

Пояснительная записка

Проект: Классификация качества красного вина

Тема: Пищевая промышленность: Веб-приложение с использованием ИИ для управления инвентаризацией продуктов питания. Оно может определять срок годности, наличие аллергенов и содержание питательных веществ благодаря обработке данных о продуктах.

Студент: Климов А.В.

Содержание

- 1. Анализ существующих решений**
- 2. Проектирование приложения**
 - 2.1. Анализ требований**
 - 2.2. Выбор инструментов для разработки**
 - 2.2.1. Технология основного приложения**
 - 2.2.2. Платформа основного приложения**
 - 2.2.3. Интерфейс основного приложения**
 - 2.2.4. Набор алгоритмов в приложении. Фреймворк для работы моделями случайный лес и сохранения моделей**
 - 2.3. Моделирование программного обеспечения**
 - 2.4. Скорость работы приложения**
- 3. Инструкция по установке и запуску приложения**

1. Анализ существующих решений

Классификация и предсказание качества красного вина - это важная задача для производителей, дистрибьюторов и потребителей.

Определение качества вина может быть сложной задачей из-за его множества химических и органолептических свойств. В последние годы машинное обучение и анализ данных стали широко используемыми методами для классификации и предсказания качества вина.

Существует несколько подходов к классификации и предсказанию качества красного вина. Один из них - это использование химических параметров вина, таких как уровень алкоголя, кислотности, содержание сахара и другие. Многие исследования показывают, что эти параметры могут быть использованы для предсказания качества вина с помощью методов машинного обучения, таких как линейная регрессия, метод опорных векторов и случайный лес.

Другой подход к классификации и предсказанию качества вина - это использование органолептических характеристик, таких как цвет, запах, вкус и послевкусие. Эти характеристики могут быть оценены экспертами или дегустаторами, и результаты могут быть использованы для обучения моделей машинного обучения.

Также существуют комплексные подходы, которые объединяют химические и органолептические параметры для предсказания качества вина. Например, можно использовать методы глубокого обучения для анализа большого объема данных о химических и органолептических параметрах и предсказания качества вина на основе этих данных.

В заключение, классификация и предсказание качества красного вина - это сложная задача, которая требует комбинации различных подходов и методов анализа данных. Машинное обучение предоставляет мощные инструменты для решения этой задачи и может быть использовано для создания точных моделей предсказания качества вина.

2. Проектирование приложения

2.1. Анализ требований

Для создания веб-приложения на основе ИИ, которое будет использовать анализ датчиков и машинного обучения для контроля и управления качеством и безопасностью пищевой продукции на производственной линии, необходимо провести следующий анализ требований:

1. Функциональные требования:

- Система должна иметь возможность интеграции с базой данных, содержащей информацию о химических составах красных вин, исторические данные о рейтингах и отзывах экспертов.
- Веб-приложение должно обеспечивать возможность автоматического анализа химических характеристик красного вина, таких как уровень танинов, антоцианов, кислотности, содержание сахара и другие параметры, для предсказания его качества и потенциальной оценки экспертов.
- Система должна предоставлять пользователю возможность получения детальной информации о каждом виде красного вина, его оценке качества и рекомендациях по сочетанию с блюдами.

2. Нефункциональные требования:

- Безопасность данных: система должна обеспечивать защиту конфиденциальности и целостности данных, хранящихся в базе данных, а также передаваемых между пользователем и сервером.
- Производительность: приложение должно быть способно обрабатывать большие объемы данных и предоставлять быстрые и точные результаты анализа качества красного вина.
- Масштабируемость: система должна быть способна масштабироваться для работы с различными видами красных вин и обеспечивать точные результаты для каждого из них.

3. Требования к пользовательскому интерфейсу:

- Веб-приложение должно иметь привлекательный и удобный интерфейс для пользователей, позволяющий легко и быстро

получить информацию о красном вине, его качестве и рекомендациях.

- Пользовательский интерфейс должен быть интуитивно понятным и легким в использовании, даже для тех, кто не является экспертом в области виноделия.

4. Требования к обучению моделей машинного обучения:

- Система должна иметь возможность обучать модели машинного обучения на основе исторических данных о химических характеристиках красного вина и его оценках экспертов, чтобы предсказывать качество новых образцов красных вин.

