

UNIWERSYTET RZESZOWSKI
WYDZIAŁ NAUK ŚCIĘŻYCH I TECHNICZNYCH
INSTYTUT INFORMATYKI



Jakub Siłka

*Karol Brudniak
Paweł Powęska*

131509, 125106, 131496

Informatyka

*System predykcji ryzyka cukrzycy typu 2 z wykorzystaniem uczenia
maszynowego i AI*

Projekt Inżynierski

Praca wykonana pod kierunkiem
mgr inż. Wojciech Gałka

Rzeszów 2026

Spis treści

1. Wstęp i Architektura Systemu	7
1.1. Wprowadzenie	7
1.2. Architektura Systemu	7
1.3. Stos Technologiczny	7
2. Przegląd Aktualnych Rozwiązań	9
2.1. Formularz American Diabetes Association.....	9
2.2. Kalkulator platformy HaloDoctor	9
3. Analiza Wymagań	11
3.1. Wymagania funkcjonalne i niefunkcjonalne	11
3.1.1. Wymagania funkcjonalne	11
3.1.2. Wymagania niefunkcjonalne	11
3.2. Diagram przypadków użycia.....	11
3.2.1. Przypadek użycia: Szybka symulacja ryzyka.....	11
3.2.2. Przypadek użycia: Rejestracja	12
3.2.3. Przypadek użycia: Logowanie	12
3.2.4. Przypadek użycia: Symulacja ryzyka oraz porady	12
3.2.5. Przypadek użycia: Przeglądanie historii predykcji.....	13
4. Modele Uczenia Maszynowego	14
4.1. Przegląd Modeli	14
4.2. Dane Treningowe i Preprocessing	14
4.2.1. Cechy wejściowe.....	14
4.2.2. Preprocessing	14
4.3. Szczegółowa Ewaluacja i Problem Niezbalansowania	15
4.3.1. Pułapka Dokładności (Accuracy Paradox)	15
4.4. Implementacja Predykcji	15
5. Wyjaśnialność i Integracja AI	17
5.1. Wyjaśnialność Modeli (XAI) z SHAP.....	17
5.1.1. Implementacja SHAP	17
5.2. Generatywna AI - Google Gemini.....	17
5.2.1. Prompt Engineering	17
6. Implementacja i API.....	19
6.1. Backend (Flask)	19
6.1.1. Struktura API.....	19
6.1.2. Model Danych (SQLAlchemy)	19
6.2. Frontend (React)	19
6.2.1. Widoki Aplikacji	20

6.3. Podsumowanie.....	23
Bibliografia.....	24
Spis rysunków	25
Spis tabel.....	26
Spis listingów	27
Streszczenie.....	28
Oświadczenie studenta o samodzielności pracy	29

