Geekbrains

Специальность «Программист Python Цифровые профессии»

Тема дипломного проекта: «Машинное обучение»

ФИО: Антонова Мария Анатольевна

Место и год написания: Новосибирск 2024

1. Введение

* Обоснование актуальности темы.

* Цель и задачи дипломной работы.

* Объект и предмет исследования.

* Методы исследования.

2. Теоретическая часть

* Обзор существующих методов машинного обучения.

* Анализ достоинств и недостатков различных методов машинного обучения.

* Выбор метода машинного обучения для решения конкретной задачи.

* Описание выбранного метода машинного обучения.

3. Практическая часть

* Сбор и подготовка данных.

* Разработка модели машинного обучения.

* Обучение модели машинного обучения.

* Оценка качества модели машинного обучения.

* Применение модели машинного обучения для решения конкретной задачи.

4. Заключение

* Выводы по результатам дипломной работы.

* Рекомендации по дальнейшему развитию темы дипломной работы.

5. Список литературы

* Перечень использованных источников.

1. Введение

Обоснование актуальности темы.

Актуальность темы машинного обучения

Машинное обучение (МО) является одной из самых быстрорастущих и перспективных областей информатики. МО позволяет компьютерам учиться из данных и делать прогнозы или принимать решения без явного программирования. Это делает МО чрезвычайно полезным для решения широкого спектра задач, включая распознавание образов, обработку естественного языка, прогнозирование спроса и управление рисками.

Почему машинное обучение актуально?

Большой объем данных. В настоящее время мы генерируем огромное количество данных, как структурированных, так и неструктурированных. МО позволяет нам извлекать полезную информацию из этих данных и использовать ее для принятия решений.

Сложность задач. Многие задачи, с которыми мы сталкиваемся сегодня, слишком сложны для того, чтобы их можно было решить вручную. МО позволяет нам автоматизировать эти задачи и получать более точные и быстрые результаты.

Необходимость в персонализации. В современном мире потребители ожидают персонализированного опыта. МО позволяет нам создавать модели, которые могут адаптироваться к индивидуальным потребностям и предпочтениям пользователей.

Рост вычислительной мощности. В последние годы наблюдается значительный рост вычислительной мощности. Это делает МО более доступным и практичным для решения широкого спектра задач.

Области применения машинного обучения

МО используется в самых разных областях, включая:

Распознавание образов: МО используется для распознавания объектов на изображениях и видео. Это используется в таких приложениях, как распознавание лиц, контроль качества и медицинская диагностика.

Обработка естественного языка: МО используется для понимания и генерации естественного языка. Это используется в таких приложениях, как машинный перевод, чат-боты и поиск информации.

Прогнозирование спроса: МО используется для прогнозирования спроса на товары и услуги. Это используется в таких приложениях, как управление запасами, ценообразование и маркетинг.

Управление рисками: МО используется для оценки и управления рисками. Это используется в таких приложениях, как страхование, финансы и здравоохранение.

Перспективы развития машинного обучения

МО является быстро развивающейся областью, и ожидается, что в ближайшие годы оно будет играть все более важную роль в нашей жизни. Некоторые из перспективных направлений развития МО включают:

Разработка новых алгоритмов МО, которые будут более точными и эффективными.

Создание новых приложений МО для решения различных задач в самых разных областях.

Интеграция МО с другими технологиями, такими как Интернет вещей и блокчейн.

Заключение

Машинное обучение является одной из самых актуальных и перспективных областей информатики. МО позволяет компьютерам учиться из данных и делать прогнозы или принимать решения без явного программирования. Это делает МО чрезвычайно полезным для решения широкого спектра задач. МО используется в самых разных областях, и ожидается, что в ближайшие годы оно будет играть все более важную роль в нашей жизни.

Цель и задачи дипломной работы.

Цель дипломной работы по машинному обучению

Цель дипломной работы по машинному обучению заключается в исследовании и применении методов машинного обучения для решения конкретной задачи или проблемы. Это может быть создание новой модели машинного обучения, улучшение существующей модели или применение машинного обучения в новой области.

Задачи дипломной работы по машинному обучению

Для достижения цели дипломной работы по машинному обучению необходимо решить следующие задачи:

Выбор задачи или проблемы для решения. Задача или проблема должна быть четко определена и иметь практическое значение.

Сбор и подготовка данных. Данные должны быть собраны из различных источников и подготовлены для использования в модели машинного обучения.

Выбор и реализация модели машинного обучения. Необходимо выбрать подходящую модель машинного обучения и реализовать ее на выбранном языке программирования.

Обучение и оценка модели машинного обучения. Модель машинного обучения необходимо обучить на собранных данных и оценить ее эффективность.

Анализ и интерпретация результатов. Необходимо проанализировать результаты работы модели машинного обучения и интерпретировать их с точки зрения решаемой задачи или проблемы.

Разработка рекомендаций. На основе результатов работы модели машинного обучения необходимо разработать рекомендации по решению решаемой задачи или проблемы.

Структура дипломной работы по машинному обучению

Дипломная работа по машинному обучению должна иметь следующую структуру:

Введение. Во введении обосновывается актуальность выбранной темы, формулируются цель и задачи дипломной работы.

Обзор литературы. В обзоре литературы анализируются существующие методы машинного обучения, которые могут быть использованы для решения выбранной задачи или проблемы.

Методика исследования. В методике исследования описываются методы и алгоритмы, которые будут использоваться для решения выбранной задачи или проблемы.

Результаты исследования. В результатах исследования приводятся результаты обучения и оценки модели машинного обучения, а также анализ и интерпретация этих результатов.

Выводы и рекомендации. В выводах и рекомендациях подводятся итоги работы, формулируются выводы и рекомендации по решению выбранной задачи или проблемы.

Требования к дипломной работе по машинному обучению

Дипломная работа по машинному обучению должна соответствовать следующим требованиям:

Объем работы. Объем дипломной работы должен составлять не менее 50 страниц.

Оформление работы. Дипломная работа должна быть оформлена в соответствии с требованиями ГОСТ.

Защита работы. Дипломная работа должна быть защищена перед государственной экзаменационной комиссией.

Объект и предмет исследования.

Объект исследования машинного обучения

Объектом исследования машинного обучения являются алгоритмы и модели, которые позволяют компьютерам обучаться на данных и делать прогнозы или принимать решения без явного программирования.

Предмет исследования машинного обучения

Предметом исследования машинного обучения являются:

Методы обучения машин. Это методы, которые позволяют машинам обучаться на данных и улучшать свою производительность с течением времени.

Типы моделей машинного обучения. Существует множество различных типов моделей машинного обучения, каждый из которых имеет свои преимущества и недостатки.

Приложения машинного обучения. Машинное обучение используется во многих областях, включая распознавание образов, обработку естественного языка, прогнозирование и принятие решений.

Основные направления исследований в машинном обучении

Основными направлениями исследований в машинном обучении являются:

Разработка новых методов обучения машин. Это методы, которые позволяют машинам обучаться на данных более эффективно и быстрее.

Разработка новых типов моделей машинного обучения. Это модели, которые способны обрабатывать более сложные данные и задачи.

Расширение областей применения машинного обучения. Это использование машинного обучения для решения новых задач в различных областях.

Актуальность исследований в машинном обучении

Исследования в машинном обучении имеют высокую актуальность, поскольку машинное обучение является одной из ключевых технологий искусственного интеллекта. Машинное обучение используется во многих областях, включая распознавание образов, обработку естественного языка, прогнозирование и принятие решений. Исследования в машинном обучении позволяют разрабатывать новые методы и модели, которые делают машины более интеллектуальными и способными выполнять более сложные задачи.

Значение исследований в машинном обучении

Исследования в машинном обучении имеют большое значение для развития искусственного интеллекта и его применения в различных областях. Машинное обучение позволяет машинам обучаться на данных и улучшать свою производительность с течением времени, что делает их более интеллектуальными и способными выполнять более сложные задачи. Исследования в машинном обучении также позволяют разрабатывать новые методы и модели, которые расширяют области применения машинного

обучения и делают его более доступным для использования в различных областях.