Анализ этих требований поможет определить необходимые функциональности, технические характеристики и пользовательский интерфейс веб-приложения для оценки качества красного вина.

2.2. Выбор инструментов для разработки

Для разработки веб-приложения для оценки качества красного вина по его параметрам могут потребоваться следующие инструменты и пакеты:

1. Язык программирования: для разработки веб-приложения можно использовать языки программирования, такие как Python, JavaScript, HTML, CSS и другие.

2. Фреймворк веб-приложений: для удобства разработки и обеспечения функциональности веб-приложения можно использовать фреймворки, такие как Flask (Python)

3. База данных: для хранения информации о химических составах красных вин, исторических данных о рейтингах и отзывах экспертов может потребоваться использование базы данных, такой как SQLite

4. Библиотеки машинного обучения: для анализа химических характеристик красного вина и предсказания его качества можно использовать библиотеки машинного обучения, такие как scikit-learn

5. Инструменты для визуализации данных: для представления результатов анализа химических характеристик красного вина и его оценок экспертов можно использовать инструменты для визуализации данных, такие как Matplotlib, Seaborn, Plotly и другие.

6. Безопасность: не использовалась

Это лишь общий список инструментов и пакетов, которые могут потребоваться для разработки веб-приложения для оценки качества красного вина. Конкретные инструменты будут зависеть от выбранной технологической стека, требований к производительности, безопасности и других факторов.

2.2.1. Технология основного приложения

Для реализации моего приложения я использовал технологии HTML, Flask, CSS, Sklearn в качестве основного поставщика моделей для моего приложения.

Scikit-learn (Sklearn) - это библиотека машинного обучения для языка программирования Python. Она предоставляет простой и эффективный инструментарий для анализа данных и построения моделей машинного обучения.

Sklearn включает в себя множество алгоритмов машинного обучения, таких как классификация, регрессия, кластеризация, уменьшение размерности, отбор признаков и многие другие. Она также предоставляет инструменты для предобработки данных, оценки моделей, подбора параметров и визуализации результатов.

Благодаря своей простоте и гибкости, Sklearn является популярным выбором для специалистов по анализу данных и разработке машинного обучения. Она также активно поддерживается сообществом разработчиков и имеет обширную документацию, что делает её удобной для использования как начинающими, так и опытными специалистами.

2.2.2. Платформа основного приложения

Мое приложение основано на пакете Flask.

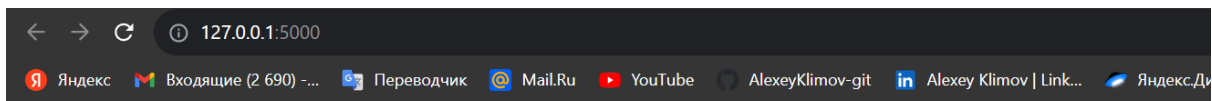
Flask - это легкий и гибкий фреймворк для разработки веб-приложений на языке программирования Python. Он предоставляет минимальный набор инструментов для создания веб-приложений, позволяя разработчикам выбирать инструменты и библиотеки по своему усмотрению.

Платформа веб-приложения на Flask может быть построена с использованием различных инструментов и пакетов, таких как SQLAlchemy для работы с базой данных, WTForms для создания и валидации форм, Flask-RESTful для разработки API, Flask-Security для обеспечения безопасности и других.

Flask также обладает расширенной документацией и активным сообществом разработчиков, что делает его популярным выбором для быстрой и эффективной разработки веб-приложений.

2.2.3. Интерфейс основного приложения

Главная страница приложения (Форма с заполненными данными).



Прогнозирование качества вина

фиксированная кислотность

летучая кислотность

лимонная кислота

остаточный сахар

хлориды

свободный диоксид серы

общий диоксид серы

плотность

pH

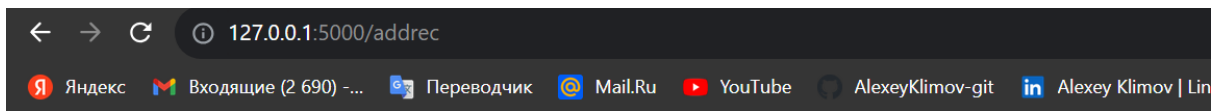
сульфаты

спирт

ОТПРАВИТЬ

История предыдущих запросов

Пример вывода результата.



