**ИТОГОВАЯ АТТЕСТАЦИЯ**
**ПО ПРОГРАММЕ**
**ПРОФЕССИОНАЛЬНОЙ ПЕРЕПОДГОТОВКИ**
«Искусственный интеллект и глубокое обучение»

Итоговая аттестация слушателей проводится в формате демонстрационного экзамена с участием представителей профильных организаций работодателей. На демонстрационном экзамене слушатели решают практико-ориентированные задачи в сфере искусственного интеллекта. При разработке практико-ориентированных задач использовался комплексный подход, обеспечивающий контроль формирования у слушателей профессиональных компетенций (Таблица 3).

## Примеры практико-ориентированных задач для итоговой аттестации слушателей по программе профессиональной переподготовки «Искусственный интеллект и глубокое обучение»

### Общая формулировка практико-ориентированной задачи:

Решение необходимо предоставить в виде Jupyter notebook, выложить на Git и по окончании программы представить в виде презентации итоговый анализ с выводами. План анализа данных состоит из следующих шагов:

1. Напишите название работы и кратко опишите суть решаемой задачи.
2. Загрузите данные и опишите, что они из себя представляют, их особенности при их наличии.
3. Продемонстрируйте пример данных.
4. Выберите алгоритмы и методы, которые будут решать вашу задачу, обоснуйте этот выбор.
5. Выберите метрики, поиск оптимальных значений которых будет осуществляться, обоснуйте этот выбор.
6. Осуществите подготовку данных для работы с выбранными инструментами, при необходимости осуществляя фильтрацию, обоснуйте данную предобработку.
7. Разбейте данные на обучающую и тестовую выборки.
8. Обучите выбранные методы, получив оптимальные значения метрик, опишите как получилось достичь данных показателей, обоснуйте оптимальность решения.
9. Оцените качество полученных результатов.
10. Сделайте общие выводы по решению поставленной задачи.

### Описание кейсов для практико-ориентированных задач

1. Оценка стоимости товаров по различным параметрам (машинное обучение, задача регрессии).

   1.1. Задача: построить модель регрессии стоимости товара по различным параметрам товара. Ранжировать параметры товаров по внесенному вкладу в стоимость товара.

   1.2. Датасеты:

      1.2.1. Цена автомобилей BMW при перепродаже (входные параметры: данные об автомобиле и его использовании; выходной параметр: стоимость автомобиля): https://www.kaggle.com/datasets/danielkyrka/bmw-pricing-challenge
      1.2.2. Цены на ноутбуки (входные параметры: данные о ноутбуке; выходной параметр: стоимость ноутбука): https://www.kaggle.com/datasets/huzdaria/laptop-pricing/data
      1.2.3. Цены на аренду недвижимости (входные параметры: данные о недвижимости; выходной параметр: стоимость недвижимости): https://www.kaggle.com/datasets/ivanchvez/ny-rental-properties-pricing

1.3. Библиотеки для решения: scikit-learn, catboost, lightgbm, xgboost.

1.4. Рекомендуемые модели: линейная регрессия, полиномиальная регрессия, kNN, SVR, дерево решений, RF, Adaboost, GBT, CatBoost, LightGBM, XGBoost.

1.5. Рекомендации для анализа данных: используйте различные метрики качества, такие как MAPE, R^2, MSE, MAE.

2. Распределение клиентов на группы (задача кластеризации).

2.1. Задача: разбейте объекты в выборке на несколько кластеров, найдите оптимальное количество кластеров. Визуализируйте полученные кластера. Выявите наиболее значимые признаки для полученной кластеризации. Обобщите полученное разбиение в виде правил и иерархии.

2.2. Датасеты:

2.2.1. Данные по клиентам продуктовой фирмы (входные параметры: данные о покупателях; выходной параметр: группы покупателей): https://www.kaggle.com/code/karnikakapoor/customer-segmentation-clustering

2.2.2. Данные по кредитным картам (входные параметры: данные о тратах с кредитных карт; выходной параметр: группы клиентов кредитных карт): https://www.kaggle.com/datasets/arjunbhasin2013/ccdata

2.2.3. Данные по онлайн-продажам (входные параметры: данные о покупках клиентов; выходной параметр: группы клиентов): https://www.kaggle.com/datasets/carrie1/ecommerce-data

2.3. Библиотеки для решения: scikit-learn, scipy.

