

**Приложение 2 к РПД Технологии обработки информации
09.03.02 Информационные системы и технологии
Направленность (профиль) – Программно-аппаратные комплексы
Форма обучения – заочная
Год набора - 2017**

**ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ
АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

1. Общие сведения

1.	Кафедра	Информатики и вычислительной техники
2.	Направление подготовки	09.03.02 Информационные системы и технологии
3.	Направленность (профиль)	Программно-аппаратные комплексы
4.	Дисциплина (модуль)	Технологии обработки информации
5.	Форма обучения	заочная
6.	Год набора	2017

2. Перечень компетенций

- | |
|---|
| <ul style="list-style-type: none">– способность разрабатывать, согласовывать и выпускать все виды проектной документации (ПК-10);– способность оптимизировать функционирование баз данных (ПСК-1). |
|---|

3. Критерии и показатели оценивания компетенций на различных этапах их формирования

Этап формирования компетенции (разделы, темы дисциплины)	Формируемая компетенция	Критерии и показатели оценивания компетенций			Формы контроля сформированности компетенций
		Знать:	Уметь:	Владеть:	
1. Введение в предмет. Основные процедуры обработки данных.	ПК-10 ПСК-1	основные виды и задачи обработки данных; основные процедуры обработки данных; принципы создания хранилищ и витрин данных; правила Кодда для хранилищ данных; принципы организации OLAP	осуществлять математическую и информационную постановку задач по обработке информации; использовать алгоритмы обработки информации для различных приложений;	Терминологической базой	Тест
2. Большие данные и машинное обучение.	ПК-10 ПСК-1	понятие больших данных и их свойства; постановку задачи классификации и регрессии	выполнять постановку задачи машинного обучения	навыками предобработки данных, используя классы библиотеки Pandas	Лабораторная работа, реферат, тест
3. Метрические методы классификации	ПК-10 ПСК-1	формализацию задачи; понятие обобщенного метрического классификатора; алгоритмы метрической классификации; метод отбора эталонов, алгоритм STOLP	применять метрические алгоритмы классификации для анализа данных	навыками применения алгоритма k взвешенных ближайших соседей, используя классы из библиотеки scikit-learn	Лабораторная работа Тест
4. Логические методы классификации	ПК-10 ПСК-1	основные принципы построения логических алгоритмов классификации; критерии информативности: простые критерии, статистический критерий, энтропийный критерий; алгоритм построения дерева ID 3	применять алгоритм ID 3 для классификации данных	навыками работы с алгоритмами классификации на основе деревьев решений, используя классы из библиотеки scikit-learn	Лабораторная работа
5. Линейные методы классификации	ПК-10 ПСК-1	основные принципы построения линейных алгоритмов классификации; алгоритм стохастического градиента; метод SVM	использовать линейные методы для классификации данных	навыками работы с линейными алгоритмами классификации, используя классы из библиотеки scikit-learn	Лабораторная работа

4. Критерии и шкалы оценивания

4.1. Тест

Процент правильных ответов	до 50	51-60	61-80	81-100
Количество баллов за ответы	0	1	2	3

4.2. Реферат

Характеристики выполнения реферата	Баллы
1. Новизна реферированного текста: актуальность проблемы и темы; новизна и самостоятельность в постановке проблемы, в формулировании нового аспекта выбранной для анализа проблемы; наличие авторской позиции, самостоятельность суждений.	3
2. Степень раскрытия сущности проблемы: соответствие плана теме реферата; соответствие содержания теме и плану; полнота и глубина раскрытия основных понятий; обоснованность способов и методов работы с материалом; умение работать с литературой, систематизировать и структурировать материал; умение обобщать, сопоставлять различные точки зрения по рассматриваемому вопросу, аргументировать основные положения и выводы.	2
3. Обоснованность выбора источников: круг, полнота использования литературных источников по теме; привлечение новейших работ (журнальные публикации, материалы сборников научных трудов и т.д.).	2
4. Соблюдение требований к оформлению: правильное оформление ссылок на используемую литературу; грамотность и культура изложения; владение терминологией и понятийным аппаратом; соблюдение требований к объему работы; культура оформления: выделение абзацев; использование информационных технологий.	2
5. Грамотность: отсутствие орфографических и синтаксических ошибок, стилистических погрешностей; опечаток, сокращений слов, кроме общепринятых; наличие литературного стиля изложения.	2
Максимальное количество баллов	11

4.3. Выполнение лабораторной работы

10 баллов выставляется, если обучающийся выполнил полностью все задания указанные в лабораторной работе и может аргументировано пояснить ход своего решения.