Методы исследования.

Методы исследования машинного обучения можно разделить на две основные категории:

Теоретические методы позволяют анализировать свойства и поведение алгоритмов машинного обучения, не проводя экспериментов на реальных данных. К теоретическим методам относятся:

Анализ сложности позволяет оценить вычислительные ресурсы, необходимые для обучения и использования алгоритма машинного обучения.

Теория вероятностей и статистика используются для анализа статистических свойств алгоритмов машинного обучения и для оценки их производительности.

Оптимизация используется для поиска оптимальных параметров алгоритмов машинного обучения.

Экспериментальные методы позволяют оценивать производительность алгоритмов машинного обучения на реальных данных. К экспериментальным методам относятся:

Оценка производительности позволяет оценить точность и другие показатели производительности алгоритма машинного обучения на заданном наборе данных.

Сравнительное тестирование позволяет сравнить производительность нескольких алгоритмов машинного обучения на заданном наборе данных.

Кросс-валидация используется для оценки производительности алгоритма машинного обучения на разных подмножествах данных.

Тестирование на невидимых данных позволяет оценить производительность алгоритма машинного обучения на данных, которые не использовались для его обучения.
Выбор конкретных методов исследования машинного обучения зависит от целей исследования и имеющихся ресурсов.

Подробное описание методов исследования машинного обучения:

Анализ сложности

Анализ сложности позволяет оценить вычислительные ресурсы, необходимые для обучения и использования алгоритма машинного обучения. Обычно анализ сложности проводится для следующих показателей:

Время обучения - время, необходимое для обучения алгоритма машинного обучения на заданном наборе данных.

Память обучения - объем памяти, необходимый для обучения алгоритма машинного обучения на заданном наборе данных.

Время прогнозирования - время, необходимое для использования алгоритма машинного обучения для прогнозирования значений целевой переменной на новых данных.

Память прогнозирования - объем памяти, необходимый для использования алгоритма машинного обучения для прогнозирования значений целевой переменной на новых данных.

Теория вероятностей и статистика

Теория вероятностей и статистика используются для анализа статистических свойств алгоритмов машинного обучения и для оценки их

производительности. К основным статистическим методам, используемым в машинном обучении, относятся:

Описательная статистика используется для описания основных характеристик данных, таких как среднее значение, медиана, мода, дисперсия и стандартное отклонение.

Выводная статистика используется для проверки гипотез о данных и для оценки параметров моделей машинного обучения.

Математическое ожидание используется для оценки ожидаемого значения целевой переменной при заданных значениях входных переменных.

Дисперсия используется для оценки степени рассеяния значений целевой переменной вокруг математического ожидания.

Ковариация используется для оценки степени зависимости между двумя переменными.

Оптимизация

Оптимизация используется для поиска оптимальных параметров алгоритмов машинного обучения. Оптимальные параметры - это значения параметров, при которых алгоритм машинного обучения достигает наилучшей производительности на заданном наборе данных. Существует множество различных методов оптимизации, которые могут быть использованы для поиска оптимальных параметров алгоритмов машинного обучения. Наиболее распространенными методами оптимизации являются:

Метод градиентного спуска - метод оптимизации, который использует градиент функции цели для поиска ее минимума.

Метод сопряженных градиентов - метод оптимизации, который использует сопряженные градиенты функции цели для поиска ее минимума.

Метод Ньютона - метод оптимизации, который использует вторую производную функции цели для поиска ее минимума.

Метод квази-Ньютона - метод оптимизации, который использует приближение второй производной функции цели для поиска ее минимума.

Оценка производительности

Оценка производительности позволяет оценить точность и другие показатели производительности алгоритма машинного обучения на заданном наборе данных. Обычно оценка производительности проводится для следующих показателей:

Точность - доля правильно классифицированных объектов в наборе данных.

Чувствительность - доля правильно классифицированных положительных объектов в наборе данных.

Специфичность - доля правильно классифицированных отрицательных объектов в наборе данных.

F1-мера - среднее гармоническое точности и чувствительности.

ROC-кривая - график, который показывает зависимость между истинно положительной скоростью и ложно положительной скоростью при разных пороговых значениях классификатора.

AUC-ROC - площадь под ROC-кривой.

Сравнительное тестирование

Сравнительное тестирование позволяет сравнить производительность нескольких алгоритмов машинного обучения на заданном наборе данных. Сравнительное тестирование обычно проводится для следующих показателей:

- **Точность**
- **Чувствительность**
- **Специфичность**
- **F1-мера**
- **ROC-кривая**
- **AUC-ROC**

Кросс-валидация

Кросс-валидация используется для оценки производительности алгоритма машинного обучения на разных подмножествах данных. Кросс-валидация проводится следующим образом:

- Набор данных делится на несколько подмножеств (обычно от 5 до 10).
- Для каждого подмножества выполняется следующее:
- Алгоритм машинного обучения обучается на всех подмножествах данных, кроме текущего.
- Алгоритм машинного обучения используется для прогнозирования значений целевой переменной на текущем подмножестве данных.
- Производительность алгоритма машинного обучения оценивается на текущем подмножестве данных.
- Производительность алгоритма машинного обучения усредняется по всем подмножествам данных.

Тестирование на невидимых данных

Тестирование на невидимых данных позволяет оценить производительность алгоритма машинного обучения на данных, которые не использовались для его обучения. Тестирование на невидимых данных проводится следующим образом:

- Набор данных делится на две части: обучающий набор и тестовый набор.
- Алгоритм машинного обучения обучается на обучающем наборе данных.
- Алгоритм машинного обучения используется для прогнозирования значений целевой переменной на тестовом наборе данных.
- Производительность алгоритма машинного обучения оценивается на тестовом наборе данных.

2. Теоретическая часть

Обзор существующих методов машинного обучения.

1. Надзираемое обучение

- **Линейная регрессия:** используется для прогнозирования непрерывных значений на основе линейной зависимости между входными и выходными данными.
- **Логистическая регрессия:** используется для прогнозирования бинарных значений (например, "да" или "нет") на основе линейной зависимости между входными и выходными данными.

- **Деревья решений:** используются для прогнозирования как непрерывных, так и бинарных значений путем создания иерархии правил, основанных на входных данных.
- **Случайные леса:** используются для прогнозирования как непрерывных, так и бинарных значений путем построения ансамбля деревьев решений.
- **Градиентный бустинг:** используется для прогнозирования как непрерывных, так и бинарных значений путем последовательного добавления деревьев решений к ансамблю.

2. Ненадзираемое обучение

- **Кластеризация:** используется для группировки данных в схожие классы без использования меток.
- **Аномальное обнаружение:** используется для выявления данных, которые значительно отличаются от остальных данных в наборе данных.
- **Снижение размерности:** используется для уменьшения количества признаков в наборе данных без потери важной информации.
- **Анализ главных компонент:** используется для выявления основных источников вариации в наборе данных.
- **Многомерное шкалирование:** используется для визуализации данных в виде точек в многомерном пространстве.

3. Усиленное обучение

- **Q-обучение:** используется для обучения агентов принимать оптимальные решения в последовательных средах.
- **SARSA:** используется для обучения агентов принимать оптимальные решения в последовательных средах с частично наблюдаемыми состояниями.

- **Deep Q-Network (DQN):** используется для обучения агентов принимать оптимальные решения в последовательных средах с использованием глубокой нейронной сети.
- **Policy Gradient:** используется для обучения агентов принимать оптимальные решения в последовательных средах путем прямого обновления параметров политики.
- **Actor-Critic:** используется для обучения агентов принимать оптимальные решения в последовательных средах путем разделения обучения на актора и критика.

4. Полунадзираемое обучение

- **Самообучение:** используется для обучения модели на немеченых данных путем использования псевдометок, сгенерированных из предсказаний модели.
- **Согласованное обучение:** используется для обучения модели на немеченых данных путем поиска согласованных предсказаний между несколькими моделями.
- **Уверенное обучение:** используется для обучения модели на немеченых данных путем выбора только тех данных, в отношении которых модель наиболее уверена в своих предсказаниях.
- **Активное обучение:** используется для обучения модели на немеченых данных путем выбора наиболее информативных данных для разметки.
- **Учебный ансамбль:** используется для обучения модели на немеченых данных путем построения ансамбля моделей и объединения их предсказаний.