2.4. Рекомендуемые модели: KMeans, иерархическая кластеризация, AffinityPropagation, MeanShift, SpectralClustering, Ward, DBSCAN.

2.5. Рекомендации для анализа данных: используйте различные метрики качества кластеризации (внешние и внутренние меры оценки качества), такие как Minkowski Score, Purity, F-мера, компактность кластеров, Silhouete.

3. Определение болезней по медицинским данным (задача классификации).

3.1. Задача: постройте классификатор заболевания на основе предложенных данных, выделите наиболее важные признаки для классификации, визуализируйте боксплоты и swarmplot с результатами работы для отдельных признаков. Проведите классический анализ тестирования гипотез отличия групп по отдельным признакам.

3.2. Датасеты:

3.2.1. Данные по гепатиту С (входные параметры: данные о пациенте и лабораторные пробы; выходной параметр: диагноз по гепатиту С): https://www.kaggle.com/datasets/fedesoriano/hepatitis-c-dataset

3.2.2. Данные по ожирению (входные параметры: данные о пациенте; выходной параметр: диагноз по ожирению): https://www.kaggle.com/datasets/fedesoriano/body-fat-prediction-dataset

3.2.3. Данные по циррозу печени (входные параметры: данные о пациенте и лабораторные пробы; выходной параметр: диагноз по циррозу печени): https://www.kaggle.com/datasets/fedesoriano/cirrhosis-prediction-dataset

3.3. Библиотеки для решения: scikit-learn, catboost, lightgbm, xgboost.

3.4. Рекомендуемые модели: kNN, SVM, дерево решений, RF, Adaboost, GBT, CatBoost, LightGBM, XGBoost, ET.

3.5. Рекомендации для анализа данных: используйте различные метрики качества, такие как Accuracy, Precision, Recall, F-мера.

4. Климат (машинное обучение, задача регрессии)

4.1. Задача: визуализировать месячную температуру/осадки по годам, построить регрессионную модель глобального потепления, оценить устойчивость модели к выбросам и оценить достоверность модели.

4.2. Датасеты:
- 4.2.1. Данные суточных временных рядов средней температуры, влажности, скорости ветра, среднего давления (входные параметры: данные о погоде в прошлом; выходной параметр: данные о погоде в будущем): https://www.kaggle.com/datasets/sumanthvrao/daily-climate-time-series-data
- 4.2.2. Данные о температуре поверхности Земли (входные параметры: данные о температуре поверхности Земли прошлом; выходной параметр: данные о температуре поверхности Земли в будущем): https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data

4.3. Библиотеки для решения: scikit-learn, catboost, lightgbm, xgboost

4.4. Рекомендуемые модели: линейная регрессия, полиномиальная регрессия, SVR, дерево решений, RF, Adaboost, GBT, CatBoost, LightGBM, XGBoost, ElasticNet.

4.5. Рекомендации для анализа данных: используйте различные метрики качества, такие как MAPE, R^2, MSE, MAE.

5. Демография (машинное обучение, задача кластеризации)

5.1. Задача: визуализация, прогнозирование изменения численности населения, кластеризация стран по группам

5.2. Датасеты:
- 5.2.1. Данные о российской демографии 1990-2017 гг. (входные параметры: данные о показателях демографии в прошлом; выходной параметр: группы регионов по демографии) https://www.kaggle.com/datasets/dwdkills/russian-demography
- 5.2.2. Набор данных по странам 2020 г. (входные параметры: данные о жизни в различных странах; выходной параметр: группы стран по качеству жизни) https://www.kaggle.com/datasets/dumbgeek/countries-dataset-2020
- 5.2.3. Гендерная статистика (входные параметры: данные о жизни в различных странах; выходной параметр: группы стран по качеству жизни) https://www.kaggle.com/datasets/salehahmedrony/gender-statistics

5.3. Библиотеки для решения: scikit-learn, scipy.

