

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«ТЮМЕНСКИЙ ИНДУСТРИАЛЬНЫЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ

Заведующий кафедрой

_____ Данилов О. Ф.

«_____» _____ 20__ г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

дисциплины: **Анализ данных и машинное обучение**

направление подготовки: **09.03.04 Программная инженерия**

направленность: **Разработка программно-информационных систем**

форма обучения: **очная**

Фонд оценочных средств рассмотрен на заседании кафедры интеллектуальных систем и технологий для направления 09.03.04 Программная инженерия направленность (профиль) «Разработка программно-информационных систем»

1. Формы аттестации по дисциплине

1.1. Форма промежуточной аттестации: зачет – 6 семестр

Способ проведения промежуточной аттестации: устный опрос

1.2. Формы текущей аттестации:

Таблица 2.1

№ п/п	Форма обучения
	ОФО
1	Защита практического задания
2	Устный опрос

2. Результаты обучения по дисциплине, подлежащие проверке при проведении текущей и промежуточной аттестации

Таблица 2.1

№ п/п	Структура дисциплины		Код ИДК	Оценочные средства	
	Номер раздела	Наименование раздела		Текущая аттестация	Промежуточная аттестация
1	1	Введение в курс	31-34, У1-У3, В1-В3	Опрос Защита практического задания	Вопросы к зачету
2	2	Этапы жизненного цикла решения	31-34, У1-У3, В1-В3	Защита практического задания	Вопросы к зачету
3	3	Подготовка данных	31-34, У1-У3, В1-В3	Опрос Защита практического задания	Вопросы к зачету
4	4	Информационное обучение	31-34, У1-У3, В1-В3	Защита практического задания	Вопросы к зачету
5	5	Обучение на основе сходства	31-34, У1-У3, В1-В3	Опрос Защита практического задания	Вопросы к зачету
6	6	Вероятностное обучение	31-34, У1-У3, В1-В3	Защита практического задания	Вопросы к зачету
7	7	Обучение на основе ошибок	31-34, У1-У3, В1-В3	Опрос Защита практического задания	Вопросы к зачету
8	8	Оценка моделей	31-34, У1-У3, В1-В3	Опрос Защита практического задания	Вопросы к зачету
9	9	Перспективы	31-34, У1-У3, В1-В3	Защита практического задания	Вопросы к зачету

3. Фонд оценочных средств

3.1. Фонд оценочных средств, позволяющие оценить результаты обучения по дисциплине, включает в себя оценочные средства для текущей аттестации и промежуточной аттестации.

3.2. Фонд оценочных средств для текущей аттестации включает:

- задания для практических работ – 9 шт. (Приложение 1);
- вопросы для подготовки к устным опросам – 25 шт. (Приложение 2)

3.3. Фонд оценочных средств для промежуточной аттестации включает:

- комплект вопросов к зачету по дисциплине – 13 шт. (Приложение 3).

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«ТЮМЕНСКИЙ ИНДУСТРИАЛЬНЫЙ УНИВЕРСИТЕТ»

Задания для практических работ**Работа № 1.** Введение в курс

Задание. Дать определение понятиям «Анализ данных» и «Машинное обучение». Рассмотреть нормативно-правовые акты, применяемые в данной области. Представить схему методов машинного обучения, применяемых в настоящее время.

Контрольные вопросы.

1. Дайте определение понятия «Анализ данных»
2. Дайте определение понятия «Машинное обучение»
3. Какие нормативно-правовые акты могут быть применены в области анализа данных и машинного обучения.
4. Какие методы машинного обучения, применяемые при анализе данных вы можете назвать.

Работа № 2 Этапы жизненного цикла решения.

Задание. Составить основной pipeline машинного обучения и dataflow. Указать особенности архитектурных решений. Указать облачные платформы

Контрольные вопросы.

1. Какие основные операции должны быть выполнены в процессе машинного обучения?
2. Существуют ли отличия pipeline машинного обучения и data flow?
3. Какие особенности архитектурных решений pipeline машинного обучения вы можете назвать?
4. Какие существуют платформы для скачивания датасетов?

Работа № 3. Подготовка данных

Задание. Предложить методы анализа выбросов, учитывая особенности данных. Сделать анализ выбросов и удалить их. Проанализировать матрицу корреляций оценок по различным критериям. Выявить значимые корреляции. Проанализировать описательные статистики. Сравнить результаты и производительность вычислений в различных системах.

Контрольные вопросы.

1. Какие способы трансформации данных вы знаете?
2. Какие показатели входят в описательную статистику?
3. Какие инструменты анализа данных вы знаете?
4. Назовите аномалии данных.

