

Государственное образовательное учреждение
"Приднестровский государственный университет им. Т.Г. Шевченко"

Инженерно-технический институт

Кафедра информационных технологий

УТВЕРЖДАЮ

Заведующий кафедрой ИТ

 Ю.А. Столяренко

«28» августа 2023 г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

по дисциплине

РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ

Направление подготовки

2.09.03.02 Информационные системы и технологии

Профиль подготовки

Безопасность информационных систем

Квалификация (степень)

выпускника:

бакалавр

Форма обучения:

очная, заочная

Год набора:

2021 г.

Разработал:

к.т.н., доцент кафедры ИТ

 /В.С. Попускайло

«28» августа 2023 г.

Тирасполь, 2023

Паспорт фонда оценочных средств по учебной дисциплине

1. В результате изучения дисциплины «Рекомендательные системы» у обучающегося должны быть сформированы следующие компетенции:

Категория (группа) компетенций	Код и наименование	Код и наименование индикатора достижения универсальной компетенции
Общепрофессиональные компетенции и индикаторы их достижения		
	ОПК-1. Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	ИД-1 _{ОПК-1} Знать основы высшей математики, физики, экологии, инженерной графики, информатики и программирования ИД-2 _{ОПК-1} Уметь решать стандартные профессиональные задачи с применением естественнонаучных и общеинженерных знаний, методов математического анализа и моделирования ИД-3 _{ОПК-1} Владеть методами теоретического и экспериментального исследования объектов профессиональной деятельности
	ОПК-8. Способен применять математические модели, методы и средства проектирования информационных и автоматизированных систем	ИД-1 _{ОПК-8} Знать: методологию и основные методы математического моделирования, классификацию и условия применения моделей, основные методы и средства проектирования информационных и автоматизированных систем, инструментальные средства моделирования и проектирования информационных и автоматизированных систем. ИД-2 _{ОПК-8} Уметь: применять на практике математические модели, методы и средства проектирования и автоматизации систем на практике. ИД-3 _{ОПК-8} Иметь навыки: моделирования и проектирования информационных и автоматизированных систем.

2. Программа оценивания контролируемой компетенции:

Текущая аттестация	Контролируемые модули, разделы (темы) дисциплины их название	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
РУБЕЖНЫЙ КОНТРОЛЬ	Раздел 1-Раздел 5	ОПК-1, ОПК-8	Лабораторные работы №1-5
РУБЕЖНАЯ АТТЕСТАЦИЯ	Раздел 6-Раздел 12		Лабораторные работы №6-9 Тестирование
Промежуточная аттестация		Код контролируемой компетенции (или ее части)	Наименование оценочного средства
№1		ОПК-1, ОПК-8	зачет

3. Показатели и критерии оценивания компетенции по этапам формирования, описание шкал оценивания

Этапы оценивания компетенции	Показатели достижения заданного уровня освоения компетенции	Критерии оценивания результатов обучения			
		2	3	4	5
Первый этап	ИД-1 ОПК-1 Знать основы высшей математики, физики, экологии, инженерной графики, информатики и программирования	Не знает	Знает основные понятия но не знает способы использования в профессиональной деятельности	Знает основы высшей математики, физики, экологии, инженерной графики, информатики и программирования, но не может применить знания в полной мере в профессиональной деятельности	Знает основы высшей математики, физики, экологии, инженерной графики, информатики и программирования и может использовать в профессиональной деятельности
Второй этап	ИД-2 ОПК-1 Уметь решать стандартные профессиональные задачи с применением естественнонаучных и общинженерных знаний, методов математического анализа и моделирования	Не умеет	Уметь решать некоторые стандартные профессиональные задачи с применением естественнонаучных и общинженерных знаний, методов математического анализа и моделирования	Умеет решать стандартные профессиональные задачи с применением естественнонаучных и общинженерных знаний, методов математического анализа и моделирования, но не в полной мере	Умеет решать стандартные профессиональные задачи с применением естественнонаучных и общинженерных знаний, методов математического анализа и моделирования
Третий этап	ИД-3 ОПК-1 Владеть методами теоретического и	Не владеет	Владеет методами теоретического и	Владеет методами теоретического и	Владеет методами теоретического и