5.4. Рекомендуемые модели: KMeans, иерархическая кластеризация, AffinityPropagation, MeanShift, SpectralClustering, Ward, DBSCAN.

5.5. Рекомендации для анализа данных: используйте различные метрики качества кластеризации (внешние и внутренние меры оценки качества), такие как Minkowski Score, Purity, F-мера, компактность кластеров.

6. Прогнозирование заболеваемости на основе статистических данных (машинное обучение, задача регрессии)

6.1. Задача: построить модель регрессии различных заболеваний по данным о пациенте. Ранжировать данные о пациенте по влиянию на итоговый диагноз.

6.2. Датасеты:
6.2.1. Сердечные заболевания (входные параметры: данные о пациенте; выходной параметр: диагноз):
https://www.kaggle.com/datasets/mirzahasnine/heart-disease-dataset
6.2.2. Диабет (входные параметры: данные о пациенте; выходной параметр: диагноз):
https://www.kaggle.com/datasets/mahdiehhajian/diabetes-prevalence
6.2.3. Рак лёгких (входные параметры: данные о пациенте; выходной параметр: диагноз):
https://www.kaggle.com/datasets/thedevastator/cancer-patients-and-air-pollution-a-new-link

6.3. Библиотеки для решения: scikit-learn, catboost, lightgbm, xgboost

6.4. Рекомендуемые модели: линейная регрессия, полиномиальная регрессия, SVR, дерево решений, RF, Adaboost, GBT, CatBoost, LightGBM, XGBoost, ElasticNet.

6.5. Рекомендации для анализа данных: используйте различные метрики качества, такие как MAPE, R^2, MSE, MAE.

7. Определение сегмента товара по различным параметрам (машинное обучение, задача классификации)

7.1. Задача: постройте классификатор сегмента товаров на основе предложенных данных, выделите наиболее важные признаки для классификации, визуализируйте боксплоты и swarmplot с результатами работы для отдельных признаков. Проведите классический анализ тестирования гипотез отличия групп по отдельным признакам.

7.2. Датасеты:
7.2.1. Электроника с Amazon (входные параметры: название электронного продукта, категория, к которой принадлежит продукт, цена со скидкой, рекомендуемая цена производителя; выходной параметр: сегмент товара):
https://www.kaggle.com/datasets/akeshkumarhp/electronics-products-amazon-10k-items
7.2.2. Еда и аллергены в ней (входные параметры: название продукта питания, ингредиенты, присутствующие в продукте питания, сгруппированные по различным столбцам, аллергены, связанные с продуктом питания; выходной параметр: сегмент продукта питания):
https://www.kaggle.com/datasets/uom190346a/food-ingredients-and-allergens
7.2.3. Фастфуд (входные параметры: данные о еде, выходной параметр: сегмент продукта питания):
https://www.kaggle.com/datasets/ulrikthygepedersen/fastfood-nutrition

7.3. Библиотеки для решения: scikit-learn, catboost, lightgbm, xgboost.

7.4. Рекомендуемые модели: kNN, SVM, дерево решений, RF, Adaboost, GBT, CatBoost, LightGBM, XGBoost, ET.

7.5. Рекомендации для анализа данных: используйте различные метрики качества, такие как Accuracy, Precision, Recall, F-мера.

8. Распознавание рукописных символов (задача снижения размерности)

8.1. Задача: визуализировать пространство признаков с использованием методов снижения размерности для данных изображений рукописных символов. Определить оптимальное число компонент в уменьшенном пространстве. Оценить качество

разбиения групп в полученном пространстве признаков с использованием кластеризации.

8.2. Датасеты:
  8.2.1. Изображения нарисованных от руки цифр (входные параметры: изображения цифр, выходной параметр: цифра): https://www.kaggle.com/competitions/digit-recognizer/data
  8.2.2. Изображения рукописных русских букв (входные параметры: изображения букв, выходной параметр: буква): https://www.kaggle.com/datasets/olgabelitskaya/classification-of-handwritten-letters

8.3. Библиотеки для решения: numpy, scikit-learn, scipy, opencv.

