МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ГЕОДЕЗИИ И КАРТОГРАФИИ
(МИИГАиК)

Факультет геоинформатики и информационной безопасности
Кафедра информатики и геоинформационных технологий

**ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ**

Сертификат: 5BFEE1D286E9C95A0C0D2C611E434BD5
Владелец: Степанченко Алексей Леонидович
Действителен: с 23.09.2026 по 17.12.2027

«УТВЕРЖДАЮ»
Проректор
А. Л. Степанченко
18 апреля 2025 г.

# ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ (МОДУЛЮ) «Б1.О.11 ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ»

Уровень высшего образования: магистратура

Направление подготовки: 09.04.03 Прикладная информатика

Направленность (профиль) подготовки: Анализ больших пространственных данных

Квалификация (степень) выпускника: магистр

Форма обучения: очная

Год набора (приема на обучение): 2025

Срок получения образования: 2 года

Объем: в зачетных единицах: 4 з.е.
в академических часах: 144 ак.ч.

Москва, 2025

**Разработчики:**
Старший преподаватель кафедры Информатики и
геоинформационных технологий Стрельцов А. С.

Фонд оценочных средств по дисциплине (модулю) составлен в соответствии с требованиями ФГОС ВО по направлению подготовки 09.04.03 Прикладная информатика, утвержденного приказом Минобрнауки от 19.09.2017 № 916

Согласование и утверждение

| № | Подразделение | Ответственное лицо | ФИО | Виза | Дата, протокол (при наличии) |
|---|---|---|---|---|---|
| 1 | факультет Геоинформатики и информационной безопасности | Председатель ученого совета факультета | Матерухин А. В. | Рассмотрено | 14.10.2024, № 10-24 |

## 1. Планируемые результаты обучения, соотнесенные с планируемыми результатами освоения образовательной программы

ОПК-2 Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач

ОПК-2.3 Разрабатывает программные средства на основе методов интеллектуального анализа данных для решения профессиональных задач

*Знать:*

ОПК-2.3/Зн1 Теоретические основы и математический аппарат методов интеллектуального анализа данных (Data Mining), включая алгоритмы классификации, кластеризации, регрессионного анализа, ассоциативных правил, а также принципы их программной реализации с использованием библиотек Python (NumPy, Pandas, Scikit-learn).

ОПК-2.3/Зн2 Методы предобработки, визуализации и исследовательского анализа данных (EDA), метрики оценки качества моделей машинного обучения, техники валидации и оптимизации гиперпараметров для создания эффективных программных решений.

ОПК-2.3/Зн3 Архитектуру и принципы построения программных систем интеллектуального анализа данных, включая конвейеры обработки данных (ML pipelines), методы развертывания моделей и создания пользовательских интерфейсов.

*Уметь:*

ОПК-2.3/Ум1 Разрабатывать программные модули для решения задач интеллектуального анализа данных с использованием алгоритмов машинного обучения (KNN, деревья решений, случайный лес, SVM, наивный Байес, ансамблевые методы).

ОПК-2.3/Ум2 Проектировать и реализовывать комплексные программные решения для анализа структурированных и неструктурированных данных, включая обработку текстовой информации, временных рядов и пространственных данных.

ОПК-2.3/Ум3 Интегрировать разработанные программные средства интеллектуального анализа данных в существующие информационные системы, создавать API и веб-интерфейсы для взаимодействия с аналитическими моделями.

*Владеть:*

ОПК-2.3/Нв1 Навыками программирования на языке Python и использования специализированных библиотек (NumPy, Pandas, Matplotlib, Plotly, Scikit-learn) для разработки программных средств интеллектуального анализа данных.

ОПК-2.3/Нв2 Методами построения, обучения, валидации и оптимизации моделей машинного обучения, а также навыками оценки их эффективности и интерпретации результатов для решения профессиональных задач.

ОПК-2.3/Нв3 Технологиями развертывания программных решений интеллектуального анализа данных, включая сериализацию моделей, создание REST API с FastAPI, разработку интерактивных веб-приложений с Streamlit.

