

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

Физико-технический институт

Кафедра информационных технологий

УТВЕРЖДАЮ

Заведующий кафедрой ИТ



Ю.А. Столяренко

«28» августа 2023 г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

по дисциплине

БОЛЬШИЕ ДАННЫЕ

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 ИТиАУПП,



/В.С. Попукайло

«28» августа 2023 г.

Тирасполь, 2023

Паспорт фонда оценочных средств по учебной дисциплине

1. В результате изучения дисциплины «Статистическая обработка данных» у обучающегося должны быть сформированы следующие компетенции:

Категория универсальных компетенций	Код и наименование универсальной компетенции	Код и наименование индикатора достижения универсальной компетенции
Универсальные компетенции выпускников и индикаторы их достижения		
Разработка и реализация проектов	УК-2. Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	ИД-1 _{УК-2} Знать: виды ресурсов и ограничений для решения профессиональных задач: основные методы оценки разных способов решения задач: действующее законодательство и правовые нормы, регулирующие профессиональную деятельность
		ИД-2 _{УК-2} Уметь: проводить анализ поставленной цели и формулировать задачи, которые необходимо решить для ее достижения: анализировать альтернативные варианты решений для достижения намеченных результатов: использовать нормативно-правовую документацию в сфере профессиональной деятельности
		ИД-3 _{УК-2} Владеть: методиками разработки цели и задач проекта: методами оценки потребности в ресурсах, продолжительности и стоимости проекта: навыками работы с нормативно-правовой документацией.
Самоорганизация и саморазвитие (в том числе здоровьесбережение)	УК-6. Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни	ИД-1 _{УК-6} Знать: основные приемы эффективного управления собственным временем: основные методики самоконтроля, саморазвития и самообразования на протяжении всей жизни.
		ИД-2 _{УК-6}

		Уметь: эффективно планировать и контролировать собственное время; использовать методы саморегуляции. саморазвития и самообучения. ИД-Зук-6 Владеть: методами управления собственным временем; технологиями приобретения, использования и обновления социокультурных и профессиональных знаний, умений и навыков; методиками саморазвития и самообразования в течение всей жизни.
--	--	---

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 2 Раздел 3	УК-6, УК-6	Лабораторные работы №1-5
РУБЕЖНАЯ АТТЕСТАЦИЯ	Раздел 4 Раздел 5 Раздел 6		Лабораторные работы №6-9 Тестирование
Промежуточная аттестация		Код контролируемой компетенции (или ее части)	Наименование оценочного средства
№1		УК-6, УК-6	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 _{УК-2} Знать: виды ресурсов и ограничений для решения профессиональных задач: основные методы оценки разных способов решения задач: действующее законодательство и правовые нормы, регулирующие профессиональную деятельность	Не знает	Знает виды ресурсов	Знает виды ресурсов и ограничений для решения профессиональных задач	Знает виды ресурсов и ограничений для решения профессиональных задач: основные методы оценки разных способов решения задач: действующее законодательство и правовые нормы, регулирующие профессиональную деятельность
Второй этап	ИД-2 _{УК-2} Уметь: проводить анализ поставленной цели и формулировать задачи, которые необходимо решить для ее достижения: анализировать альтернативные варианты решений для достижения намеченных результатов: использовать нормативно-правовую документацию в сфере профессиональной деятельности	Не умеет	Умет: проводить анализ поставленной цели	Умеет проводить анализ поставленной цели и формулировать задачи	Умеет проводить анализ поставленной цели и формулировать задачи, которые необходимо решить для ее достижения: анализировать альтернативные варианты решений для достижения намеченных результатов: использовать нормативно-правовую документацию в сфере профессиональной деятельности
Третий этап	ИД-3 _{УК-2}	Не владеет	Владеет методиками разработки	Владеет методиками разработки	Владеет методиками разработки