8.4. Рекомендуемые модели: tSNE, PCA, UMAP

8.5. Рекомендации для анализа данных: используйте различные метрики качества

9. Социология (восстановление неизвестных параметров распределения)

9.1. Задача: используя предложенные данные определите класс распределений, наиболее точно подходящий для рассматриваемой выборки, определите параметры этого распределения и качество фитирования. Визуализируйте данные

9.2. Датасеты:
  9.2.1. Вступительные экзамены в Индонезии (входные параметры: название экзамена, место сдачи, тип экзамена, , выходной параметр: получение параметров распределения сдачи экзамена): https://www.kaggle.com/datasets/ekojsalim/indonesia-college-entrance-examination-utbk-2019
  9.2.2. Данные о миграции (входные параметры: данные о миграции, выходной параметр: получение параметров распределения миграции): https://www.kaggle.com/datasets/sukhmandeepsinghbrar/migration-data
  9.2.3. Мировая популяция (входные параметры: данные о населении, выходной параметр: получение параметров распределения населения): https://www.kaggle.com/datasets/allanwandia/world-population

9.3. Библиотеки для решения: numpy, scikit-learn, scipy.

9.4. Рекомендуемые модели: набор классических распределений

9.5. Рекомендации для анализа данных:

10. Экономика (машинное обучение, задача кластеризации)

10.1. Задача: разбейте объекты в выборке на несколько кластеров, найдите оптимальное количество кластеров. Визуализируйте полученные кластера. Выявите наиболее значимые признаки для полученной кластеризации. Обобщите полученное разбиение в виде правил и иерархии.

10.2. Датасеты:
  10.2.1. Банковский маркетинг (входные параметры: данные о клиенте, выходной параметр: группы клиентов): https://www.kaggle.com/datasets/berkayalan/bank-marketing-data-set/data
  10.2.2. Индикаторы глобальной экономики (входные параметры: данные об экономике стран, выходной параметр: группы стран): https://www.kaggle.com/datasets/prasad22/global-economy-indicators
  10.2.3. Обзор зарплат в Аргентине (входные параметры: данные о зарплатах, выходной параметр: группы трудящихся):

https://www.kaggle.com/datasets/aletbm/argentina-salary-survey-sysarmy

10.3. Библиотеки для решения: scikit-learn, scipy.

10.4. Рекомендуемые модели: KMeans, иерархическая кластеризация, AffinityPropagation, MeanShift, SpectralClustering, Ward, DBSCAN.

10.5. Рекомендации для анализа данных: используйте различные метрики качества, такие как Precision, Recall, FBeta.

Таблица 4

Формы и методы контроля и оценки результатов освоения модулей

| № п/п | Наименование процедуры | Основные показатели оценки | Компетенция | Формы и методы контроля и оценки |
|---|---|---|---|---|
| 1 | Промежуточная аттестация. Модуль 1. Введение в математическую статистику | Владеет навыками статистического анализа данных и исследования вероятностных распределений в табличном процессоре | ПК-170 | Зачет/Устный опрос |
| 2 | Промежуточная аттестация. Модуль 2. Программирование на Python | Владеет базовыми алгоритмами и простейшими структурами данных, а также практикой применения базовых возможностей и библиотек языка для решения прикладных задач. | ПК-36 | Зачет / Лабораторная работа |
| 3 | Промежуточная аттестация. Модуль 3. Основы технологии машинного обучения и искусственного интеллекта | Владеет навыками проведения полного цикла работ по анализу данных от сбора данных до интерпретации полученных результатов и подготовки соответствующих отчетов с помощью искусственного интеллекта. Владеет навыками обеспечения информационной безопасности в сфере искусственного интеллекта | ПК-36 ПК-37 ПК-170 | Дифференцированный зачет / Лабораторная работа |
| 4 | Промежуточная аттестация. Модуль 4. Современные нейронные сети и компьютерное | Знает различные типы глубоких нейросетевых моделей (модели, обучаемые с учителем: полносвязные сети, сверточные нейронные | ПК-37 | Дифференцированный зачет /Лабораторная работа |