## 2. Показатели и критерии оценивания компетенций

| Планируемые результаты освоения программы | Уровень овладения | Критерии оценивания (перечень сформированных ИДК/ЗУВ по уровням освоения) |
|---|---|---|

| | | |
|---|---|---|
| ОПК-2 Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач | Повышенный | ИДК ОПК-2.3 Разрабатывает программные средства на основе методов интеллектуального анализа данных для решения профессиональных задач<br><br>Знать:<br>ОПК-2.3/Зн1 Теоретические основы и математический аппарат методов интеллектуального анализа данных (Data Mining), включая алгоритмы классификации, кластеризации, регрессионного анализа, ассоциативных правил, а также принципы их программной реализации с использованием библиотек Python (NumPy, Pandas, Scikit-learn).<br>ОПК-2.3/Зн2 Методы предобработки, визуализации и исследовательского анализа данных (EDA), метрики оценки качества моделей машинного обучения, техники валидации и оптимизации гиперпараметров для создания эффективных программных решений.<br>ОПК-2.3/Зн3 Архитектуру и принципы построения программных систем интеллектуального анализа данных, включая конвейеры обработки данных (ML pipelines), методы развертывания моделей и создания пользовательских интерфейсов.<br>Уметь:<br>ОПК-2.3/Ум1 Разрабатывать программные модули для решения задач интеллектуального анализа данных с использованием алгоритмов машинного обучения (KNN, деревья решений, случайный лес, SVM, наивный Байес, ансамблевые методы).<br>ОПК-2.3/Ум2 Проектировать и реализовывать комплексные программные решения для анализа структурированных и неструктурированных данных, включая обработку текстовой информации, временных рядов и пространственных данных.<br>ОПК-2.3/Ум3 Интегрировать разработанные программные средства интеллектуального анализа данных в существующие информационные системы, создавать API и веб-интерфейсы для взаимодействия с аналитическими моделями.<br>Владеть:<br>ОПК-2.3/Нв1 Навыками программирования на языке Python и использования специализированных библиотек (NumPy, Pandas, Matplotlib, Plotly, Scikit-learn) для разработки программных средств интеллектуального анализа данных.<br>ОПК-2.3/Нв2 Методами построения, обучения, валидации и оптимизации моделей машинного обучения, а также навыками оценки их эффективности и интерпретации результатов для решения профессиональных задач.<br>ОПК-2.3/Нв3 Технологиями развертывания программных решений интеллектуального анализа данных, включая сериализацию моделей, создание REST API с FastAPI, разработку интерактивных веб-приложений с Streamlit. |

| | | |
|---|---|---|
| | Базовый | ИДК ОПК-2.3 Разрабатывает программные средства на основе методов интеллектуального анализа данных для решения профессиональных задач<br><br>Знать:<br>ОПК-2.3/Зн3 Архитектуру и принципы построения программных систем интеллектуального анализа данных, включая конвейеры обработки данных (ML pipelines), методы развертывания моделей и создания пользовательских интерфейсов.<br>ОПК-2.3/Зн2 Методы предобработки, визуализации и исследовательского анализа данных (EDA), метрики оценки качества моделей машинного обучения, техники валидации и оптимизации гиперпараметров для создания эффективных программных решений.<br>ОПК-2.3/Зн1 Теоретические основы и математический аппарат методов интеллектуального анализа данных (Data Mining), включая алгоритмы классификации, кластеризации, регрессионного анализа, ассоциативных правил, а также принципы их программной реализации с использованием библиотек Python (NumPy, Pandas, Scikit-learn).<br>Уметь:<br>ОПК-2.3/Ум3 Интегрировать разработанные программные средства интеллектуального анализа данных в существующие информационные системы, создавать API и веб-интерфейсы для взаимодействия с аналитическими моделями.<br>ОПК-2.3/Ум2 Проектировать и реализовывать комплексные программные решения для анализа структурированных и неструктурированных данных, включая обработку текстовой информации, временных рядов и пространственных данных.<br>ОПК-2.3/Ум1 Разрабатывать программные модули для решения задач интеллектуального анализа данных с использованием алгоритмов машинного обучения (KNN, деревья решений, случайный лес, SVM, наивный Байес, ансамблевые методы). |

| | | |
|---|---|---|
| | Пороговый | ИДК ОПК-2.3 Разрабатывает программные средства на основе методов интеллектуального анализа данных для решения профессиональных задач<br><br>Знать:<br>ОПК-2.3/Зн1 Теоретические основы и математический аппарат методов интеллектуального анализа данных (Data Mining), включая алгоритмы классификации, кластеризации, регрессионного анализа, ассоциативных правил, а также принципы их программной реализации с использованием библиотек Python (NumPy, Pandas, Scikit-learn).<br>ОПК-2.3/Зн2 Методы предобработки, визуализации и исследовательского анализа данных (EDA), метрики оценки качества моделей машинного обучения, техники валидации и оптимизации гиперпараметров для создания эффективных программных решений.<br>ОПК-2.3/Зн3 Архитектуру и принципы построения программных систем интеллектуального анализа данных, включая конвейеры обработки данных (ML pipelines), методы развертывания моделей и создания пользовательских интерфейсов. |