5 баллов выставляется, если обучающийся выполнил не менее 85 % заданий указанных в лабораторной работе, и может аргументировано пояснить ход своего решения и указать.

2 балла выставляется, если обучающийся решил не менее 50% заданий указанных в лабораторной работе, и может аргументировано пояснить ход своего решения.

0 баллов выставляется, если обучающийся не может аргументировано пояснить ход своего решения.

В случае если сроки сдачи работ превышены, количество баллов сокращается на 50%.

4.4. Подготовка опорного конспекта

Подготовка материалов опорного конспекта является эффективным инструментом систематизации полученных обучающимся знаний в процессе изучения дисциплины.

Составление опорного конспекта представляет собой вид внеаудиторной самостоятельной работы обучающегося по созданию краткой информационной структуры, обобщающей и отражающей суть материала лекции, темы учебника. Опорный конспект призван выделить главные объекты изучения, дать им краткую характеристику, используя символы, отразить связь с другими элементами. Основная цель опорного конспекта – облегчить запоминание. В его составлении используются различные базовые понятия, термины, знаки (символы) — опорные сигналы. Опорный конспект может быть представлен системой взаимосвязанных геометрических фигур, содержащих блоки концентрированной информации в виде ступенек логической лестницы; рисунка с дополнительными элементами и др.

Критерии оценки опорного конспекта	Максимальное количество баллов
– подготовка материалов опорного конспекта по изучаемым темам дисциплины только в текстовой форме;	7
– подготовка материалов опорного конспекта по изучаемым темам дисциплины в текстовой форме, которая сопровождается схемами, табличной информацией, графиками, выделением основных мыслей с помощью цветов, подчеркиваний.	15

5. Типовые контрольные задания и методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

5.1. Типовое тестовое задание

1. При каком подходе к измерению информации используется тезаурусная мера?

- 1: Семантический;
- 2: Прагматический;
- 3: Статистический;

2. Какая операция над гиперкубом определяет переход от детального представления данных к агрегированному?

1. консолидация;
2. срез;
3. вращение;

3. Какое понятие определяет следующее высказывание «Множественная перспектива, состоящая из нескольких независимых измерений, вдоль которых могут быть проанализированы определенные совокупности данных»?

1. реляционная модель данных;
2. многомерное представление данных;
3. хранилище данных;

4. К какому типу задач машинного обучения относится задача предсказания цены жилья по его характеристикам?

- 1: Классификация на два класса;
- 2: Классификация на M непересекающихся классов;
- 3: Классификация на M пересекающихся классов;
- 4: Восстановление регрессии;

5. К какому типу относятся признаки «Цвет глаз»?

- 1: *Бинарный;*
- 2: *Количественный;*
- 3: *Номинальный (категориальный);*
- 4: *Порядковый;*

6. К какому типу задач машинного обучения, относится задача в которой необходимо определить независимые группы и их характеристики во всем множестве анализируемых данных?

1. *задача классификации;*
2. *задача регрессии;*
3. *задача кластеризации;*

7. К какому типу задач машинного обучения, относится задача в которой необходимо определить зависимости между объектами или событиями?

1. *задача распознавания образов;*
2. *задача поиска ассоциативных правил;*
3. *задача нормализации;*