Этапы оценивания компетенции	Показатели достижения заданного уровня освоения компетенции	Критерии оценивания результатов обучения			
		2	3	4	5
	экспериментального исследования объектов профессиональной деятельности		ского и экспериментального исследования объектов профессиональной деятельности, но не владеет ими в междисциплинарном контексте	ческого и экспериментального исследования объектов профессиональной деятельности, но ошибается в обработке их результатов	экспериментального исследования объектов профессиональной деятельности, в том числе в новой или незнакомой среде и в междисциплинарном контексте
Первый этап	ИД-1 _{ОПК-8} Знать: методологию и основные методы математического моделирования, классификацию и условия применения моделей, основные методы и средства проектирования информационных и автоматизированных систем, инструментальные средства моделирования и проектирования информационных и автоматизированных систем	Не знает	Знает основные методы но не знает способы использования в профессиональной деятельности	Знает методологию и основные методы математического моделирования, но не может применять знания в полной мере в профессиональной деятельности	Знает методологию и основные методы математического моделирования, классификацию и условия применения моделей, основные методы и средства проектирования информационных и автоматизированных систем, инструментальные средства моделирования и проектирования информационных и автоматизированных систем и может использовать в профессиональной деятельности
Второй этап	ИД-2 _{ОПК-8} Уметь: применять на практике математические модели, методы и средства проектирования и автоматизации систем.	Не умеет	Уметь применять на практике некоторые стандартные математические модели, методы и средства проектирования и автоматизации систем	Умеет применять на практике стандартные математические модели, методы и средства проектирования и автоматизации систем, но не в полной мере	Умеет применять на практике математические модели, методы и средства проектирования и автоматизации систем, и решать задачи из профессиональной области
Третий этап	ИД-3 _{ОПК-8} Иметь навыки: моделирования и проектирования информационных и автоматизированных систем.	Не владеет	Владеет методами моделирования и проектирования информационных и автоматизированных систем, но	Владеет методами моделирования и проектирования информационных и автоматизированных систем	Владеет методами моделирования и проектирования информационных и автоматизированных систем, в том числе в новой

Этапы оценивания компетенции	Показатели достижения заданного уровня освоения компетенции	Критерии оценивания результатов обучения			
		2	3	4	5
			не владеет ими в междисциплинарном контексте	систем, но ошибается в незнакомых предметных областях	или незнакомой среде и в междисциплинарном контексте

4. Шкала оценивания

Согласно Положению «О порядке организации аттестации в ИТИ ПГУ им. Т.Г. Шевченко, итоговая оценка представляет собой сумму баллов, полученных студентом по итогу освоения дисциплины (модуля):

Оценка в традиционной шкале	Оценка в 100-балльной шкале	Буквенные эквиваленты оценок в шкале 3Е (% успешно аттестованных)
5 (отлично)	88–100	A (отлично) – 88-100 баллов
4 (хорошо)	70–87	B (очень хорошо) – 80-87 баллов
		C (хорошо) – 70-79 баллов
3 (удовлетворительно)	50–69	D (удовлетворительно) – 60-69 баллов
		E (посредственно) – 50-59 баллов
2 (неудовлетворительно)	0–49	Fx – неудовлетворительно, с возможной пересдачей – 21-49 баллов
		F – неудовлетворительно, с повторным изучением дисциплины – 0-20 баллов

Расшифровка уровня знаний, соответствующего полученным баллам, дается в таблице, указанной ниже

A	“Отлично” - теоретическое содержание курса освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, все предусмотренные программой обучения учебные задания выполнены, качество их выполнения оценено числом баллов, близким к максимальному.
B	“Очень хорошо” - теоретическое содержание курса освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом в основном сформированы, все предусмотренные программой обучения учебные задания выполнены, качество выполнения большинства из них оценено числом баллов, близким к максимальному.
C	“Хорошо” - теоретическое содержание курса освоено полностью, без пробелов, некоторые практические навыки работы с освоенным материалом сформированы недостаточно, все предусмотренные программой обучения учебные задания выполнены, качество выполнения ни одного из них не оценено минимальным числом баллов, некоторые виды заданий выполнены с ошибками.
D	“Удовлетворительно” - теоретическое содержание курса освоено частично, но пробелы не носят существенного характера, необходимые практические навыки работы с освоенным материалом в основном сформированы, большинство предусмотренных программой обучения учебных заданий выполнено, некоторые из выполненных заданий, возможно, содержат ошибки.
E	“Посредственно” - теоретическое содержание курса освоено частично, некоторые практические навыки работы не сформированы, многие предусмотренные программой обучения учебные задания не выполнены, либо качество выполнения некоторых из них оценено числом баллов, близким к минимальному.