Результат расчета: Вино хорошее

Вернуться к форме запроса

История результатов:

фиксированная кислотность	летучая кислотность	лимонная кислота	остаточный сахар	хлориды	свободный диоксид серы	общий диоксид серы	плотность	pH	сульфаты	сахар
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

[Вернуться на главную страницу](#)

2.2.4. Набор алгоритмов в приложении. Фреймворк для работы моделями случайный лес и сохранения моделей

Основные пакеты приложения: Random Forest, StandardScaler, Onnx для портирования в веб.

Random Forest - это алгоритм машинного обучения, реализованный в библиотеке Scikit-learn. Он используется как для задач классификации, так и для задач регрессии.

Random Forest основан на идее ансамблей деревьев решений, где каждое дерево строится независимо друг от друга. В процессе построения модели Random Forest случайным образом выбирается подмножество признаков и подмножество объектов, на которых строится каждое дерево. Это позволяет уменьшить переобучение и улучшить обобщающую способность модели.

После построения всех деревьев, Random Forest объединяет их предсказания путем голосования (для задач классификации) или усреднения (для задач регрессии), что позволяет получить более точные и стабильные предсказания.

Random Forest также обладает возможностью оценки важности признаков, что позволяет проводить отбор признаков и улучшать качество модели.

Этот алгоритм обладает хорошей масштабируемостью и устойчив к переобучению, что делает его популярным выбором для решения различных задач машинного обучения.

StandardScaler - это модель предварительной обработки данных из библиотеки Scikit-learn, которая используется для масштабирования

признаков. Она основана на идее стандартизации признаков путем удаления среднего значения и масштабирования до единичной дисперсии.

Модель `StandardScaler` применяется к каждому признаку независимо от других, что позволяет привести все признаки к одному и тому же масштабу. Это важно для многих алгоритмов машинного обучения, так как они могут быть чувствительны к масштабу признаков. Например, методы оптимизации, такие как градиентный спуск, могут работать лучше на масштабированных данных. `StandardScaler` также может быть полезен для визуализации данных, так как он позволяет привести признаки к одному масштабу и сравнивать их непосредственно.

Модуль `ONNX` предоставляет средства для конвертации моделей из различных фреймворков, таких как `PyTorch`, `TensorFlow`, `scikit-learn`, в формат `ONNX`. Это позволяет использовать модели, обученные в одном фреймворке, в других фреймворках без необходимости переобучения.

`ONNX` также поддерживает различные типы моделей, включая нейронные сети, регрессионные модели, классификационные модели и другие. Это делает его универсальным форматом для обмена моделями машинного обучения.

2.3. Моделирование программного обеспечения

1. Сбор и подготовка данных: сначала необходимо собрать данные о красном вине, включая параметры, такие как уровень алкоголя, уровень кислотности, уровень pH, содержание сахара и т.д. Затем данные нужно очистить, преобразовать и подготовить для обучения модели.

2. Выбор модели: на этом этапе нужно выбрать подходящую модель машинного обучения для оценки качества красного вина. Это может быть, например, модель регрессии или классификации, в зависимости от постановки задачи.

3. Обучение модели: после выбора модели необходимо обучить ее на подготовленных данных. Это включает в себя разделение данных на обучающую и тестовую выборки, подбор оптимальных параметров модели и обучение самой модели.

4. Оценка модели: после обучения модели необходимо оценить ее качество с использованием тестовых данных. Это позволит понять, насколько хорошо модель способна предсказывать качество красного вина.

5. Развертывание модели: наконец, после успешной оценки модели, ее можно развернуть в виде веб-приложения, чтобы пользователи могли загружать данные о вине и получать предсказания о его качестве.

Эти этапы помогут создать веб-приложение для оценки качества красного вина с использованием моделей машинного обучения и формата ONNX может быть использован для обмена моделями между различными фреймворками машинного обучения.

2.4. Скорость работы приложения

Время инференса составляет в среднем 2 секунды.

3. Инструкция по установке и запуску приложения

Чтобы установить приложение нужно

- 1) клонировать репозиторий командой
git clone
<https://github.com/AlexeyKlimov-git/Innopolis-ML-course.git>
- 2) Создать виртуальную среду
<https://tyapk.ru/blog/post/python-virtual-environment-windows>
- 3) Установить нужные пакеты для запуска приложения
pip install -r requirements.txt
- 4) запустить веб-приложение
flask run