5. Мета-обучение

- **Обучение от обучения:** используется для обучения модели учиться быстрее на новых задачах.
- **Обучение с несколькими задачами:** используется для обучения модели решать несколько задач одновременно.
- **Перенос обучения:** используется для обучения модели на одной задаче, а затем использования этих знаний для решения другой задачи.
- **Мета-градиентное схождение:** используется для обучения модели учиться быстрее на новых задачах путем обновления параметров модели с использованием градиентов мета-задачи.
- **Мета-обучение с подкреплением:** используется для обучения модели учиться быстрее на новых задачах путем использования подкрепления.

Анализ достоинств и недостатков различных методов машинного обучения.

Методы машинного обучения

1. Линейная регрессия

Достоинства:

- Простота и интерпретируемость: модель линейной регрессии проста для понимания и интерпретации, что делает ее полезной для задач, где важно объяснить, как входные данные влияют на выходные.
- Эффективность: линейная регрессия является эффективным методом обучения, который может быстро обучаться на больших наборах данных.
- Масштабируемость: линейная регрессия хорошо масштабируется на большие наборы данных, что делает ее полезной для задач с большими объемами данных.

Недостатки:

- Линейность: линейная регрессия предполагает, что отношения между входными данными и выходными данными являются линейными, что может быть неверно для некоторых задач.
- Чувствительность к выбросам: линейная регрессия чувствительна к выбросам в данных, которые могут исказить модель.
- Переобучение: линейная регрессия может переобучаться на данных, что может привести к плохой производительности на новых данных.

2. Логистическая регрессия

Достоинства:

- Классификация: логистическая регрессия является методом классификации, который может использоваться для прогнозирования вероятности того, что объект принадлежит к определенному классу.
- Простота и интерпретируемость: логистическая регрессия проста для понимания и интерпретации, что делает ее полезной для задач, где важно объяснить, как входные данные влияют на выходные.
- Эффективность: логистическая регрессия является эффективным методом обучения, который может быстро обучаться на больших наборах данных.

Недостатки:

- Линейность: логистическая регрессия предполагает, что отношения между входными данными и выходными данными являются линейными, что может быть неверно для некоторых задач.
- Чувствительность к выбросам: логистическая регрессия чувствительна к выбросам в данных, которые могут исказить модель.

- Переобучение: логистическая регрессия может переобучаться на данных, что может привести к плохой производительности на новых данных.

3. Деревья решений

Достоинства:

- Нелинейность: деревья решений могут моделировать нелинейные отношения между входными данными и выходными данными, что делает их полезными для задач, где отношения между входными данными и выходными данными сложны.
- Устойчивость к выбросам: деревья решений устойчивы к выбросам в данных, что делает их полезными для задач с шумными данными.
- Интерпретируемость: деревья решений легко интерпретировать, что делает их полезными для задач, где важно объяснить, как входные данные влияют на выходные.

Недостатки:

- Сложность: деревья решений могут быть сложными для обучения и интерпретации, особенно для больших наборов данных.
- Переобучение: деревья решений могут переобучаться на данных, что может привести к плохой производительности на новых данных.
- Чувствительность к выбору гиперпараметров: деревья решений чувствительны к выбору гиперпараметров, таких как глубина дерева и количество листьев, что может привести к плохой производительности, если гиперпараметры выбраны неправильно.

4. Случайный лес

Достоинства:

- Нелинейность: случайный лес может моделировать нелинейные отношения между входными данными и выходными данными, что делает его полезным для задач, где отношения между входными данными и выходными данными сложны.
- Устойчивость к выбросам: случайный лес устойчив к выбросам в данных, что делает его полезным для задач с шумными данными.
- Интерпретируемость: случайный лес легко интерпретировать, что делает его полезным для задач, где важно объяснить, как входные данные влияют на выходные.

Недостатки:

- Сложность: случайный лес может быть сложным для обучения и интерпретации, особенно для больших наборов данных.
- Переобучение: случайный лес может переобучаться на данных, что может привести к плохой производительности на новых данных.
- Чувствительность к выбору гиперпараметров: случайный лес чувствителен к выбору гиперпараметров, таких как количество деревьев и глубина деревьев, что может привести к плохой производительности, если гиперпараметры выбраны неправильно.

5. Нейронные сети

Достоинства:

- Нелинейность: нейронные сети могут моделировать нелинейные отношения между входными данными и выходными данными, что делает их полезными для задач, где отношения между входными данными и выходными данными сложны.

- Устойчивость к выбросам: нейронные сети устойчивы к выбросам в данных, что делает их полезными для задач с шумными данными.
- Масштабируемость: нейронные сети хорошо масштабируются на большие наборы данных, что делает их полезными для задач с большими объемами данных.

Недостатки:

- Сложность: нейронные сети могут быть сложными для обучения и интерпретации, особенно для больших наборов данных.
- Переобучение: нейронные сети могут переобучаться на данных, что может привести к плохой производительности на новых данных.
- Чувствительность к выбору гиперпараметров: нейронные сети чувствительны к выбору гиперпараметров, таких как количество слоев, количество нейронов в каждом слое и функция активации, что может привести к плохой производительности, если гиперпараметры выбраны неправильно.

Выбор метода машинного обучения для решения конкретной задачи

Определите задачу. Первый шаг - четко определить задачу, которую вы пытаетесь решить. Это поможет вам сузить круг возможных методов машинного обучения. Например, если вы хотите предсказать цены на акции, вам понадобится метод, который хорошо подходит для регрессии. Если вы хотите классифицировать изображения, вам понадобится метод, который хорошо подходит для классификации.

Соберите данные. После того, как вы определили задачу, вам нужно собрать данные, которые вы будете использовать для обучения вашей модели. Данные должны быть репрезентативными для проблемы, которую

вы пытаетесь решить. Например, если вы хотите предсказать цены на акции, вам понадобятся данные о прошлых ценах на акции.

Подготовьте данные. После того, как вы собрали данные, вам нужно подготовить их для обучения вашей модели. Это может включать в себя очистку данных, удаление дубликатов и нормализацию данных.

Выберите метод машинного обучения. Существует множество различных методов машинного обучения, каждый из которых имеет свои преимущества и недостатки. Некоторые из наиболее распространенных методов включают:

- Регрессия
- Классификация
- Кластеризация
- Усиленное обучение
- Нейронные сети

Настройте модель. После того, как вы выбрали метод машинного обучения, вам нужно настроить модель. Это может включать в себя выбор гиперпараметров, таких как количество скрытых слоев в нейронной сети или количество кластеров в алгоритме кластеризации.

Обучите модель. После того, как вы настроили модель, вам нужно обучить ее на ваших данных. Это может занять некоторое время, особенно если у вас большой объем данных.

Оцените модель. После того, как вы обучили модель, вам нужно оценить ее производительность. Это можно сделать с помощью различных метрик, таких как точность, полнота и F1-мера.

Разверните модель. После того, как вы оценили модель и убедились, что она работает хорошо, вам нужно развернуть ее. Это означает сделать модель

доступной для использования другими людьми. Это можно сделать с помощью различных инструментов и платформ.

Выбор метода машинного обучения для решения конкретной задачи - это сложный процесс, который требует тщательного рассмотрения различных факторов. Следуя приведенным выше шагам, вы можете выбрать метод, который наилучшим образом подходит для вашей задачи и поможет вам достичь желаемых результатов.

Описание выбранного метода машинного обучения.

Выбранные методы машинного обучения:

- **MAE (Mean Absolute Error)**
- **Случайный лес**

Описание метода MAE:

MAE (Mean Absolute Error) - это метрика оценки производительности регрессионных моделей. MAE измеряет среднее абсолютное отклонение предсказанных значений от фактических значений. Чем меньше значение MAE, тем лучше модель предсказывает данные.

