. Метод обратного распространения ошибки.

47. Сверточные нейронные сети: структура, принцип работы, области применения.
48. Рекуррентные нейронные сети: структура, принцип работы, области применения.

Критерии оценки:

- оценка «отлично» выставляется, если обучающийся в полном объеме усвоил материал программы учебной дисциплины, раскрыл теоретическое содержание вопросов билета, не затрудняется в ответах на дополнительные вопросы экзаменатора, успешно выполнил практическое задание, продемонстрировав необходимые навыки и умения правильного применения теоретических знаний в практической деятельности;
- оценка «хорошо» выставляется, если обучающийся знает материал программы учебной дисциплины, правильно, по существу и последовательно излагает содержание вопросов билета, в целом правильно выполнил практическое задание, владеет основными умениями и навыками, при ответе не допустил существенных ошибок и неточностей;
- оценка «удовлетворительно» выставляется, если обучающийся усвоил только основные положения материала программы учебной дисциплины, содержание вопросов билета изложил поверхностно, без должного обоснования, допускает неточности и ошибки, недостаточно правильные формулировки, нарушает последовательность в изложении материала, практическое задание выполнил не в полном объеме, испытывает затруднения при ответе на часть дополнительных вопросов;
- оценка «неудовлетворительно» выставляется, если обучающийся не знает основных положений материала программы учебной дисциплины, при ответе на вопросы билета допускает существенные ошибки, не выполнил практическое задание, не смог ответить на большинство дополнительных вопросов или отказался отвечать.