Работа №4. Информационное обучение

Задание. Изучить особенности анализа главных компонент. Рассмотреть набор данных, выданный преподавателем. Описать исследуемые данные. Выполнить анализ главных компонент для одного вида данных, затем для нескольких и сравнить их между собой. Выполнить анализ главных компонент для всего набора данных. Выделите главные факторы, дайте интерпретацию.

Контрольные вопросы.

1. Что такое анализ главных компонент? Какова его цель?
2. Что такое singular value decomposition?
3. Что такое latent semantic analysis?

Работа №5. Обучение на основе сходства

Задание. Отличить фальшивые банкноты от настоящих. Проанализировать результаты для набора данных Banknote Authentication, в чем разница базовых настроек алгоритма в разных инструментах? Найти наилучшие параметры для данного датасета, используя технику кросс-валидации.

Контрольные вопросы.

1. Какие способы решения данной задачи вы можете предложить?
2. Опишите сущность метода классификации SVM.
3. Назовите положительные и отрицательные стороны метода SVM.

Работа 6. Вероятностное обучение

Задание. Имеется коллекция новостных статей и множество рубрик новостей. Необходимо построить модель, которая для заданной статьи определяет к какой рубрике она относится.

Контрольные вопросы.

1. Что такое Байесовский классификатор?
2. Опишите сущность метода векторного представления текстов.
3. Что такое наивный Байес?
4. Что такое multinomial Naïve Bayes?

Работа 7. Обучение на основе ошибок

Задание. Решить задачу прогнозирования цен на рынке недвижимости. Оценить функцию потерь.

Контрольные вопросы.

1. С помощью какого метода машинного обучения была решена эта задача?
2. Опишите сущность применяемого в решении метода машинного обучения.
3. Как найти коэффициенты регрессии?
4. Как оценить качество регрессионной модели?

Работа 8. Оценка моделей

Задание. Для моделей, полученных в лабораторных работах 5-7, рассчитать показатели их качества.

Контрольные вопросы.

1. Какие показатели качества используются для оценки адекватности модели, полученной методом классификации? Опишите их смысл.
2. Какие показатели качества используются для оценки адекватности модели, полученной методом регрессии? Опишите их смысл.
3. Какие ещё показатели качества вы знаете и для чего они используются?

Работа 9. Перспективы

Задание. Опишите перспективы развития машинного обучения и анализа данных.

Контрольные вопросы.

1. Какие методы машинного обучения вы можете выделить как наиболее перспективные?
2. Какие перспективы в анализе данных вы можете выделить?

Критерии оценки

Оценивается весь комплекс лабораторных работ. Максимальное количество баллов за цикл работ – 45.

31-45 баллов выставляется обучающемуся, если он усвоил теоретический материал курса, правильно понял цели и задания работ. Выполнил весь цикл лабораторных работ по дисциплине и предоставил функциональное, корректное решение. Успешно защитил все лабораторные работы.

20-30 баллов выставляется обучающемуся, если он усвоил теоретический материал курса, правильно понял цели и задания работ. Выполнил весь цикл лабораторных работ по дисциплине. Функциональность и корректность результатов имеет небольшие замечания.

Защитил свои разработки, по существу отвечая на вопросы как практического, так и теоретического содержания, допускает 1-2 негрубых ошибки при ответе.

11-19 баллов выставляется обучающемуся, если он усвоил теоретический материал курса, правильно понял цели и задания работ. Выполнил часть (более 60%) лабораторных работ по дисциплине. Функциональность и корректность решения содержит ошибки. Защитил свою разработку, отвечая на вопросы как практического, так и теоретического содержания, допускает грубые ошибки при ответе.

До 11 баллов выставляется обучающемуся, если он выполнил часть (менее 60%) лабораторных работ по дисциплине. Функциональность и корректность решения содержит грубые ошибки. Разработал не все документы или разработал с ошибками. Защитил свою разработку, допуская грубые ошибки при ответе.

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«ТЮМЕНСКИЙ ИНДУСТРИАЛЬНЫЙ УНИВЕРСИТЕТ»

Вопросы для подготовки к устным опросам

Раздел 1. «Введение»

1. Что такое «машинное обучение»?
2. Что такое «анализ данных»?
3. Назовите основные виды машинного обучения?
4. Назовите примеры методов по каждому виду машинного обучения?
5. Опишите особенности применения каждого вида машинного обучения.

Раздел 3. «Подготовка данных».

1. Какие показатели входят в описательную статистику?
2. Какие методы выявления выбросов в выборке данных вы знаете?
3. Какие способы трансформации данных вы знаете? Опишите их
4. Какие аномалии данных вы знаете?
5. Как устранять различные виды аномалий в данных?