| | | | | |
|---|---|---|---|---|
| | зрение | сети, рекуррентные нейронные сети; модели, обучаемые без учителя: автокодировщики, ограниченные машины Больцмана).<br><br>Знает постановки задач обучения глубоких нейросетевых моделей; основные показатели качества решения классических задач компьютерного зрения (классификация изображений, детектирование объектов, семантическая сегментация изображений).<br><br>Знает общую схему решения задач компьютерного зрения с использованием методов глубокого обучения. | | |
| 5 | Итоговая аттестация | Способен решать актуальные задачи данной сферы (сфер) деятельности, требующих применения и (или) технологий машинного обучения и других инструментов работы с данными, относящимися к технологиям искусственного интеллекта. | ПК-36<br>ПК-37<br>ПК-170 | Экзамен/практико-ориентированные задачи |

Критерии оценки промежуточной аттестации - устный опрос

| | |
|---|---|
| Зачтено | Обучающему засчитывается результат ответа при устном опросе, если обучающийся дает развернутый ответ, который представляет собой связное, логически последовательное сообщение на заданную тему, показывает его умение применять определения, правила в конкретных случаях. |
| Не зачтено | Обучающийся обнаруживает незнание большей части соответствующего вопроса, допускает ошибки в формулировке определений и правил, искажающие их смысл, беспорядочно и неуверенно излагает материал. |

Критерии оценки промежуточной аттестации – лабораторная работа (зачет)

| | |
|---|---|
| Зачтено | При выполнении задания выполнены все этапы задачи. Либо при выполнении задания выполнены все этапы алгоритма, но отдельные части и аргументация не уточнены или частично не были представлены. Либо не выполнены все этапы алгоритма, допущены логические ошибки и полученный результат не обоснован. |
| Не зачтено | Слушатель не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки и не умеет применять базовые термины и знания при решении типовых практических задач. |

Критерии оценки промежуточной аттестации – лабораторная работа (дифференцированный зачет)

| | |
|---|---|
| Отлично | При выполнении задания выполнены все этапы задачи. |
| Хорошо | При выполнении задания выполнены все этапы алгоритма, но отдельные части и аргументация не уточнены или частично не были представлены. |
| Удовлетворительно | Не выполнены все этапы алгоритма, допущены логические ошибки и полученный результат не обоснован. |
| Неудовлетворительно | Слушатель не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки и не умеет применять базовые термины и знания при решении типовых практических задач. |

Критерии оценки итоговой аттестации

| | |
|---|---|
| Отлично | Слушатель имеет глубокие знания учебного материала по теме практической задачи, показывает усвоение взаимосвязи основных понятий, используемых в работе, смог ответить на все уточняющие и дополнительные вопросы. Слушатель демонстрирует знания теоретического и практического материала в области искусственного интеллекта и анализа данных, определяет взаимосвязи между показателями задачи, даёт правильный алгоритм решения. |
| Хорошо | Слушатель показал знание учебного материала, смог ответить почти полно на все заданные дополнительные и уточняющие вопросы. Слушатель демонстрирует знания теоретического и практического материала в области искусственного интеллекта и анализа данных, допуская незначительные неточности при решении практической задачи, имея неполное понимание междисциплинарных связей при правильном выборе алгоритма решения задания. |
| Удовлетворительно | Слушатель в целом освоил материал практической работы, ответил не на все уточняющие и дополнительные вопросы. Слушатель затрудняется с правильным выполнением предложенной задачи, даёт неполный ответ, требующий наводящих вопросов преподавателя, выбор алгоритма решения задачи возможен при наводящих вопросах преподавателя. |
| Неудовлетворительно | Слушатель имеет существенные пробелы в знаниях основного учебного материала практической работы, который полностью не раскрыл содержание вопросов, не смог ответить на уточняющие и дополнительные вопросы. Слушатель даёт неверную оценку ситуации, неправильно выбирает алгоритм действий. |