FX	“Условно неудовлетворительно” - теоретическое содержание курса освоено частично, необходимые практические навыки работы не сформированы, большинство предусмотренных программой обучения учебных заданий не выполнено, либо качество их выполнения оценено числом баллов, близким к минимальному; при дополнительной самостоятельной работе над материалом курса возможно повышение качества выполнения учебных заданий.
F	“Безусловно неудовлетворительно” - теоретическое содержание курса не освоено, необходимые практические навыки работы не сформированы, все выполненные учебные задания содержат грубые ошибки, дополнительная самостоятельная работа над материалом курса не приведет к какому-либо значимому повышению качества выполнения учебных заданий.

5. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций при изучении учебной дисциплины в процессе освоения образовательной программы

5.1 Образец индивидуального задания к лабораторной работе №1

1. Написать функцию, в соответствии с индивидуальным вариантом, принимающую на вход аргументы, вводимые пользователем с клавиатуры, выполняющую определённые действия и сохраняющую результат в файл.

5.2. Образец индивидуального задания к лабораторной работе №2

Проведите исследовательский анализ данных предложенной Базы данных. В качестве отчёта о работе оформите `ipynb notebook` или `Rmd` файл, в котором будут ответы на следующие вопросы:

1. Сколько заказов попало в выборку?
2. Есть ли в данных пропуски, в каких колонках?
3. Какие 5 товаров были самыми дешёвыми/ самыми дорогими?
4. Какая самая популярная категория товара?
5. В каком количестве заказов есть товар, который стоил более 50% от стоимости чека?

5.3 Образец индивидуального задания к лабораторной работе №3

1. Постройте график зависимости времени работы алгоритма `Apriori` библиотеки `DAAL` от значения `minsup` при фиксированном `minconf = 0.2`.
2. Постройте график зависимости времени работы алгоритма `Apriori` библиотеки `DAAL` от значения `minconf` при фиксированном `minsup = 0.02`.
3. Реализуйте в `Python` алгоритм `Apriori` с процедурой построения ассоциативных правил. Сравните время работы реализованного алгоритма с алгоритмом `Apriori`, реализованным в `DAAL`.

5.4 Образец индивидуального задания к лабораторной работе №4

1. Загрузить данные из файла «Данные для расчетов».
2. Сделать срез из данных, такой, чтобы в исследуемом `data.frame` остались только столбцы с номерами: [№ по журналу; № по журналу+10]
3. Проверить предпосылки для применения классического однофакторного дисперсионного анализа.
4. Провести параметрический или непараметрический дисперсионный анализ, в зависимости от полученных результатов на предыдущем шаге.
5. При отклонении гипотезы о равенстве всех математических ожиданий провести попарные сравнения изучаемых групп.
6. Визуализировать полученные различия при помощи `boxplot + violinplot`.
7. Сделать выводы о проделанной работе.

5.5 Образец индивидуального задания к лабораторной работе №5

1. Загрузить набор данных для своего варианта, ознакомиться с его содержимым.
2. Построить график корреляционного поля для каждого фактора.
3. Построить уравнение парной линейной регрессии для каждого фактора
4. Проверить значимость каждого из полученных уравнений регрессии. Показать уравнения регрессии с заданным в варианте доверительным интервалом на графиках.
5. Построить прогнозы по каждому из уравнений парной регрессии для заданных в варианте значений факторов.
6. Построить уравнение множественной линейной регрессии и получить корреляционную матрицу.
7. Построить прогноз по уравнению множественной регрессии для заданных в варианте значений факторов.

5.6 Образец индивидуального задания к лабораторной работе №6

1. Загрузить набор данных для своего варианта, ознакомиться с его содержимым.
2. Построить модель логистической регрессии для заданной целевой переменной.
3. Оцените качество полученной модели. Определите, какой метрикой лучше пользоваться в данном случае.
4. Постройте ROC кривую для полученной модели.
5. Постройте дерево решений для данных своего варианта.
6. Оцените качество полученной модели. Определите, какой метрикой лучше пользоваться в данном случае.
7. Постройте ROC кривую для полученной модели.
8. Сделать выводы о проделанной работе.