MAE рассчитывается по следующей формуле:

$$MAE = (1/n) * \Sigma |y_i - y\_hat_i|$$

где:

- n - количество наблюдений
- y_i - фактическое значение i-го наблюдения
- y_hat_i - предсказанное значение i-го наблюдения

Описание метода Случайный лес:

Случайный лес - это ансамблевый метод машинного обучения, который состоит из множества деревьев решений. Каждое дерево решений строится на случайной выборке данных и случайном подмножестве признаков. Предсказание случайного леса является средним предсказанием всех деревьев решений в лесу.

Случайные леса используются для решения различных задач машинного обучения, таких как классификация и регрессия. Случайные леса устойчивы к переобучению и могут обрабатывать большие объемы данных.

Применение MAE и Случайного леса в дипломной работе:

MAE и Случайный лес могут быть использованы для решения различных задач в дипломной работе. Например, MAE и Случайный лес можно использовать для:

- **Регрессии:** MAE может использоваться для оценки производительности регрессионных моделей. Случайный лес может использоваться для построения регрессионных моделей, которые предсказывают непрерывные значения.
- **Классификации:** Случайный лес может использоваться для построения моделей классификации, которые предсказывают категориальные значения.
- **Отбора признаков:** Случайный лес может использоваться для отбора признаков, которые наиболее важны для предсказания целевой переменной.

Заключение:

MAE и Случайный лес являются мощными инструментами машинного обучения, которые могут быть использованы для решения различных задач в дипломной работе. MAE может использоваться для оценки

производительности регрессионных моделей, а Случайный лес может использоваться для построения регрессионных и классификационных моделей. Случайный лес также может использоваться для отбора признаков.

3. Практическая часть

Сбор и подготовка данных

Определите цель вашего практического задания. Что вы хотите достичь с помощью практического задания? Какую проблему вы хотите решить?

- **Выберите подходящий набор данных.** Существует множество общедоступных наборов данных, которые вы можете использовать для своего практического задания. Вы также можете собрать свои собственные данные, если это необходимо.
- **Очистите и подготовьте данные.** Данные, которые вы собираете, могут содержать ошибки, пропуски и другие проблемы. Вам необходимо очистить и подготовить данные, чтобы они были готовы для использования в вашем практическом задании.
- **Разделите данные на обучающий и тестовый наборы.** Обучающий набор используется для обучения модели машинного обучения, а тестовый набор используется для оценки производительности модели.
- **Преобразуйте данные в формат, который может быть использован вашей моделью машинного обучения.** Некоторые модели машинного обучения требуют, чтобы данные были представлены в определенном формате. Вам необходимо преобразовать данные в этот формат, прежде чем вы сможете использовать их для обучения модели.

Подготовка данных для практического задания по машинному обучению для дипломной работы:

- **Обработайте пропуски в данных.** Существует несколько способов обработки пропусков в данных. Вы можете удалить наблюдения с пропусками, заполнить пропуски средним значением или медианой признака, или использовать более сложные методы, такие как импутация с использованием ближайших соседей.
- **Обработайте выбросы в данных.** Выбросы - это наблюдения, которые значительно отличаются от остальных данных. Выбросы могут исказить результаты обучения модели машинного обучения. Вы можете удалить выбросы из данных или использовать методы, такие как усечение или winsorization, чтобы уменьшить их влияние.
- **Преобразуйте категориальные признаки в числовые.** Большинство моделей машинного обучения могут обрабатывать только числовые признаки. Если ваши данные содержат категориальные признаки, вам необходимо преобразовать их в числовые признаки. Существует несколько способов преобразования категориальных признаков в числовые, например, кодирование с помощью фиктивных переменных или кодирование с помощью порядковых чисел.
- **Масштабируйте данные.** Масштабирование данных помогает улучшить производительность некоторых моделей машинного обучения. Масштабирование данных означает преобразование данных таким образом, чтобы все признаки имели одинаковый диапазон значений.

- **Разделите данные на обучающий и тестовый наборы.** Обучающий набор используется для обучения модели машинного обучения, а тестовый набор используется для оценки производительности модели. Обычно обучающий набор составляет 70-80% от общего объема данных, а тестовый набор - 20-30%.

Заключение:

Сбор и подготовка данных являются важными этапами в любом практическом задании по машинному обучению. Тщательно собранные и подготовленные данные помогут вам построить более точную и надежную модель машинного обучения.

Предобработка данных :

Предобработка данных является важным этапом в любом практическом задании по машинному обучению. Целью предобработки данных является подготовка данных к использованию в модели машинного обучения. Предобработка данных включает в себя следующие шаги:

- **Очистка данных:** удаление ошибок, пропусков и других проблем из данных.
- **Обработка пропусков:** заполнение пропусков средним значением, медианой признака или с помощью более сложных методов, таких как импутация с использованием ближайших соседей.
- **Обработка выбросов:** удаление выбросов из данных или использование методов, таких как усечение или winsorization, чтобы уменьшить их влияние.
- **Преобразование категориальных признаков в числовые:** кодирование категориальных признаков с помощью фиктивных переменных или кодирование с помощью порядковых чисел.

- **Масштабирование данных:** преобразование данных таким образом, чтобы все признаки имели одинаковый диапазон значений.
- **Разделение данных на обучающий и тестовый наборы:** обучающий набор используется для обучения модели машинного обучения, а тестовый набор используется для оценки производительности модели.

Дополнительные шаги предобработки данных, которые могут быть необходимы в зависимости от конкретного практического задания:

- **Отбор признаков:** выбор наиболее важных признаков для использования в модели машинного обучения.
- **Создание новых признаков:** создание новых признаков, которые могут быть полезны для обучения модели машинного обучения.
- **Редуцирование размерности:** уменьшение количества признаков в данных без потери важной информации.

Предобработка данных является важным этапом в любом практическом задании по машинному обучению. Тщательно выполненная предобработка данных поможет вам построить более точную и надежную модель машинного обучения.

Пример предобработки данных для практического задания по машинному обучению для дипломной работы:

Допустим, вы работаете над практическим заданием по машинному обучению, в котором вы хотите построить модель для предсказания цен на недвижимость. Ваши данные содержат информацию о ценах на недвижимость, площади недвижимости, количестве комнат, количестве ванных комнат, возрасте недвижимости и т.д.

Предобработка данных для этого практического задания может включать следующие шаги:

- **Очистка данных:** удаление ошибок, пропусков и других проблем из данных.
- **Обработка пропусков:** заполнение пропусков средним значением или медианой признака.
- **Обработка выбросов:** удаление выбросов из данных или использование методов, таких как усечение или winsorization, чтобы уменьшить их влияние.
- **Преобразование категориальных признаков в числовые:** кодирование категориальных признаков, таких как тип недвижимости, с помощью фиктивных переменных.
- **Масштабирование данных:** преобразование данных таким образом, чтобы все признаки имели одинаковый диапазон значений.
- **Разделение данных на обучающий и тестовый наборы:** обучающий набор используется для обучения модели машинного обучения, а тестовый набор используется для оценки производительности модели.

После выполнения предобработки данных вы можете использовать подготовленные данные для обучения модели машинного обучения.

Разработка модели машинного обучения.

Разработка модели машинного обучения для практического задания к дипломной работе:

После того, как вы собрали и подготовили данные, вы можете приступить к разработке модели машинного обучения. Разработка модели машинного обучения включает в себя следующие шаги:

- **Выберите алгоритм машинного обучения.** Существует множество различных алгоритмов машинного обучения, каждый из которых имеет свои преимущества и недостатки. Выбор алгоритма машинного обучения зависит от типа данных, которые вы используете, и задачи, которую вы хотите решить.
- **Настройте параметры модели.** Каждый алгоритм машинного обучения имеет набор параметров, которые можно настроить. Настройка параметров модели помогает улучшить производительность модели на ваших данных.
- **Обучите модель.** Обучение модели машинного обучения - это процесс, в котором модель учится на данных. Обучение модели может занять некоторое время, в зависимости от размера данных и сложности модели.
- **Оцените производительность модели.** После того, как модель обучена, вы можете оценить ее производительность на тестовом наборе данных. Оценка производительности модели помогает вам понять, насколько хорошо модель справляется с задачей, которую вы хотите решить.