Ключ: 1-1, 2-1, 3-2, 4-4, 5-3, 6-3, 7-2

5.2. Примерные темы рефератов

1. Онтологии и онтологические системы. Модели онтологии и онтологической системы.
2. Задача классификации. Методы построения деревьев решения. Методика «разделяй и властвуй».
3. Задача классификации. Методы построения деревьев решений. Алгоритм покрытия.
4. Задача классификации. Методы построения деревьев решений. Алгоритм ID 3.
5. Задача классификации. Методы построения деревьев решений. Алгоритм C4.5.
6. Задача классификации. Методы построения правил классификации. Алгоритм построения 1 – правил.
7. Задача классификации. Методы построения правил классификации. Метод Naive Bayes.
8. Задача кластеризации. Алгоритм k-means.
9. Задача кластеризации. Алгоритм Fuzzy C-Means.
10. Задача кластеризации. Алгоритм EM.
11. Информационный поиск в текстах. Information Retrieval.
12. Поиск ассоциативных правил. Алгоритм Apriori.
13. Секвенциальный анализ. Алгоритм AprioriALL.
14. Секвенциальный анализ. Алгоритм GSP.
15. Методы кластеризации текстовых документов.
16. Задача аннотирования текстов. Методы извлечения фрагментов для аннотации.
17. Преобразование MTF.
18. Алгоритм сжатия BWT.
19. Алгоритм построения 1-правил.
20. Метод Naive-Base.
21. Прогнозирование временных рядов.

5.3. Пример задания на лабораторную работу

Тема. Предобработка данных в Pandas

Задание для самостоятельной работы:

1. Используйте библиотеку Pandas выполните задания по предварительной обработке данных.
2. Анализ данных по доходу населения UCI Adult. В задании предлагается с помощью Pandas ответить на несколько вопросов по данным репозитория UCI Adult. Список вопросов:
 - Каков средний возраст (признак age) женщин?
 - Какова доля граждан Германии (признак native-country)?
 - Постройте гистограмму распределения (bar plot) образования людей (признак education).
 - Каковы средние значения и среднеквадратичные отклонения возраста тех, кто получает более 50К в год (признак salary) и тех, кто получает менее 50К в год?
 - Правда ли, что люди, которые получают больше 50k, имеют как минимум высшее образование? (признак education - Bachelors, Prof-school, Assoc-acdm, Assoc-voc, Masters или Doctorate)
 - Выведите статистику возраста для каждой расы (признак race) и каждого пола. Используйте groupby и describe. Найдите таким образом максимальный возраст мужчин расы Amer-Indian-Eskimo.
 - Среди кого больше доля зарабатывающих много (>50К): среди женатых или холостых мужчин (признак marital-status)? Женатыми считаем тех, у кого marital-status начинается с Married (Married-civ-spouse, Married-spouse-absent или Married-AF-spouse), остальных считаем холостыми.
 - Какое максимальное число часов человек работает в неделю (признак hours-per-week)? Сколько людей работают такое количество часов и каков среди них процент зарабатывающих много?
 - Посчитайте среднее время работы (hours-per-week) зарабатывающих мало и много (salary) для каждой страны (native-country).
 - Анализ данных по пассажирам Титаника. В задании предлагается с помощью Pandas ответить на несколько вопросов по данным репозитория UCI Titanic. Список вопросов:
 - Какое количество мужчин и женщин ехало на корабле?
 - Какой части пассажиров удалось выжить? Посчитайте долю выживших пассажиров.
 - Какую долю пассажиры первого класса составляли среди всех пассажиров?
 - Какого возраста были пассажиры?
 - Посчитайте среднее и медиану возраста пассажиров.
 - Коррелируют ли число братьев/сестер с числом родителей/детей? Посчитайте корреляцию Пирсона между признаками SibSp и Parch.
 - Какое самое популярное женское имя на корабле? Извлеките из полного имени пассажира (колонок Name) его личное имя (First Name).

5.4. Вопросы к зачету

1. Основные понятия – информация, данные, знания. Виды информации. Обработка данных и ее виды. Data Mining. Классификация задач Data Mining.
2. Модели процессов обработки данных. Модель: конечные автоматы.
3. Модели процессов обработки данных. Модель: сети Петри.
4. Задачи обработки данных различных типов. Прикладные области обработки данных. Оцифровка сигналов. Теорема Котельникова.
5. Базы данных. OLTP – системы. Неэффективность OLTP для анализа данных. Определение и свойства хранилищ данных.
6. Физические и виртуальные хранилища данных (ХД). Основные проблемы создания ХД.