5.7 Образец индивидуального задания к лабораторной работе №7

1. Загрузить набор данных для своего варианта, ознакомиться с его содержимым.
2. Используя пакет caret постройте модели: наивный байесовский классификатор, машины опорных векторов, случайный лес и Adaboost.
3. Оцените качество полученных моделей при помощи метрик Ассигасу и Карра.
4. Сделать выводы о проделанной работе.

5.8 Образец индивидуального задания к лабораторной работе №8

1. Загрузить набор данных для своего варианта, ознакомиться с его содержимым.
2. Проведите процедуру кластеризации при помощи метода k-means и иерархической кластеризации.
3. Постройте дендрограмму и определите оптимальное количество кластеров.
4. Сделать выводы о проделанной работе.

5.9 Образец индивидуального задания к лабораторной работе №9

1. Загрузить набор данных для своего варианта, ознакомиться с его содержимым.
2. Проведите разведывательный анализ данных, ответьте на вопросы.
3. Постройте графики, указанные в заданиях.
4. Рассчитайте статистические критерии.
5. Оцените величину корреляционной связи между величинами.
6. Постройте математическую модель по данным.
7. Оформите отчёт в виде RMarkdown файла.

5.10 Образец тестового задания №1

Пример вопросов теста №1:

- Что такое рекомендательная система?
 - а) Система анализа сетевого трафика
 - б) Система защиты информации
 - в) Система, предлагающая персонализированный контент
 - г) Система управления данными

- Какой основной источник данных используют рекомендательные системы?
 - а) Сетевой трафик
 - б) Поведение пользователей
 - в) Случайная генерация
 - г) DNS-запросы

- Какой тип рекомендательной системы основан на действиях других пользователей?
 - а) Контентная
 - б) Гибридная
 - в) Коллаборативная
 - г) Автоматическая

- Что такое коллаборативная фильтрация?
 - а) Метод фильтрации нежелательных писем
 - б) Фильтрация по контенту страницы
 - в) Рекомендации на основе поведения похожих пользователей
 - г) Метод анонимизации

- Что делает контентная рекомендательная система?
 - а) Ищет вирусы
 - б) Учитывает предпочтения схожих пользователей
 - в) Рекомендует элементы, схожие с теми, что пользователь оценил ранее
 - г) Удаляет дубликаты

- Что означает гибридная рекомендательная система?
 - а) Объединение нескольких алгоритмов
 - б) Система, основанная только на случайных данных
 - в) Комбинация с файловой системой
 - г) Применение исключительно ручной фильтрации

- Какой алгоритм используется в коллаборативной фильтрации?
 - а) Поиск по ключевым словам
 - б) К-средних
 - в) Алгоритм ближайших соседей
 - г) Хеширование

- Что такое матрица взаимодействия в рекомендательной системе?
 - а) Таблица соединений между серверами
 - б) Список системных процессов
 - в) Матрица пользователей и объектов с оценками
 - г) Таблица HTML-страниц

- Что такое "user-based" подход?
 - а) Метод анализа кода
 - б) Метод оценки производительности
 - в) Рекомендации на основе схожих пользователей
 - г) Выдача случайных объектов

- Что такое "item-based" подход?
 - а) Фильтрация по дате
 - б) Рекомендации, основанные на схожести между элементами
 - в) Метод защиты
 - г) Генератор контента

- Что такое cold start в рекомендательных системах?
 - а) Ошибка ввода
 - б) Запуск сервера после сбоя
 - в) Ситуация, когда система не имеет данных о новом пользователе или объекте
 - г) Стартовая страница сайта

- Какой метод уменьшает размерность данных в матрице оценок?
 - а) Хэш-функция
 - б) Линейная регрессия
 - в) SVD (сингулярное разложение)
 - г) DNS

- Что используется в системе рекомендаций на основе содержимого (контентной)?
 - а) IP-адрес пользователя
 - б) Характеристики объектов
 - в) Только предпочтения друзей
 - г) Погода

- Какая метрика используется для оценки схожести пользователей?
 - а) Евклидово расстояние
 - б) Алгоритм Дейкстры
 - в) Скорость загрузки
 - г) Глубина кеша

- Что означает термин RMSE?
 - а) Реальная модель системы
 - б) Среднеквадратичная ошибка
 - в) Размер файла
 - г) Случайная выборка

- Что такое implicit feedback?
 - а) Прямые оценки пользователя
 - б) Выборка по логам
 - в) Непрямые признаки интереса (например, клики)
 - г) Спам-фильтрация