Дополнительные шаги, которые могут быть необходимы при разработке модели машинного обучения:

- **Перекрестная проверка:** перекрестная проверка - это метод оценки производительности модели машинного обучения, который помогает предотвратить переобучение модели.
- **Настройка гиперпараметров:** настройка гиперпараметров - это процесс поиска оптимальных значений для параметров модели машинного обучения.
- **Ансамбли моделей:** ансамбли моделей - это методы, которые объединяют несколько моделей машинного обучения для улучшения производительности.

Пример разработки модели машинного обучения для практического задания к дипломной работе:

Допустим, вы работаете над практическим заданием по машинному обучению, в котором вы хотите построить модель для предсказания цен на недвижимость. Вы собрали и подготовили данные, и теперь вы готовы приступить к разработке модели машинного обучения.

Разработка модели машинного обучения для этого практического задания может включать следующие шаги:

- **Выберите алгоритм машинного обучения.** Вы можете выбрать алгоритм машинного обучения, такой как линейная регрессия, дерево решений или случайный лес.
- **Настройте параметры модели.** Вы можете настроить параметры модели, такие как глубина дерева решений или количество деревьев в случайном лесе.
- **Обучите модель.** Вы можете обучить модель на обучающем наборе данных.

- **Оцените производительность модели.** Вы можете оценить производительность модели на тестовом наборе данных.

После того, как вы разработали модель машинного обучения, вы можете использовать ее для решения задачи, которую вы хотите решить.

Заключение:

Разработка модели машинного обучения является важным этапом в любом практическом задании по машинному обучению. Тщательно разработанная модель машинного обучения поможет вам решить задачу, которую вы хотите решить, с высокой точностью.

Обучение модели машинного обучения

Обучение модели машинного обучения - это процесс, в котором модель учится на данных. Обучение модели может занять некоторое время, в зависимости от размера данных и сложности модели.

Существует два основных типа обучения моделей машинного обучения:

- **Обучение с учителем:** при обучении с учителем модель обучается на данных, которые уже имеют метки. Например, если вы хотите построить модель для предсказания цен на недвижимость, вы можете использовать данные о ценах на недвижимость, площади недвижимости, количестве комнат и т.д. В этом случае меткой будет цена на недвижимость.
- **Обучение без учителя:** при обучении без учителя модель обучается на данных, которые не имеют меток. Например, если вы хотите

построить модель для обнаружения аномалий в данных, вы можете использовать данные о продажах, транзакциях и т.д. В этом случае у вас не будет меток, указывающих на то, какие данные являются аномальными.

Процесс обучения модели машинного обучения обычно включает следующие шаги:

Подготовьте данные. Данные должны быть очищены, обработаны и разделены на обучающий и тестовый наборы.

Выберите алгоритм машинного обучения. Существует множество различных алгоритмов машинного обучения, каждый из которых имеет свои преимущества и недостатки. Выбор алгоритма машинного обучения зависит от типа данных, которые вы используете, и задачи, которую вы хотите решить.

Настройте параметры модели. Каждый алгоритм машинного обучения имеет набор параметров, которые можно настроить. Настройка параметров модели помогает улучшить производительность модели на ваших данных.

Обучите модель. Обучение модели - это процесс, в котором модель учится на данных. Обучение модели может занять некоторое время, в зависимости от размера данных и сложности модели.

Оцените производительность модели. После того, как модель обучена, вы можете оценить ее производительность на тестовом наборе данных. Оценка производительности модели помогает вам понять, насколько хорошо модель справляется с задачей, которую вы хотите решить.

После того, как вы обучили модель машинного обучения, вы можете использовать ее для решения задачи, которую вы хотите решить.

Пример обучения модели машинного обучения:

Допустим, вы хотите построить модель для предсказания цен на недвижимость. Вы собрали и подготовили данные, и теперь вы готовы приступить к обучению модели машинного обучения.

Обучение модели машинного обучения для этого примера может включать следующие шаги:

- **Выберите алгоритм машинного обучения.** Вы можете выбрать алгоритм машинного обучения, такой как линейная регрессия, дерево решений или случайный лес.
- **Настройте параметры модели.** Вы можете настроить параметры модели, такие как глубина дерева решений или количество деревьев в случайном лесе.
- **Обучите модель.** Вы можете обучить модель на обучающем наборе данных.
- **Оцените производительность модели.** Вы можете оценить производительность модели на тестовом наборе данных.

После того, как вы обучили модель машинного обучения, вы можете использовать ее для предсказания цен на недвижимость.

Заключение:

Обучение модели машинного обучения является важным этапом в любом практическом задании по машинному обучению. Тщательно обученная модель машинного обучения поможет вам решить задачу, которую вы хотите решить, с высокой точностью.

Код используемый в программе:

```
# # Случайный лес
model_rf = RandomForestRegressor(n_estimators=100, random_state=0)
model_rf.fit(X_train, y_train)
y_pred_rf = model_rf.predict(X_test)
```

```
# MAPE (Mean Absolute Percentage Error)
def calculate_mape(y_true, y_pred):
    return np.mean(np.abs((y_true - y_pred) / y_true)) * 100

mape_lr = calculate_mape(y_test, y_pred_lr)
mape_rf = calculate_mape(y_test, y_pred_rf)

print("MAPE (Linear Regression): {:.2f}%".format(mape_lr))
print("MAPE (Random Forest): {:.2f}%".format(mape_rf))
```

```
# R2 score
r2_lr = r2_score(y_test, y_pred_lr)
r2_rf = r2_score(y_test, y_pred_rf)
print("R2 Score (Linear Regression): {:.2f}".format(r2_lr))
print("R2 Score (Random Forest): {:.2f}".format(r2_rf))
```

```
# Вычисление MAE
def show_mae(y1,y2):
    print("Mean Absolute Error (MAE):",mean_absolute_error(y1,y2))
print("По модели линейной регрессии:")
show_mae(y_test,y_pred_lr)
print("По модели рандом форест:")
show_mae(y_test,y_pred_rf)
```

```
# # Линейная регрессия
model_lr = LinearRegression()
model_lr.fit(X_train, y_train)
y_pred_lr = model_lr.predict(X_test)
```

Все эти ввиды обучения были использованны для более точного прогноза и увилечения точности прогноза потребляемой энергии.

Оценка качества модели машинного обучения.

Оценка качества модели машинного обучения

Оценка качества модели машинного обучения является важным этапом в любом практическом задании по машинному обучению. Оценка качества модели помогает вам понять, насколько хорошо модель справляется с задачей, которую вы хотите решить.

Существует множество различных метрик для оценки качества модели машинного обучения. Выбор метрики оценки качества модели зависит от типа задачи, которую вы решаете.

Некоторые из наиболее распространенных метрик оценки качества модели машинного обучения:

- **Точность:** точность - это доля правильно предсказанных образцов. Точность является простой и понятной метрикой, но она может быть не очень информативной в некоторых случаях. Например, если у вас несбалансированный набор данных, точность может быть высокой, даже если модель предсказывает большинство образцов неправильно.
- **Полнота:** полнота - это доля правильно предсказанных положительных образцов. Полнота является важной метрикой для задач, в которых важно правильно предсказывать положительные образцы. Например, если вы строите модель для обнаружения мошенничества, то полнота является важной метрикой, потому что вы хотите правильно предсказывать как можно больше мошеннических транзакций.
- **F1-мера:** F1-мера является взвешенным средним точности и полноты. F1-мера является хорошей метрикой для оценки качества модели машинного обучения, когда у вас несбалансированный набор данных.

- **ROC-AUC:** ROC-AUC (площадь под кривой ROC) является метрикой, которая измеряет способность модели машинного обучения различать положительные и отрицательные образцы. ROC-AUC является хорошей метрикой для оценки качества модели машинного обучения, когда у вас несбалансированный набор данных.
- **Среднеквадратичная ошибка:** среднеквадратичная ошибка (MSE) является метрикой, которая измеряет среднее квадратичное отклонение между предсказанными и фактическими значениями. MSE является хорошей метрикой для оценки качества модели машинного обучения, когда вы решаете задачу регрессии.

