

Dokumentacja Aplikacji do Analizy Danych

Aplikacja usprawnia proces analizy zestawów danych, umożliwiając ich łatwe ładowanie, wstępne przetwarzanie oraz interaktywną wizualizację za pośrednictwem intuicyjnego interfejsu użytkownika.

Główne funkcje

Ładowanie danych

Obsługuje ładowanie zestawów danych w formatach CSV oraz JSON.

Wstępne przetwarzanie danych

Umożliwia obsługę brakujących wartości poprzez ich zastąpienie średnią, wypełnienie określoną wartością lub usunięcie wierszy.

Pozwala na zmianę nazw lub usuwanie zbędnych kolumn w celu przygotowania danych do dalszej analizy.

Interaktywna wizualizacja danych

Generowanie różnych typów wykresów, takich jak wykresy punktowe, liniowe, histogramy czy mapy cieplne.

Wykorzystanie technik redukcji wymiarów, takich jak PCA i UMAP, w celu wizualizacji danych o wysokiej liczbie wymiarów.

Technologie oraz decyzje projektowe

Wykorzystane technologie

Streamlit: Ułatwia tworzenie interaktywnych i intuicyjnych aplikacji internetowych.

Pandas: Efektywnie wspiera manipulację danymi oraz procesy wstępnego przetwarzania.

Plotly: Zapewnia dynamiczne, interaktywne oraz estetyczne wizualizacje danych.

Scikit-learn: Dostarcza narzędzi do wstępnego przetwarzania danych oraz implementacji PCA.

UMAP-learn: Wspiera nieliniową redukcję wymiarów, umożliwiając zaawansowaną analizę danych.

Decyzje projektowe

Aby zapewnić modularność, skalowalność i łatwość w rozwoju aplikacji, zastosowano szereg wzorców projektowych:

Wzorzec fabryki

FileReaderFactory: Dynamicznie tworzy odpowiedni czytnik danych (CSV lub JSON) w zależności od typu pliku.

DimReductionFactory: Inicjalizuje wybraną metodę redukcji wymiarów (PCA lub UMAP).

Wzorzec strategii

PlotStrategy: Umożliwia dynamiczny wybór strategii generowania wykresów (np. punktowy, liniowy, histogram).

Wzorzec fasady

DataAnalysisFacade: Zapewnia uproszczony interfejs do głównych funkcji aplikacji, takich jak ładowanie danych, wstępne przetwarzanie i wizualizacja.

Wzorzec metody szablonowej

DimReduction: Udostępnia standardowy przebieg pracy dla redukcji wymiarów, umożliwiając specyficzne implementacje (PCA/UMAP).

Wzorzec singleton

PlotContext oraz **DataAnalysisFacade:** Zapewniają, że tylko jedna instancja tych klas jest aktywna w danym czasie, co gwarantuje spójność oraz integralność stanu aplikacji.

Rozszerzalność

Architektura aplikacji pozwala na łatwe rozszerzanie poprzez:

Dodanie nowych formatów plików (np. Excel, Parquet).

Wprowadzenie dodatkowych typów wizualizacji danych.

Implementację nowych technik redukcji wymiarów oraz metod wstępnego przetwarzania.

Przykładowe użycie:

1. **Ładowanie danych:** Prześlij swój zestaw danych w formacie CSV lub JSON lub wybierz jeden z wbudowanych zestawów danych przykładowych.
2. **Wstępne przetwarzanie danych:** Skorzystaj z opcji dostępnych w bocznym panelu, aby obsługiwać brakujące wartości, zmieniać nazwy kolumn lub usuwać niepotrzebne.
3. **Wizualizacja danych:** Wybierz typ wykresu lub zastosuj redukcję wymiarów, aby interaktywnie analizować dane o wysokiej liczbie wymiarów.
4. **Eksport wyników:** Opcjonalnie, pobierz generowane wizualizacje do późniejszego wykorzystania.

Dzięki modularnej, interaktywnej architekturze aplikacja pozwala użytkownikom na płynne i efektywne eksplorowanie oraz analizowanie zestawów danych, stanowiąc solidną podstawę do przyszłych rozszerzeń.