## 3. Шкала оценивания

| Оценка | Критерии оценивания (перечень сформированных ИДК/ЗУВ) |
|---|---|

| | |
|---|---|
| Отлично | ОПК-2.3 Разрабатывает программные средства на основе методов интеллектуального анализа данных для решения профессиональных задач<br>ОПК-2.3/Зн1 Теоретические основы и математический аппарат методов интеллектуального анализа данных (Data Mining), включая алгоритмы классификации, кластеризации, регрессионного анализа, ассоциативных правил, а также принципы их программной реализации с использованием библиотек Python (NumPy, Pandas, Scikit-learn).<br>ОПК-2.3/Зн2 Методы предобработки, визуализации и исследовательского анализа данных (EDA), метрики оценки качества моделей машинного обучения, техники валидации и оптимизации гиперпараметров для создания эффективных программных решений.<br>ОПК-2.3/Зн3 Архитектуру и принципы построения программных систем интеллектуального анализа данных, включая конвейеры обработки данных (ML pipelines), методы развертывания моделей и создания пользовательских интерфейсов.<br>ОПК-2.3/Ум1 Разрабатывать программные модули для решения задач интеллектуального анализа данных с использованием алгоритмов машинного обучения (KNN, деревья решений, случайный лес, SVM, наивный Байес, ансамблевые методы).<br>ОПК-2.3/Ум2 Проектировать и реализовывать комплексные программные решения для анализа структурированных и неструктурированных данных, включая обработку текстовой информации, временных рядов и пространственных данных.<br>ОПК-2.3/Ум3 Интегрировать разработанные программные средства интеллектуального анализа данных в существующие информационные системы, создавать API и веб-интерфейсы для взаимодействия с аналитическими моделями.<br>ОПК-2.3/Нв1 Навыками программирования на языке Python и использования специализированных библиотек (NumPy, Pandas, Matplotlib, Plotly, Scikit-learn) для разработки программных средств интеллектуального анализа данных.<br>ОПК-2.3/Нв2 Методами построения, обучения, валидации и оптимизации моделей машинного обучения, а также навыками оценки их эффективности и интерпретации результатов для решения профессиональных задач.<br>ОПК-2.3/Нв3 Технологиями развертывания программных решений интеллектуального анализа данных, включая сериализацию моделей, создание REST API с FastAPI, разработку интерактивных веб-приложений с Streamlit. |

| | |
|---|---|
| Хорошо | ОПК-2.3 Разрабатывает программные средства на основе методов интеллектуального анализа данных для решения профессиональных задач<br>ОПК-2.3/Зн1 Теоретические основы и математический аппарат методов интеллектуального анализа данных (Data Mining), включая алгоритмы классификации, кластеризации, регрессионного анализа, ассоциативных правил, а также принципы их программной реализации с использованием библиотек Python (NumPy, Pandas, Scikit-learn).<br>ОПК-2.3/Зн2 Методы предобработки, визуализации и исследовательского анализа данных (EDA), метрики оценки качества моделей машинного обучения, техники валидации и оптимизации гиперпараметров для создания эффективных программных решений.<br>ОПК-2.3/Зн3 Архитектуру и принципы построения программных систем интеллектуального анализа данных, включая конвейеры обработки данных (ML pipelines), методы развертывания моделей и создания пользовательских интерфейсов.<br>ОПК-2.3/Ум1 Разрабатывать программные модули для решения задач интеллектуального анализа данных с использованием алгоритмов машинного обучения (KNN, деревья решений, случайный лес, SVM, наивный Байес, ансамблевые методы).<br>ОПК-2.3/Ум2 Проектировать и реализовывать комплексные программные решения для анализа структурированных и неструктурированных данных, включая обработку текстовой информации, временных рядов и пространственных данных.<br>ОПК-2.3/Ум3 Интегрировать разработанные программные средства интеллектуального анализа данных в существующие информационные системы, создавать API и веб-интерфейсы для взаимодействия с аналитическими моделями. |
| Удовлетворительно | ОПК-2.3 Разрабатывает программные средства на основе методов интеллектуального анализа данных для решения профессиональных задач<br>ОПК-2.3/Зн1 Теоретические основы и математический аппарат методов интеллектуального анализа данных (Data Mining), включая алгоритмы классификации, кластеризации, регрессионного анализа, ассоциативных правил, а также принципы их программной реализации с использованием библиотек Python (NumPy, Pandas, Scikit-learn).<br>ОПК-2.3/Зн2 Методы предобработки, визуализации и исследовательского анализа данных (EDA), метрики оценки качества моделей машинного обучения, техники валидации и оптимизации гиперпараметров для создания эффективных программных решений.<br>ОПК-2.3/Зн3 Архитектуру и принципы построения программных систем интеллектуального анализа данных, включая конвейеры обработки данных (ML pipelines), методы развертывания моделей и создания пользовательских интерфейсов. |
| Неудовле-творительно | Результаты обучения не сформированы на достаточном уровне |