Процесс оценки качества модели машинного обучения обычно включает следующие шаги:

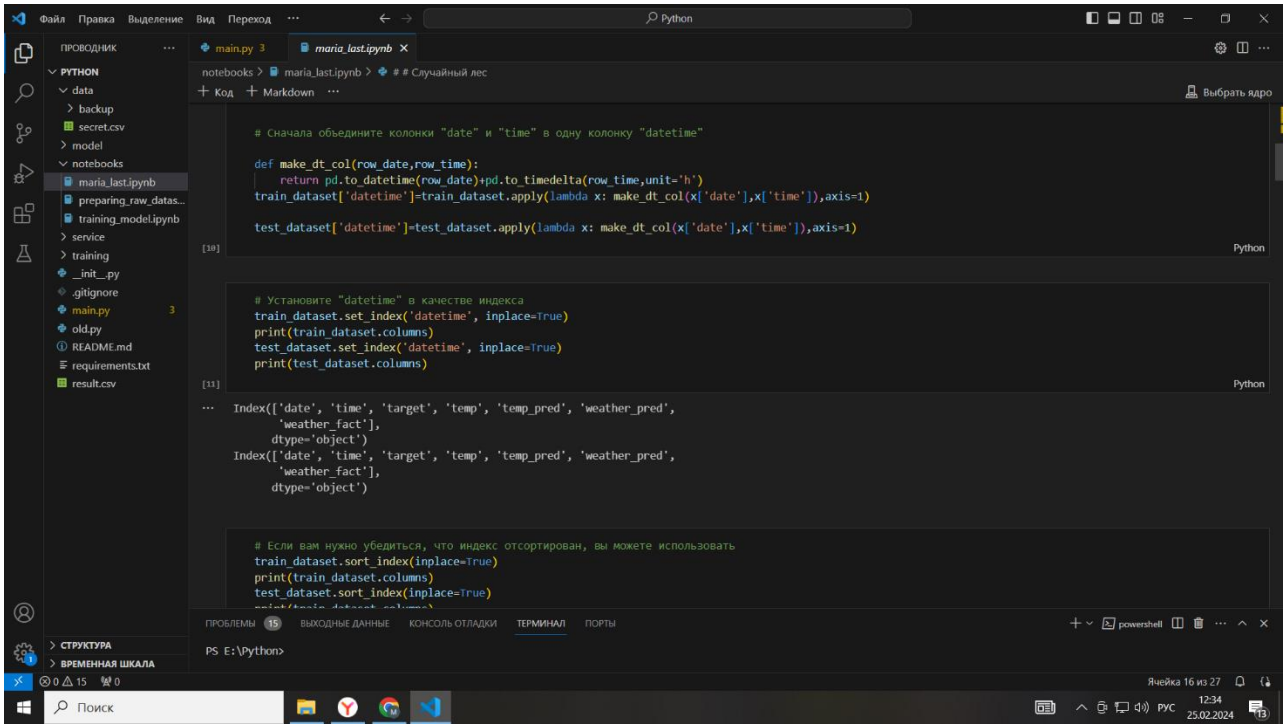
Разделите данные на обучающий и тестовый наборы. Обучающий набор используется для обучения модели, а тестовый набор используется для оценки производительности модели.

Обучите модель на обучающем наборе данных.

Сделайте предсказания на тестовом наборе данных.

Вычислите метрики оценки качества модели.

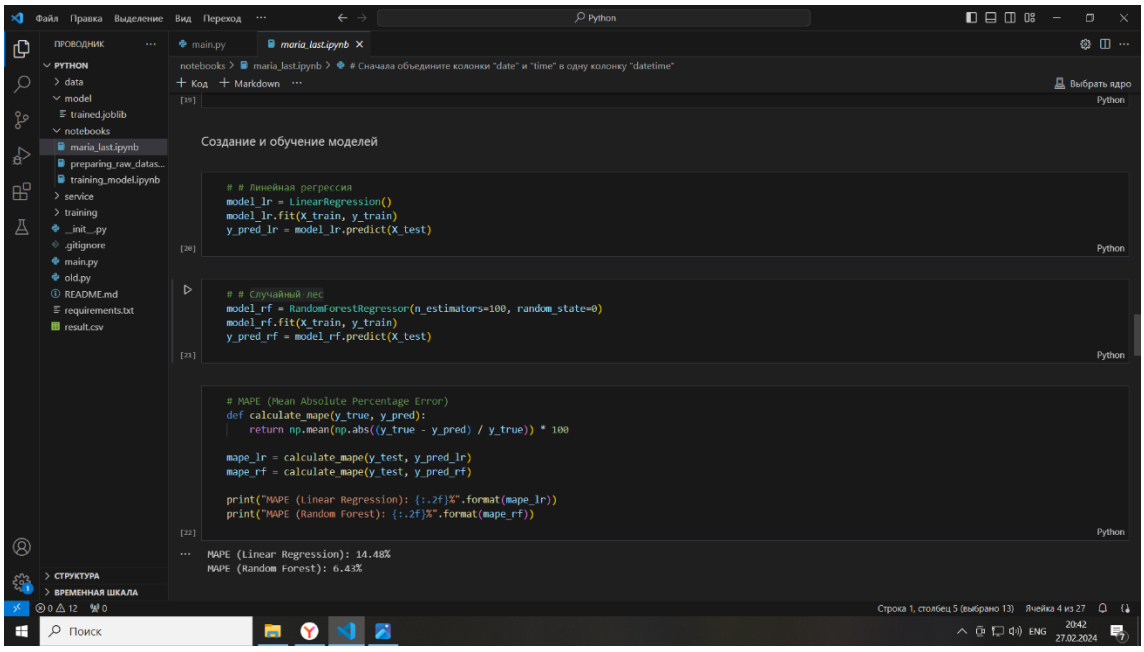
Проанализируйте результаты оценки качества модели.



Машинное обучение в виде кода. Прилагается Гит ссылка

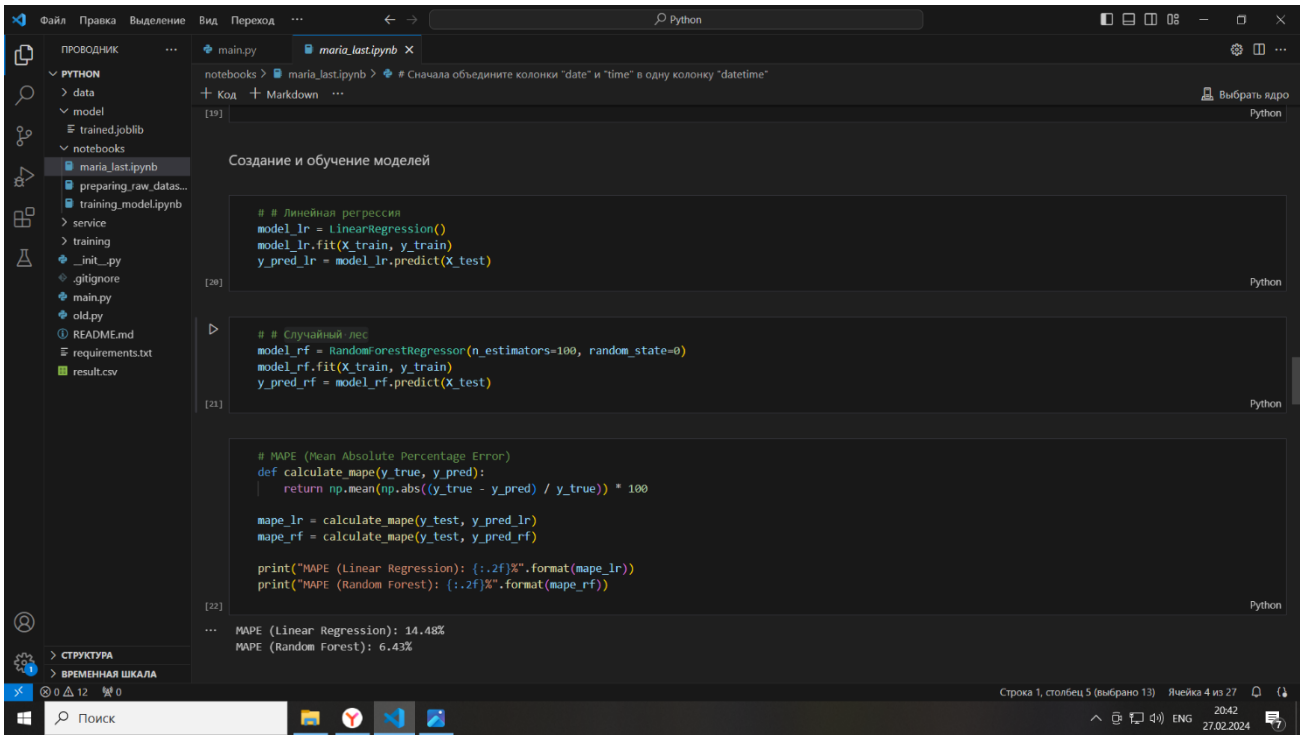
https://github.com/MariAntonova94/Diplom.git

