

Diabetes Insight Analyzer

Data: 23.01.2024

Członkowie Zespołu Projektowego:

Remigiusz Pręgowski, Jakub Guliński, Oleksii Haida, Remigiusz Pręgowski

Cel Projektu

Projekt ten ma na celu stworzenie zaawansowanego narzędzia analitycznego, które wykorzystuje techniki uczenia maszynowego do przeprowadzenia szczegółowej analizy danych związanych z cukrzycą. Celem jest nie tylko identyfikacja potencjalnych przypadków cukrzycy, ale także zrozumienie wzorców i czynników ryzyka związanych z tą chorobą. Dzięki temu narzędziu, specjaliści medyczni oraz badacze będą mogli lepiej przewidywać wystąpienie cukrzycy, dostosowywać terapie i prowadzić bardziej celowane badania epidemiologiczne. Projekt ma również na celu poprawę świadomości i edukacji dotyczącej cukrzycy poprzez udostępnianie łatwo zrozumiałych analiz i raportów. W ten sposób "Diabetes Insight Analyzer" ma przyczynić się do lepszego zrozumienia i zarządzania cukrzycą na poziomie indywidualnym i społecznym.

Sposób Użytkowania

Po wchodzeniu na stronę internetową użytkownik widzi :

1. Wybór Operacji:

- **Run Pipeline:** Uruchom pełny proces analizy danych, który może obejmować czyszczenie danych, wstępną obróbkę i ewentualnie trenowanie modelu.
- **Load New Model:** Załaduj nowy model uczenia maszynowego do systemu, aby umożliwić przewidywanie na podstawie najnowszych danych i algorytmów.
- **Generate Synthetic Data:** Wygeneruj syntetyczne dane, które mogą być użyte do testowania modelu bez konieczności używania prawdziwych danych pacjentów, co jest przydatne w przypadku zachowania prywatności danych.

2. Wprowadzenie Danych Pacjenta:

Gender: Wprowadź płeć pacjenta.

Age: Podaj wiek pacjenta.

Hypertension: Zaznacz, czy pacjent cierpi na nadciśnienie.

co ja robie

run pipeline

load new model

generate synthetic data

gender

age

hypertension

heart_disease

smoking history

bmi

hba1c_level

blood_glucose_level

predict

nie wiadomo czy jest chory

Heart Disease: Zaznacz, czy pacjent ma chorobę serca.

Smoking History: Określ historię palenia pacjenta.

BMI: Wprowadź wskaźnik masy ciała pacjenta.

HbA1c Level: Podaj poziom HbA1c, który jest markerem kontroli cukrzycy.

Blood Glucose Level: Podaj poziom glukozy we krwi.

3. Przewidywanie:

Po wprowadzeniu wszystkich wymaganych danych, kliknij przycisk Predict (Prognozuj), aby system mógł przewidzieć ryzyko cukrzycy u pacjenta.

4. Interpretacja Wyników:

Po uruchomieniu predykcji, system przetworzy dane i na podstawie modelu przewidzi, czy pacjent jest chory. Wynik zostanie zaprezentowany użytkownikowi, zazwyczaj w postaci prawdopodobieństwa lub klasyfikacji ryzyka. Model może zaklasyfikować model do takich klas: „**Jest chory**”, „**Jest nie chory**”. W każdym razie na początku będzie się wyświetlać „Nie wiadomo czy jest chory”, co oznacza że model jeszcze nie jest gotowy do klasyfikacji.

co ja robie

run pipeline

load new model

generate synthetic data

gender

1

age

1

hypertension

1

heart_disease

1

smoking history

1

bmi

1

hba1c_level

150

blood_glucose_level

1

predict

jest chory

co ja robie

run pipeline

load new model

generate synthetic data

gender

1

age

1

hypertension

1

heart_disease

1

smoking history

1

bmi

1

hba1c_level

1

blood_glucose_level

1

predict

jest nie chory

Funkcjonalności

Funkcjonalności czynią projekt kompleksowym narzędziem do analizy ryzyka cukrzycy, dostosowanym do potrzeb badaczy i specjalistów medycznych.

Modelowanie Cukrzycy: Umożliwia tworzenie modeli predykcyjnych dla diagnozowania cukrzycy, wykorzystując różne techniki, takie jak regresja logistyczna, Autogluon i random forest classifier.

Interaktywna Analiza Danych: Notatniki Jupyter umożliwiają analizę danych, eksperymentowanie z modelami i wizualizację wyników.

Interfejs API: Umożliwia łatwy dostęp do modeli przez sieć, umożliwiając użytkownikom wprowadzanie danych i odbieranie prognoz.

Automatyzacja Procesu: Integracja z Kedro zapewnia zarządzanie przepływem pracy, automatyzację pipeline'ów i testowanie.

Architektura Rozwiązania

Projekt składa się z kilku kluczowych komponentów:

- **Integracja Źródeł Danych:** Dane są zbierane z wielu źródeł, w tym elektronicznych kart pacjentów, aplikacji do monitorowania zdrowia i urządzeń IoT. Integracja tych danych jest kluczowa do stworzenia kompleksowego obrazu stanu zdrowia pacjentów.
- **Procesy ETL/ELT:** Zastosowanie zaawansowanych procesów ETL (Extract, Transform, Load) lub ELT (Extract, Load, Transform), które przetwarzają i porządkują dane do analizy. Transformacja obejmuje normalizację, deduplikację i wzbogacenie danych.
- **Magazyn Danych i Jezioro Danych:** Używanie magazynów danych do przechowywania strukturyzowanych danych oraz jezior danych dla niestrukturyzowanych danych. To umożliwia elastyczność w analizie i przechowywaniu dużych ilości danych.
- **Data Science i Machine Learning:** Platforma do nauki o danych i uczenia maszynowego, gdzie dane są wykorzystywane do tworzenia i trenowania modeli predykcyjnych, oraz do eksperymentów i wizualizacji.
- **Interfejs API i Aplikacje:** Dostępne są interfejsy API do komunikacji z modelami uczenia maszynowego oraz aplikacje umożliwiające użytkownikom łatwy dostęp do prognoz i analiz.
- **Konsumenci Danych:** Grupy docelowe korzystające z danych, w tym specjaliści medyczni i badacze, do różnorodnych celów, jak raportowanie operacyjne czy analiza ad hoc.