**4. Контрольные мероприятия по дисциплине**

| Вид контроля | Форма контроля/Оценочное средство |
|---|---|
| Текущий контроль | Просмотр выполненных работ |
| Промежуточная аттестация | Экзамен |

| № | Наименование раздела | Контролируем | Вид контроля/ используемые оценочные материалы |
|---|---|---|---|

| п/п | Наименование раздела | ые ИДК | Текущий | Промежут. аттестация |
|---|---|---|---|---|
| 1 | Математические основы анализа данных | ОПК-2.3 | Просмотр выполненных работ | Экзамен |
| 2 | Визуализация и анализ данных | ОПК-2.3 | Просмотр выполненных работ | Экзамен |
| 3 | Основы теории машинного обучения | ОПК-2.3 | | Экзамен |
| 4 | Методы обучения на основе расстояний | ОПК-2.3 | Просмотр выполненных работ | Экзамен |
| 5 | Линейные модели | ОПК-2.3 | Просмотр выполненных работ | Экзамен |
| 6 | Древовидные модели | ОПК-2.3 | Просмотр выполненных работ | Экзамен |
| 7 | Ансамблевые методы | ОПК-2.3 | Просмотр выполненных работ | Экзамен |
| 8 | Вероятностные и дистанционные методы | ОПК-2.3 | Просмотр выполненных работ | Экзамен |
| 9 | Практическое применение машинного обучения | ОПК-2.3 | Просмотр выполненных работ | Экзамен |

## 5. Порядок проведения промежуточной аттестации

Порядок проведения промежуточной аттестации регламентируется локальными нормативными актами Университета:

- Положением о формах, периодичности, порядке текущего контроля успеваемости и промежуточной аттестации обучающихся по основным образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры в федеральном государственном бюджетном образовательном учреждении высшего образования «Московский государственный университет геодезии и картографии», п.6 «Организация и проведение промежуточной аттестации».
- Положением о порядке осуществления образовательного процесса, проведения текущего контроля успеваемости и промежуточной аттестации обучающихся с применением электронного обучения и дистанционных образовательных технологий в федеральном государственном бюджетном образовательном учреждении высшего образования «Московский государственный университет геодезии и картографии», п.8 «Проведение промежуточной аттестации с применением ЭО и ДОТ».

*Промежуточная аттестация: Экзамен, 3 семестр.*

Экзамен по дисциплине проводится в формате тестирования для итоговой оценки знаний.

Студентам будет предложен тест, состоящий из тридцати вопросов закрытого типа с несколькими вариантами ответов. На выполнение всей работы отводится шестьдесят минут, в течение которых запрещается использование любых вспомогательных материалов и электронных устройств.

Итоговая оценка выставляется в зависимости от процента правильных ответов. Для получения оценки «отлично» необходимо верно ответить на 90-100% вопросов (27-30 правильных ответов). Оценка «хорошо» ставится за 70-89% правильных ответов (21-26), а «удовлетворительно» — за 55-69% (17-20). Результат ниже 55% правильных ответов признается неудовлетворительным.