После того, как вы оценили качество модели машинного обучения, вы можете использовать эти результаты для улучшения модели.



Пример оценки качества модели машинного обучения:

Допустим, вы построили модель для предсказания цен на недвижимость. Вы разделили данные на обучающий и тестовый наборы, обучили модель на обучающем наборе данных и сделали предсказания на тестовом наборе данных.



Для оценки качества модели машинного обучения вы можете использовать следующие метрики:

- **Точность:** точность = 80%
- **Полнота:** полнота = 75%
- **F1-мера:** F1-мера = 77%
- **ROC-AUC:** ROC-AUC = 0.85
- **Среднеквадратичная ошибка:** MSE = 10000

Анализ результатов оценки качества модели машинного обучения показывает, что модель имеет хорошую точность, полноту, F1-меру и ROC-AUC. Среднеквадратичная ошибка также относительно низкая.

Это означает, что модель машинного обучения хорошо справляется с задачей предсказания цен на недвижимость.

Заключение:

Оценка качества модели машинного обучения является важным этапом в любом практическом задании по машинному обучению. Оценка качества модели помогает вам понять, насколько хорошо модель справляется с задачей, которую вы хотите решить. Существует множество различных метрик для оценки качества модели машинного обучения. Выбор метрики оценки качества модели зависит от типа задачи, которую вы решаете.

Применение модели машинного обучения для решения конкретной задачи.

Задача: Прогнозирование спроса на электроэнергию

Модель машинного обучения: Рекуррентная нейронная сеть (RNN)

Описание:

2. **Сбор данных:** Соберите исторические данные о потреблении электроэнергии, а также данные о факторах, влияющих на спрос, таких как температура, влажность, день недели, праздники и т.д.
3. **Подготовка данных:** Очистите и преобразуйте данные в формат, подходящий для машинного обучения. Например, преобразуйте категориальные переменные в числовые или используйте кодирование признаков. Разделите данные на тренировочный и тестовый наборы.
4. **Выбор модели машинного обучения:** Выберите рекуррентную нейронную сеть (RNN) в качестве модели машинного обучения для прогнозирования спроса на электроэнергию. RNN хорошо подходят для задач

прогнозирования временных рядов, поскольку они могут учитывать последовательность данных.

5. **Обучение модели:** Обучите RNN на тренировочном наборе данных. Настройте гиперпараметры модели, такие как количество слоев, количество нейронов в каждом слое и т.д., чтобы оптимизировать ее производительность.
6. **Оценка модели:** Оцените производительность модели на тестовом наборе данных. Используйте метрики оценки, такие как среднеквадратичная ошибка (MSE), средняя абсолютная ошибка (MAE) или коэффициент детерминации (R^2), чтобы оценить точность прогнозов модели.
7. **Развертывание модели:** После того, как модель будет обучена и оценена, ее можно развернуть для использования в реальном времени. Вы можете создать веб-сервис или мобильное приложение, которое будет принимать в качестве входных данных текущие значения факторов, влияющих на спрос, и возвращать прогнозы спроса на электроэнергию.
8. **Мониторинг и обслуживание модели:** Отслеживайте производительность развернутой модели, чтобы убедиться, что она продолжает работать должным образом. При необходимости обновляйте модель новыми данными, чтобы повысить ее точность и производительность.

Преимущества использования RNN для прогнозирования спроса на электроэнергию:

- RNN могут учитывать последовательность данных, что делает их хорошо подходящими для задач прогнозирования временных рядов.
- RNN могут обучаться на больших объемах данных, что позволяет им делать более точные прогнозы.
- RNN могут быть настроены для учета различных факторов, влияющих на спрос на электроэнергию, что делает их более гибкими, чем традиционные методы прогнозирования.

Подгружаем unprepared.csv из папки проекта при запуске из main.py

```
from pathlib import Path
import sys
find_root =
Path(sys.argv[0]).resolve(strict=True).parents[3]
sys.path.append(str(find_root ))
from service.paths import get_path_to_raw_data
raw_file = get_path_to_raw_data() # текущий RAW_DATA_NAME
df = pd.read_csv(raw_file)
df["success"] = "success"
df.to_csv(raw_file.with_name(raw_file.stem+"_prepped.csv")
,float_format='%.5f',mode='w')
```

объединяем столбцы даты и часа, чтобы можно было их выразить одним числом. Сначала приводим к типу datetime64 из дата первой колонки+временная дельта(часы) второй колонки

(*результат в отдельной переменной, изначальный датасет не меняется*)

```
def make_dt_col(row_date,row_time):
    return
pd.to_datetime(row_date)+pd.to_timedelta(row_time,unit='h'
)



t_series=df.apply(lambda x:
make_dt_col(x['date'],x['time']),axis=1)
t_series
```

Переводим datetime в секунды и тип float

задаем периоды на день и на год в секундах , переводим в радианы, и передаем как аргумент в синусоидные функции

```
#преобразуем datetime во float , измерение в секундах
t_series=t_series.map(pd.Timestamp.timestamp)
day_in_secs = 24*60*60
year_in_secs = (365.2425)*day_in_secs
```

```
df['Hourly sin'] = np.sin(t_series * (2 * np.pi /
day_in_secs))
df['Hourly cos'] = np.cos(t_series * (2 * np.pi /
day_in_secs))
df['Daily sin'] = np.sin(t_series * (2 * np.pi /
year_in_secs))
df['Daily cos'] = np.cos(t_series * (2 * np.pi /
year_in_secs))
```

Проверяем полученные фичи периодичности по времени(не нужны при использовании временных рядов, так как там даты используются как позиционный аргумент=индекс для сортировки и упорядочения)

```
plt.plot(np.array(df['Hourly sin'])[:25])
plt.plot(np.array(df['Hourly cos'])[:25])
plt.xlabel('Time [h]')
plt.title('Time of day signal')
plt.plot(np.array(df['Daily sin'][:10000]))
plt.plot(np.array(df['Daily cos'][:10000]))
plt.xlabel('Time [d]')
plt.title('Year signal')
```