7. Витрины данных.
8. Данные в хранилищах данных. ETL процесс.
9. Представление данных в виде гиперкуба. Операции над гиперкубом. Пример. Технология OLAP. Тест FASMI.
10. Многомерное представление данных и многомерный куб. Представление данных в виде гиперкуба. Пример.
11. Основные понятия гиперкубов (OLAP кубов). Структура OLAP куба. Операции над гиперкубом.
12. Архитектура OLAP. Компоненты OLAP. MOLAP, ROLAP, HOLAP.
13. Задача анализа текстов. Этапы анализа. Предобработка текста.
14. Извлечение ключевых понятий из текста.
15. Классификация текстовых документов. Методы классификации текстовых документов.
16. Большие данные. Свойства больших данных.
17. Машинное обучение, формализация задачи машинного обучения.
18. Признаковое описание объекта. Ответы и типы задач машинного обучения. Модель алгоритмов. Метод обучения. Этап обучения и этап применения.
19. Функционалы качества. Сведение задачи обучения к задаче оптимизации.
20. Переобучение и обобщение. Пример переобучения (Рунге). Эмпирические оценки обобщающей способности.
21. Примеры задач машинного обучения: задачи классификации.
22. Примеры задач машинного обучения: задачи регрессии.
23. Примеры задач машинного обучения: задача ранжирования.
24. Эксперименты в машинном обучении: эксперименты на реальных и синтетических данных.
25. Формализация метрической классификации. Обобщенный метрический классификатор.
26. Метод ближайшего соседа.
27. Метод k взвешенных ближайших соседей.
28. Метод парзеновского окна.
29. Метод потенциальных функций.
30. Отбор эталонных объектов. Понятие отступа объекта. Типы объектов в зависимости от отступа.
31. Отбор эталонов, алгоритм STOLP.
32. Логическая закономерность. Основы вопросы построения логических алгоритмов классификации. Виды закономерностей.
33. Критерии информативности: простые критерии, статистический критерий, энтропийный критерий. Схема локального поиска информативных закономерностей.
34. Определение бинарного решающего дерева. Жадный алгоритм построения дерева ID 3.
35. Варианты критериев ветвления в ID 3.
36. Алгоритм ID3: достоинства и недостатки.
37. Обработка пропусков в ID 3, алгоритм обработки пропусков на этапе обучения и этапе классификации.
38. Стратегии редукции решающих деревьев.
39. Небрежные решающие деревья.
40. Бинаризация вещественного признака.

ТЕХНОЛОГИЧЕСКАЯ КАРТА ДИСЦИПЛИНЫ
ОСНОВНАЯ ОБРАЗОВАТЕЛЬНАЯ ПРОГРАММА
09.03.02 Информационные системы и технологии
Направленность (профиль) «Программно-аппаратные комплексы»
(код, направление, профиль)

ТЕХНОЛОГИЧЕСКАЯ КАРТА

Шифр дисциплины по РУП	Б1.В.ДВ.4.1				
Дисциплина	Технологии обработки информации				
Курс	4-5	семестр	8-9		
Кафедра	Информатики и вычислительной техники				
Ф.И.О. преподавателя, звание, должность	Тоичкин Николай Александрович, канд. техн. наук, доцент кафедры информатики и вычислительной техники				
Общ. трудоемкость ^{час/ЗЕТ}	72/2	Кол-во семестров	2	Форма контроля	Зачет
ЛК ^{общ./тек. сем.}	4/4	ПР/СМ ^{общ./тек. сем.}	-/-	ЛБ ^{общ./тек. сем.}	10/10
				СРС ^{общ./тек. сем.}	54/54

Компетенции обучающегося, формируемые в результате освоения дисциплины:

- способность разрабатывать, согласовывать и выпускать все виды проектной документации (ПК-10);
- способность оптимизировать функционирование баз данных (ПСК-1).

Код формируемой компетенции	Содержание задания	Количество мероприятий	Максимальное количество баллов	Срок предоставления
<i>Вводный блок</i>				
Не предусмотрен				
<i>Основной блок</i>				
ПК-10 ПСК-1	Тест	3	9	В межсессионный период
ПК-10 ПСК-1	Лабораторные работы	4	40	В межсессионный период
ПК-10 ПСК-1	Реферат	1	11	В межсессионный период
Итого:			60	
ПК-10 ПСК-1	Зачет	Вопрос 1	20	В сроки сессии
		Вопрос 2	20	
Всего			40	
Итого:			100	
<i>Дополнительный блок</i>				
ПК-10 ПСК-1	Подготовка опорного конспекта		15	В межсессионный период
Всего:			15	

Шкала оценивания в рамках балльно-рейтинговой системы МАГУ: «2» - 60 баллов и менее, «3» - 61-80 баллов, «4» - 81-90 баллов, «5» - 91-100 баллов.