## 6. Оценочные материалы текущего контроля

***Раздел 1. Математические основы анализа данных***
*Контролируемые ИДК: ОПК-2.3*

*Тема 1.1. Библиотека NumPy для численных вычислений*
Форма контроля/оценочное средство: Просмотр выполненных работ

*Тема 1.2. Статистические вычисления в NumPy*
Форма контроля/оценочное средство: Просмотр выполненных работ

***Раздел 2. Визуализация и анализ данных***
*Контролируемые ИДК: ОПК-2.3*

*Тема 2.1. Визуализация данных с Matplotlib*
Форма контроля/оценочное средство: Просмотр выполненных работ

*Тема 2.2. Интерактивная визуализация с Plotly*
Форма контроля/оценочное средство: Просмотр выполненных работ

*Тема 2.3. Работа с табличными данными в Pandas*
Форма контроля/оценочное средство: Просмотр выполненных работ

*Тема 2.4. Предобработка данных с Pandas*
Форма контроля/оценочное средство: Просмотр выполненных работ

***Раздел 4. Методы обучения на основе расстояний***
*Контролируемые ИДК: ОПК-2.3*

*Тема 4.1. Метод k-ближайших соседей (KNN)*
Форма контроля/оценочное средство: Просмотр выполненных работ

***Раздел 5. Линейные модели***
*Контролируемые ИДК: ОПК-2.3*

*Тема 5.1. Линейная регрессия*
Форма контроля/оценочное средство: Просмотр выполненных работ

*Тема 5.2. Логистическая регрессия*
Форма контроля/оценочное средство: Просмотр выполненных работ

***Раздел 6. Древовидные модели***
*Контролируемые ИДК: ОПК-2.3*

*Тема 6.1. Решающие деревья для классификации*
Форма контроля/оценочное средство: Просмотр выполненных работ

*Тема 6.2. Решающие деревья для регрессии*
Форма контроля/оценочное средство: Просмотр выполненных работ

*Тема 6.3. Критерии останова и регуляризация деревьев*
Форма контроля/оценочное средство: Просмотр выполненных работ

***Раздел 7. Ансамблевые методы***
*Контролируемые ИДК: ОПК-2.3*

*Тема 7.1. Случайный лес (Random Forest)*
Форма контроля/оценочное средство: Просмотр выполненных работ

*Тема 7.2. Градиентный бустинг*
Форма контроля/оценочное средство: Просмотр выполненных работ

*Тема 7.3. Стекинг и блендинг моделей*
Форма контроля/оценочное средство: Просмотр выполненных работ

***Раздел 8. Вероятностные и дистанционные методы***
*Контролируемые ИДК: ОПК-2.3*

*Тема 8.1. Наивный байесовский классификатор*
Форма контроля/оценочное средство: Просмотр выполненных работ

*Тема 8.2. Метод опорных векторов (SVM)*
Форма контроля/оценочное средство: Просмотр выполненных работ

***Раздел 9. Практическое применение машинного обучения***
*Контролируемые ИДК: ОПК-2.3*

*Тема 9.1. Классификация текстовых данных*
Форма контроля/оценочное средство: Просмотр выполненных работ

*Тема 9.3. Развертывание ML-моделей*
Форма контроля/оценочное средство: Просмотр выполненных работ

## 7. Оценочные материалы промежуточной аттестации

*3 семестр, Экзамен*
*Контролируемые ИДК: ОПК-2.3*
Вопросы/Задания:

1. Модель классификации показала точность (precision) 0.8 и полноту (recall) 0.6. Рассчитайте значение F1-меры. Ответ округлите до сотых.

2. Как называется явление, при котором модель показывает высокое качество на обучающей выборке, но значительно худшее качество на новых, ранее не встречавшихся данных?

3. Рассчитайте манхэттенское расстояние между точками (1, 2) и (4, 6).

4. Дополните предложение термином:
Правило NumPy, определяющее, каким образом операции могут выполняться поэлементно над массивами разной формы путём их согласованного расширения без физического копирования данных, называется _____ (broadcasting).

5. Дополните предложение термином:

Метод регуляризации линейной регрессии, штрафующий сумму абсолютных значений коэффициентов модели и способный полностью обнулять коэффициенты наименее значимых признаков, называется регуляризацией _____ (L1).

6. Объясните суть компромисса между смещением и разбросом (bias-variance tradeoff) при построении модели машинного обучения.

7. Объясните, каким образом один и тот же алгоритм k-ближайших соседей (KNN) применяется как для задачи классификации, так и для задачи регрессии, и в чём разница итогового правила агрегации ответов соседей в этих двух случаях.

8. Сравните регуляризацию Ridge (L2) и Lasso (L1) линейной регрессии, объясните, почему только Lasso может использоваться для отбора признаков.

9. Сравните предварительную обрезку (pre-pruning) и последующую обрезку (post-pruning) дерева решений как способы борьбы с переобучением.

10. Объясните, какое допущение лежит в основе наивного байесовского классификатора и почему оно называется «наивным».

11. Расположите этапы типового процесса разработки модели машинного обучения в правильной последовательности.