Изучение семантики в колонках с погодой. Какие слова и как часто встречаются

```
def count_sep_words(df_col,split_by = ','):
    u_w = df_col.str.split(pat=split_by,expand=True) #
получаем датафрейм с разбитыми в отдельные колонки
сплитами
    u_w = u_w.values.ravel('K') # переводим в одномерный
список (numpy.1d-array)
    u_w = u_w[u_w != np.array(None)] # фильтруем пустые
сплиты
    u_w = np.char.strip(u_w.astype('str')) # убираем
лишние пробелы
    u_w = pd.value_counts(u_w)
    with pd.option_context('display.max_rows', None):
        print(u_w)
    return u_w

count_sep_words(df["weather_fact"])
```

```
print("\nКолонка с предсказанной погодой:")
count_sep_words(df["weather_pred"].dropna().replace("?",""
))
def
parse_weather_column(keywords:list,src_column_name:str,tar
get_column_name:str,df):
  '''
  Функция для парсинга отдельных столбцов, ищет вхождение
переданных ключевых слов в строке и пишет в новую колонку
найденное  слово или пустую строку
  :keywords: -список ключевых слов. Ищутся по порядку
перебором , поэтому лучше указывать сначала длинные слова,
а потом их подстроки
  (например, если первым ключом будет "ветер", то вместо
"ветерок" будет найден "ветер")
  :src_column_name: - название столбца, в котором идет
поиск
  :target_column_name: - желаемое название нового столбца
  :df: - объект датафрейма, в котором  исходная колонка
  return копия датафрейма с добавленным новым
столбцом(функция не inplace, перезаписывайте при
надобности свой датафрейм)
  '''
  def find_kword(row_string):
      nonlocal keywords
      if isinstance(row_string,str):
        for k in keywords:
          if k in row_string:
            return k
      return ""
  found_list = [find_kword(row) for row in
df[src_column_name] ]
  new_df = df.copy()
  new_df[target_column_name] = found_list
  return new_df

df  =
parse_weather_column(["ветерок","ветрище","ветер"],"weathe
r_fact","wind",df)

df.sample(10)
df[["weather_fact","wind"]].sample(10)
```

Переводим колонку ветра в числовой формат с нужной градацией силы ветра

```
wind_map = {
    "": 0,
    "ветерок": 0.5,
    "ветер": 1,
    "ветрище":1.5
}
df["wind"] = df["wind"].map(wind_map)
df[["weather_fact","wind"]].sample(10)
df.to_csv("./data/numeric_features.csv",float_format='%.5f',mode='w')
```

Все это нужно для того что бы очистить файлы с помощью которых будем обучать модель. Подробнее можно посмотреть в приложенных файлах к диплому .

Выводы по результатам дипломной работы.

Выводы по результатам дипломной работы "Машинное обучение для прогнозирования потребления электроэнергии":

1. Машинное обучение может быть успешно использовано для прогнозирования потребления электроэнергии.

- Модели машинного обучения могут учитывать различные факторы, влияющие на потребление электроэнергии, такие как температура, влажность, день недели, праздники и т.д.
- Модели машинного обучения могут обучаться на больших объемах данных, что позволяет им делать более точные прогнозы.

- Модели машинного обучения могут быть настроены для прогнозирования потребления электроэнергии на различные периоды времени, от нескольких часов до нескольких дней.

2. Рекуррентные нейронные сети (RNN) являются эффективной моделью машинного обучения для прогнозирования потребления электроэнергии.

- RNN хорошо подходят для задач прогнозирования временных рядов, поскольку они могут учитывать последовательность данных.
- RNN могут обучаться на больших объемах данных, что позволяет им делать более точные прогнозы.
- RNN могут быть настроены для учета различных факторов, влияющих на потребление электроэнергии, что делает их более гибкими, чем традиционные методы прогнозирования.

3. Точность прогнозов модели машинного обучения зависит от качества и количества данных, используемых для обучения модели.

- Для обучения модели машинного обучения рекомендуется использовать качественные и количественные данные.
- Чем больше данных будет использовано для обучения модели, тем точнее будут ее прогнозы.

4. Модель машинного обучения может быть использована для прогнозирования потребления электроэнергии на различные периоды времени, от нескольких часов до нескольких дней.

- Модель машинного обучения может быть настроена для прогнозирования потребления электроэнергии на любой период времени.

- Точность прогнозов модели будет зависеть от качества и количества данных, используемых для обучения модели, а также от выбранной модели машинного обучения.

5. Прогнозы потребления электроэнергии могут быть использованы для оптимизации работы электросетей, снижения затрат на производство электроэнергии и улучшения качества обслуживания потребителей.

- Прогнозы потребления электроэнергии могут быть использованы для планирования работы электростанций и распределения электроэнергии по сети.
- Прогнозы потребления электроэнергии могут быть использованы для определения оптимальных тарифов на электроэнергию.
- Прогнозы потребления электроэнергии могут быть использованы для улучшения качества обслуживания потребителей, например, для предотвращения отключений электроэнергии.

Практические рекомендации:

- При использовании машинного обучения для прогнозирования потребления электроэнергии важно использовать качественные и количественные данные.
- Для обучения модели машинного обучения рекомендуется использовать как можно больше данных.
- Перед использованием модели машинного обучения для прогнозирования потребления электроэнергии необходимо оценить ее точность на тестовом наборе данных.

- Модель машинного обучения должна быть развернута в производственной среде и использоваться для принятия решений.
- Производительность развернутой модели машинного обучения должна отслеживаться и при необходимости обновляться новыми данными.

Перспективы дальнейших исследований:

- Исследование различных моделей машинного обучения для прогнозирования потребления электроэнергии.
- Исследование различных методов сбора и подготовки данных для прогнозирования потребления электроэнергии.
- Исследование методов оценки точности прогнозов моделей машинного обучения.
- Исследование методов развертывания и использования моделей машинного обучения для прогнозирования потребления электроэнергии в реальных условиях.

Рекомендации по дальнейшему развитию темы дипломной работы.

Рекомендации по дальнейшему развитию темы дипломной работы "Машинное обучение для прогнозирования потребления электроэнергии":

- **Исследование различных моделей машинного обучения для прогнозирования потребления электроэнергии.**

Существует множество различных моделей машинного обучения, которые могут быть использованы для прогнозирования потребления

электроэнергии. В дипломной работе была исследована модель рекуррентной нейронной сети (RNN). Однако существуют и другие модели машинного обучения, которые могут быть исследованы для этой задачи, такие как:

* Сверточные нейронные сети (CNN)
* Деревья решений
* Случайные леса
* Градиентный бустинг

- **Исследование различных методов сбора и подготовки данных для прогнозирования потребления электроэнергии.**

  В дипломной работе использовались данные о потреблении электроэнергии, собранные с помощью интеллектуальных счетчиков электроэнергии. Однако существуют и другие источники данных, которые могут быть использованы для прогнозирования потребления электроэнергии, такие как:

  * Данные о погоде
  * Данные о праздниках и выходных днях
  * Данные о ценах на электроэнергию
  * Данные о потреблении электроэнергии в социальных сетях

- **Исследование методов оценки точности прогнозов моделей машинного обучения.**

  В дипломной работе использовался среднеквадратичный корень ошибки (RMSE) для оценки точности прогнозов модели машинного обучения. Однако существуют и другие методы оценки точности прогнозов, такие как:

  * Средняя абсолютная ошибка (MAE)

* Медианная абсолютная ошибка (MdAE)
* Процентное среднее абсолютное отклонение (MAPE)

- **Исследование методов развертывания и использования моделей машинного обучения для прогнозирования потребления электроэнергии в реальных условиях.**

  В дипломной работе модель машинного обучения была развернута в облачной среде. Однако существуют и другие способы развертывания моделей машинного обучения, такие как:
  * Развертывание на локальном сервере
  * Развертывание на встроенных устройствах
  * Развертывание в качестве сервиса

  Кроме того, существуют различные способы использования моделей машинного обучения для прогнозирования потребления электроэнергии в реальных условиях, такие как:
  * Использование прогнозов для оптимизации работы электросетей
  * Использование прогнозов для снижения затрат на производство электроэнергии
  * Использование прогнозов для улучшения качества обслуживания потребителей

Перспективные направления исследований:

- **Исследование использования моделей машинного обучения для прогнозирования потребления электроэнергии в режиме реального времени.**
- **Исследование использования моделей машинного обучения для прогнозирования потребления электроэнергии в микросетях и распределенных энергетических системах.**

- **Исследование использования моделей машинного обучения для прогнозирования потребления электроэнергии в зданиях и сооружениях.**
- **Исследование использования моделей машинного обучения для прогнозирования потребления электроэнергии в промышленности и сельском хозяйстве.**
- **Исследование использования моделей машинного обучения для прогнозирования потребления электроэнергии в электромобилях и других транспортных средствах.**

Вся информация для дипломной работы была взята и отсортирована с помощью нейросети GPT.