Этапы оценивания компетенции	Показатели достижения заданного уровня освоения компетенции	Критерии оценивания результатов обучения			
		2	3	4	5
	Владеть: методиками разработки цели и задач проекта: методами оценки потребности в ресурсах. продолжительности и стоимости проекта: навыками работы с нормативно-правовой документацией.		цели и задач проекта	цели и задач проекта: методами оценки потребности в ресурсах	цели и задач проекта: методами оценки потребности в ресурсах. продолжительности и стоимости проекта: навыками работы с нормативно-правовой документацией.
Первый этап	ИД-1 _{УК-6} Знать: основные приемы эффективного управления собственным временем: основные методики самоконтроля, саморазвития и самообразования на протяжении всей жизни.	Не знает	Знает основные приемы эффективного управления	Знает основные приемы эффективного управления собственным временем	Знает основные приемы эффективного управления собственным временем: основные методики самоконтроля, саморазвития и самообразования на протяжении всей жизни.
Второй этап	ИД-2 _{УК-6} Уметь: эффективно планировать и контролировать собственное время; использовать методы саморегуляции, саморазвития и самообучения.	Не умеет	Умеет эффективно планировать и контролировать собственное время	Умеет эффективно планировать и контролировать собственное время; использовать методы саморегуляции	Умеет эффективно планировать и контролировать собственное время; использовать методы саморегуляции, саморазвития и самообучения.
Третий этап	ИД-3 _{УК-6} Владеть: методами управления собственным временем; технологиями приобретения, использования и обновления социокультурных и профессиональных знаний, умений и навыков; методиками саморазвития и самообразования в течение всей жизни.	Не владеет	Владеет методами управления собственным временем	Владеет методами управления собственным временем; технологиями приобретения, использования и обновления социокультурных и профессиональных знаний	Владеет методами управления собственным временем; технологиями приобретения, использования и обновления социокультурных и профессиональных знаний, умений и навыков; методиками саморазвития и самообразования в течение всей жизни

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 ЗЕ (% успешно аттестованных)
5 (отлично)	88–100	А (отлично) – 88-100 баллов

4 (хорошо)	70–87	В (очень хорошо) – 80-87баллов
		С (хорошо) – 70-79 баллов
3 (удовлетворительно)	50–69	D (удовлетворительно) – 60-69 баллов
		Е (посредственно) – 50-59 баллов
2 (неудовлетворительно)	0–49	Fx – неудовлетворительно, с возможной передачей – 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. Сгенерировать 100 выборок со стандартным нормальным законом распределения ($M=0$, $s=1$) и объемом выборки $N=(N_0 \text{ по варианту} + 15)$.
2. Рассчитать для каждой выборки точечные оценки среднего, дисперсии и стандартного отклонения.

3. Рассчитать средние значения для оцененных параметров. Сравнить полученные результаты с генеральными значениями параметров.

4. Рассчитать для каждой выборки интервальные значения среднего, дисперсии и стандартного отклонения, приняв за априорное утверждение, что нам не известны генеральные значения данных параметров. Расчет произвести для вероятностей $P=0,90$, $P=0,95$.

5. Сделать выводы о точности точечных и интервальных оценок неизвестных параметров распределений.

5.2. Образец индивидуального задания к лабораторной работе №2

1. Сгенерировать 1000 выборок со стандартным нормальным законом распределения и объемом выборки N , где $N = \text{№ по списку в журнале} + 20$.
2. Рассчитать для каждой выборки точечные оценки среднего.
3. Для каждой сотой выборки построить графики, позволяющие оценить вид закона распределения: гистограммы, *qq-plot*, *boxplot*, *violinplot*.
4. Построить аналогичные графики для вектора, состоящего из средних значений.
5. Сделать выводы о проделанной работе.

5.3 Образец индивидуального задания к лабораторной работе №3

1. Загрузить файл «Данные для расчетов».
2. Проверить гипотезу о нормальности для каждой выборки.
3. Построить гистограммы выборок, для которых удалось отвергнуть гипотезу о нормальности.
4. Произвести логарифмическую трансформацию и изучить её влияние на распределение данных, построив гистограммы и проверив гипотезу о нормальности.
5. Для наглядности, сгенерировать выборку из экспоненциального распределения и построить гистограммы по сгенерированным данным и после их логарифмирования.
6. Сделать выводы о проделанной работе.

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 содержанием.
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 отчет в виде RMarkdown файла.

5.10 Образец тестового задания №1

• Что обозначает термин "Большие данные" (Big Data)?

- а) Большие файлы мультимедиа
- б) Данные, которые не помещаются на жесткий диск
- в) Объемные, разнообразные и быстро поступающие данные
- г) Данные в архивном формате

• Какая характеристика **не** относится к принципу 3V в Big Data?