1. Обучение модели и подбор гиперпараметров по валидационной выборке.
2. Анализ исходных данных и выявление пропусков и выбросов.
3. Оценка качества итоговой модели на тестовой выборке.
4. Разделение данных на обучающую, валидационную и тестовую выборки.
5. Предобработка данных (обработка пропусков, кодирование категориальных признаков, масштабирование).

12. Расположите этапы подбора гиперпараметров модели методом Grid Search с использованием кросс-валидации в правильной последовательности.

1. Определение сетки (набора значений) для каждого настраиваемого гиперпараметра.
2. Выбор комбинации гиперпараметров, показавшей наилучшее среднее качество по кросс-валидации.
3. Обучение и оценка модели методом кросс-валидации для каждой возможной комбинации значений гиперпараметров из сетки.
4. Обучение окончательной модели на всей обучающей выборке с выбранной комбинацией гиперпараметров.

13. Расположите этапы построения дерева решений для задачи классификации в правильной последовательности.

1. Вычисление значения критерия разбиения (например, неопределённости Джини) для текущего узла.
2. Рекурсивное повторение процесса для каждого из двух дочерних узлов.
3. Разбиение объектов текущего узла на две группы по выбранному признаку и порогу.
4. Перебор возможных признаков и пороговых значений, дающих наибольшее уменьшение критерия разбиения.

14. Расположите этапы развёртывания обученной модели машинного обучения в виде веб-сервиса в правильной последовательности.

1. Сериализация (сохранение) обученной модели в файл при помощи pickle или joblib.
2. Создание REST API, загружающего сохранённую модель и принимающего запросы с новыми данными для предсказания.
3. Обучение и валидация модели на исторических данных.
4. Развёртывание созданного API на сервере (в том числе в облаке) для приёма запросов из сети.

15. Расположите этапы построения ансамбля моделей методом градиентного бустинга в правильной последовательности.

1. Вычисление остатков (ошибок) текущего ансамбля моделей на обучающей выборке.
2. Инициализация ансамбля простой базовой моделью (например, константным предсказанием).
3. Добавление новой обученной модели к ансамблю с некоторым весом (шагом обучения).
4. Обучение новой базовой модели, предсказывающей вычисленные остатки предыдущего шага.

16. Установите соответствие между метрикой качества модели и её определением.

MSE - Доля объектов, верно предсказанных положительными, среди всех объектов, предсказанных моделью как положительные
Precision (точность) - Средний квадрат отклонений предсказанных числовых значений от фактических
Recall (полнота) - Доля объясненной моделью дисперсии целевой переменной относительно среднего значения
$R^2$ (коэффициент детерминации) - Доля объектов, верно предсказанных положительными, среди всех фактически положительных объектов

17. Установите соответствие между метрикой расстояния и принципом её вычисления.

Евклидово расстояние - Обобщение через параметр степени, включающее евклидово и манхэттенское расстояния как частные случаи
Манхэттенское расстояние - Квадратный корень из суммы квадратов разностей координат точек по каждому измерению
Расстояние Минковского - Сумма модулей разностей координат точек по каждому измерению

18. Установите соответствие между библиотекой Python и её основным назначением в анализе данных.

NumPy - Работа с табличными данными: загрузка, группировка, объединение, обработка пропусков
Pandas - Построение статических графиков с гибкой настройкой осей, легенд и подграфиков
Plotly - Эффективные операции над многомерными числовыми массивами и линейная алгебра
Matplotlib - Построение интерактивных и анимированных графиков и дашбордов

19. Установите соответствие между типом регуляризации линейной регрессии и её особенностью.

Ridge (L2) - Комбинирует штрафы по сумме квадратов и по сумме модулей коэффициентов одновременно
Lasso (L1) - Штрафует сумму квадратов коэффициентов, плавно уменьшая их, но не обнуляя полностью
ElasticNet - Штрафует сумму модулей коэффициентов, способна полностью обнулять коэффициенты малозначимых признаков

20. Установите соответствие между древовидным алгоритмом машинного обучения и его характерным принципом построения.

Одиночное дерево решений - Ансамбль деревьев строится последовательно, каждое новое дерево исправляет ошибки предыдущего ансамбля
Случайный лес - Ансамбль деревьев строится независимо и параллельно, каждое на своей случайной подвыборке данных и признаков
Градиентный бустинг - Единственная модель строится путём рекурсивного разбиения обучающей выборки по значениям признаков