- Какой тип фильтрации лучше подходит для новых пользователей?
 - а) Коллаборативная
 - б) Контентная
 - в) Гибридная
 - г) Фильтрация по геолокации

- Какой из языков чаще используется при разработке рекомендательных систем?
 - а) Pascal
 - б) HTML
 - в) Python
 - г) PHP

- Что такое рейтинг в рекомендательной системе?
 - а) Код страницы
 - б) Статус регистрации
 - в) Оценка, которую пользователь ставит объекту
 - г) Пароль пользователя

- Что из перечисленного можно рекомендовать с помощью рекомендательных систем?
 - а) Только фильмы
 - б) Только товары
 - в) Различные объекты: фильмы, книги, товары и т.п.
 - г) Только видео

- Что такое фильтрация по памяти (memory-based)?
 - а) Метод удаления лишних данных
 - б) Подход, хранящий всю матрицу взаимодействий
 - в) Хранилище логов
 - г) Сжатие файла

- Что такое precision в оценке системы?
 - а) Точность рекомендаций (доля релевантных объектов в рекомендуемых)
 - б) Скорость вывода
 - в) Время отклика сервера
 - г) Количество пользователей в системе

- Что измеряет recall?
 - а) Количество новых пользователей
 - б) Долю релевантных объектов из всех возможных
 - в) Время построения модели
 - г) Размер базы данных

- Какой алгоритм чаще всего используется в построении модели машинного обучения для рекомендаций?
 - а) KNN
 - б) Decision Tree
 - в) Logistic Regression
 - г) SVM

- Какой подход не требует явных оценок пользователей?
 - а) Явная фильтрация
 - б) Гибридная фильтрация
 - в) Имплиcitная фильтрация
 - г) Тестовая фильтрация

- Что такое overfitting в модели рекомендательной системы?
 - а) Переобучение, плохая обобщающая способность
 - б) Недостаток данных
 - в) Отсутствие связи между элементами
 - г) Ошибка компиляции

- Что означает MAE?
 - а) Максимальная активность элементов
 - б) Средняя абсолютная ошибка

- в) Общее количество рекомендаций
- г) Метод анонимной оценки

• Что такое кластеризация в рекомендательных системах?

- а) Отправка уведомлений
- б) Выделение групп схожих пользователей или объектов
- в) Фильтрация рекламы
- г) Удаление базы

• Какой метод позволяет обрабатывать рекомендации в реальном времени?

- а) Batch-фильтрация
- б) Online-фильтрация
- в) Адаптивное кеширование
- г) Инкапсуляция

• Что делает матричная факторизация?

- а) Строит отчет по пользователям
- б) Делит пользователей по алфавиту
- в) Представляет взаимодействие пользователей и объектов в виде скрытых признаков
- г) Удаляет старые записи

Перечень вопросов к экзамену по предмету

1. Распределение Хи-квадрат.
2. Этапы анализа данных.
3. Рассчитайте точечную и интервальную оценки среднего и дисперсии.
4. Распределение Стьюдента.
5. Однофакторный дисперсионный анализ.
6. Рассчитайте величину линейной корреляционной связи между выборками.
7. Нормальное распределение.
8. Определите, являются ли значимыми различия между выборками.
9. Равномерное распределение.
10. Оценка качества классификатора.
11. Рассчитайте описательные статистики для выборки: среднее арифметическое, коэффициент вариации, выборочную дисперсию, стандартное отклонение.
12. Корреляционный анализ. Методы и подходы.
13. Определите, имеют ли распределения выборок равные дисперсии.
14. Регрессионный анализ. Цели и задачи.
15. Проверьте гипотезу о принадлежности генеральной совокупности к нормальному закону распределения по данным выборки.
16. Идея р-уровня значимости.
17. Оценка качества регрессионной модели.
18. Квантили распределения.
19. Графические способы оценки вида закона распределения.
20. Свойства статистических оценок.
21. Дисперсионный анализ. Цели и задачи.
22. Понятие мощности статистических критериев. Виды ошибок.
23. Гипотеза о наличии в данных выбросов.
24. Центральная предельная теорема.

25. Алгоритмы кластеризации.
26. Алгоритмы классификации.
27. Ограничения метода наименьших квадратов.
28. Логистическая регрессия.
29. Задача обработки естественного языка.