- а) Скорость (Velocity)
- б) Ценность (Value)
- в) Объем (Volume)
- г) Разнообразие (Variety)

- Что такое Hadoop?
 - а) СУБД
 - б) Облачная платформа
 - в) Фреймворк для обработки больших данных
 - г) Язык программирования

- Какой компонент Hadoop отвечает за распределенное хранение данных?
 - а) MapReduce
 - б) YARN
 - в) HDFS
 - г) Spark

- Что делает MapReduce?
 - а) Распределяет IP-адреса
 - б) Управляет транзакциями
 - в) Обрабатывает данные в виде пар "ключ–значение"
 - г) Сжимает данные

- Какой инструмент является системой распределенной обработки в памяти?
 - а) Hive
 - б) Pig
 - в) Spark
 - г) HBase

- Что такое HDFS?
 - а) Распределённая файловая система
 - б) Графическая панель управления
 - в) Система для создания SQL-запросов
 - г) Язык описания данных

- Что такое NoSQL?
 - а) База данных без индексов
 - б) Отказ от реляционной модели
 - в) Сжатие данных
 - г) Язык программирования

- Какая из систем относится к NoSQL?
 - а) MySQL
 - б) Oracle
 - в) MongoDB
 - г) PostgreSQL

- Для чего используется Apache Hive?
 - а) Хранение логов
 - б) Написание SQL-подобных запросов в Hadoop
 - в) Управление сетями
 - г) Построение графов

- Что такое кластер в контексте Big Data?
 - а) Один сервер с большим объемом ОЗУ
 - б) Группа соединённых серверов, работающих совместно
 - в) Данные в облаке
 - г) Локальная база данных

- Что такое ETL?
 - а) Хранилище данных
 - б) Сетевой протокол
 - в) Процесс извлечения, трансформации и загрузки данных
 - г) Язык программирования

- Какой язык используется в Apache Pig?
 - а) SQL
 - б) Pig Latin
 - в) Scala
 - г) Bash

- Что такое Data Lake?
 - а) Фреймворк обработки SQL
 - б) Система мониторинга
 - в) Централизованное хранилище всех типов данных
 - г) Программа виртуализации

- Что из перечисленного является типичной задачей Big Data?
 - а) Печать документов
 - б) Монтаж видео
 - в) Анализ поведения клиентов на миллионах записей
 - г) Установка драйверов

- Какой формат данных наиболее распространен для хранения в Hadoop?
 - а) CSV
 - б) TXT
 - в) XML
 - г) Parquet

- Что означает термин "потокковые данные" (streaming data)?
 - а) Данные, хранящиеся на оптических дисках
 - б) Постоянно поступающие в реальном времени данные
 - в) Архивированные файлы
 - г) Лог-файлы в Excel

- Что из следующего не относится к инструментам Big Data?
 - а) Apache Spark
 - б) Microsoft Excel
 - в) Hadoop
 - г) Apache Flink

- Какой язык программирования чаще всего используется с Apache Spark?
 - а) Pascal
 - б) Scala
 - в) HTML
 - г) PHP

- Что из перечисленного описывает "верификация данных"?
 - а) Проверка электронной подписи
 - б) Удаление ненужных данных
 - в) Проверка качества и достоверности данных
 - г) Шифрование

- Что означает термин Data Mining?
 - а) Хранение данных в облаке
 - б) Защита данных
 - в) Выявление скрытых закономерностей в больших объемах данных
 - г) Перенос данных между устройствами

- Что такое Apache Flink?
 - а) Облачное хранилище
 - б) Система для визуализации
 - в) Поточковая распределённая платформа обработки данных
 - г) Архиватор

- Какая технология используется для хранения колонок в базе данных?
 - а) JSON
 - б) Column-oriented storage
 - в) Tree-index
 - г) Hash-table

- Что такое "шардирование"?
 - а) Сжатие таблиц
 - б) Хеширование паролей
 - в) Разделение базы данных на части для масштабирования
 - г) Подключение облака

- Какой инструмент часто используется для визуализации Big Data?
 - а) Git
 - б) Tableau
 - в) GCC
 - г) MariaDB

- Что означает термин "масштабируемость"?
 - а) Работа только в одной сети
 - б) Встроенная защита
 - в) Возможность увеличения объема данных или узлов
 - г) Архивирование

- Какой из форматов является бинарным для Big Data?
 - а) CSV
 - б) JSON
 - в) Avro
 - г) HTML

- Что такое latency в Big Data-системах?
 - а) Количество данных
 - б) Пропускная способность сети
 - в) Задержка при обработке или доставке данных
 - г) Размер кластера

- Какой компонент отвечает за управление ресурсами в Hadoop?
 - а) Hive
 - б) Yarn
 - в) Pig
 - г) ZooKeeper

- Что из перечисленного описывает термин "обогащение данных"?
 - а) Перевод данных в архив
 - б) Добавление полезной информации к данным
 - в) Сжатие файлов
 - г) Очистка памяти