Раздел 5. «Обучение на основе сходства»

1. Метод опорных векторов (SVM).
2. Разделяющая гиперплоскость.
3. Функция ядра.
4. Трюк с ядром (Kernel Trick)
5. Методы снижения размерности. Метод главных компонент. SNE, t-SNE. UMAP.

Раздел 7. «Обучение на основе ошибок»

1. Функции потерь для задач машинного обучения.
2. Градиентный спуск.
3. Линейные модели.
4. Логистическая регрессия.
5. Концепция переобучения и недообучения.

Раздел 8. «Оценка моделей»

1. Функции оценки качества для задач классификации Accuracy, Precision, Recall.
2. Функции оценки качества для задач классификации F1, ROC AUC.
3. Функции оценки качества для задач регрессии: MSE, RMSE.
4. Функции оценки качества для задач регрессии: MAE, MAPE.
5. Методы валидации качества алгоритма.

Критерии оценки

Каждый обучающийся отвечает на 2 вопроса по теме.

Таблица 1

Баллы	0	4 (12,5)	7,5 (25)
Критерии ответа	Отсутствует ответ на вопрос или ответ дан не по теме вопроса	Дан ответ на вопрос с существенными замечаниями. Ответ является неполным.	Дан развернутый ответ на вопрос с незначительными замечаниями или их отсутствием

Суммарное количество баллов за опрос по каждой теме составляет от 0 до 15 баллов за 1 и 2 аттестации и от 0 до 25 баллов за 3 аттестацию.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«ТЮМЕНСКИЙ ИНДУСТРИАЛЬНЫЙ УНИВЕРСИТЕТ»

Вопросы к зачету по дисциплине

1. Типы задач: классификация, регрессия, прогнозирование, ранжирование, обнаружение аномалий. Методы обучения: обучение с учителем, обучение без учителя, обучение с подкреплением.
2. Функции потерь для задач машинного обучения. Градиентный спуск.
3. Функции оценки качества для задач классификации Accuracy, Precision, Recall, F1, ROC AUC. Функции оценки качества для задач регрессии: MSE, RMSE, MAE, MAPE.
4. Линейные модели. Логистическая регрессия. Концепция переобучения и недообучения. Методы валидации качества алгоритма. Регуляризация. L1/L2 регуляризация, множитель Лагранжа.
5. Решающее дерево, бинаризация признаков, алгоритм построения. Решающие деревья, случайный лес, градиентный бустинг. Методы их обучения, критерий информативности, критерий останова. Ансамблирование моделей: мажорантное голосование, блендинг, бустинг, бэггинг.
6. Метод опорных векторов (SVM). Разделяющая гиперплоскость. Функция ядра. Трюк с ядром (Kernel Trick)
7. Байесовский подход в машинном обучении. Наивный байесовский классификатор.
8. Кластеризация. Алгоритмы кластеризации. Метод k-средних. DBSCAN. Иерархическая кластеризация.
9. Методы снижения размерности. Метод главных компонент. SNE, t-SNE. UMAP.
10. Типы признаков и их обработка. Нормализация данных, масштабирование, обработка категориальных признаков. Векторизация текстовых данных с помощью bag-of-words, tf-idf.
11. Методы оптимизации. Градиентный спуск, SGD, AdaGrad, Adam, RMSProp, момент Нестерова.
12. Методы работы с временными рядами. Модель ARIMA.
13. Статистические методы интерпретации моделей машинного обучения. SHAP, LIME.

Критерии оценки

91-100 (зачтено) баллов выставляется обучающемуся, если он системно, глубоко и прочно усвоил программный материал курса; полно, логически стройно, четко и правильно его излагает. Обучающийся не затрудняется с ответами на вопросы при их видоизменении. Возможно 1-2 негрубых недочета в ответах.

76-90 (зачтено) баллов выставляется обучающемуся, если он в основном правильно, по существу излагает материал, но несколько нарушена логика и последовательность повествования, допускает 1-2 негрубых ошибки при ответе; если ответ не совсем полный, но справляется, возможно, при использовании наводящих вопросов.

61-75 (зачтено) баллов выставляется обучающемуся, если он имеет знания по основным вопросам курса (не менее 50 %), но не усвоил деталей, допускает значительные неточности в ответе или недостаточно правильные формулировки, имеют место 3-4 ошибки, нарушена логическая последовательность в изложении программного материала,

Менее 60 (не зачтено) баллов выставляется обучающемуся, если он не знает значительной части программного материала (более 50 %), допускает грубые ошибки, отсутствует логика изложения и системность в построении ответа.

В целях повышения объективности оценки знаний, умений и навыков студентов, преподаватель может задать до 3-х дополнительных вопросов по содержанию программного материала.