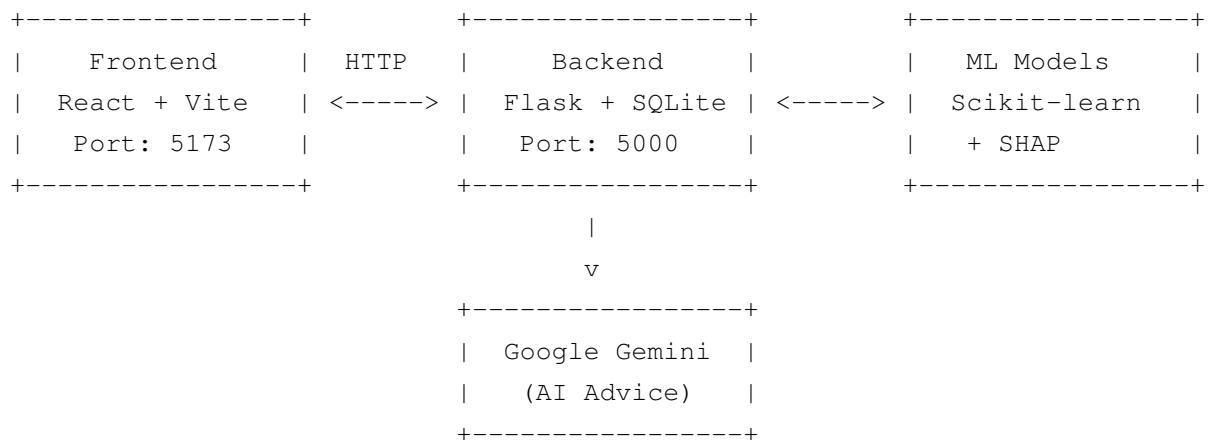
1. Wstęp i Architektura Systemu

1.1. Wprowadzenie

Cukrzyca typu 2 jest jedną z najszybciej rozwijających się chorób cywilizacyjnych. Wczesna diagnoza i świadomość czynników ryzyka są kluczowe dla skutecznej prewencji. Celem projektu *DiabetesRiskPredictor* jest stworzenie dostępnego narzędzia, które na podstawie danych ankietowych użytkownika (takich jak BMI, aktywność fizyczna, nawyki żywieniowe) oszacuje ryzyko zachorowania oraz dostarczy zrozumiałych rekomendacji zdrowotnych.

1.2. Architektura Systemu

System został zaprojektowany w architekturze klient-serwer, z wyraźnym podziałem na warstwę prezentacji, logiki biznesowej oraz modeli predykcyjnych.



Rys. 1.1. Schemat architektury systemu *DiabetesRiskPredictor*

Główne komponenty systemu to:

- **Frontend:** Aplikacja typu Single Page Application (SPA) zbudowana w oparciu o React 19 i Vite. Odpowiada za interakcję z użytkownikiem, zbieranie danych ankietowych oraz prezentację wyników.
- **Backend:** Serwer API napisany w Pythonie (Flask). Obsługuje żądania HTTP, zarządza bazą danych użytkowników, autentykacją (JWT) oraz komunikacją z modelami ML.
- **Modele ML:** Zestaw wytrenowanych modeli klasyfikacyjnych (Scikit-learn) służących do oceny ryzyka.
- **Moduł AI:** Integracja z Google Gemini API w celu generowania spersonalizowanych porad zdrowotnych w języku naturalnym.

1.3. Stos Technologiczny

W tabeli 1.1 przedstawiono szczegółowy stos technologiczny wykorzystany w projekcie.

Tabela 1.1. Technologie wykorzystane w projekcie

Warstwa	Technologia
Frontend	React 19, TypeScript, Vite, TailwindCSS (opcjonalnie)
Backend	Python 3, Flask, SQLAlchemy
Machine Learning	Scikit-learn, Pandas, NumPy, SHAP
Sztuczna Inteligencja	Google Gemini API
Autentykacja	JWT (JSON Web Tokens)
Baza Danych	SQLite (rozwojowa) / PostgreSQL (produkcyjna)

2. Przegląd Aktualnych Rozwiązań

2.1. Formularz American Diabetes Association

Internetowy kalkulatory ryzyka cukrzycy - formularz American Diabetes Association jest jednym z najczęściej spotykanych rozwiązań pozwalających na szybką samoocenę stanu zdrowia. Umożliwiają one wprowadzenie podstawowych biometrycznych, takich jak wiek, waga czy wzrost, w celu oszacowania ryzyka zachorowania w oparciu o sztywne reguły punktowe.

The image shows two screenshots of the American Diabetes Association's Type 2 Risk Test. Part (a) Część 1 is the initial form where users input their email, ZIP code, and United States. It includes a checkbox for accepting terms and conditions and a button to 'Email My Results'. Part (b) Część 2 is the results page showing a 'Low Risk' score of 4/10, a progress bar, and a summary message about risk prevention. It also includes links to 'Learn about the ADA Type 2 Diabetes Risk Test' and 'Diabetes Basics'.

Rys. 2.1. Formularz American Diabetes Association

Do głównych zalet wymienionych wyżej rozwiązań należą:

- **Szybkość i prostota:** Użytkownik otrzymuje wynik natychmiast po wypełnieniu krótkiej ankiety.
- **Dostępność:** Brak konieczności zakładania konta czy logowania.
- **Wiarygodność źródeł:** Bazują na ustandaryzowanych wytycznych medycznych.

Minusy tego formularza to:

- **Brak personalizacji i wyjaśnienia:** Wynik jest liczbą lub statusem bez wyjaśnienia, który czynnik założył na diagnozie.
- **Brak historii:** Brak możliwości zapisu wyników i śledzenia zmian w czasie.
- Formularz na ekranie końcowym zniechęcający użytkowników do sprawdzenia wyników.

2.2. Kalkulator platformy HaloDoctor

Drugim analizowanym rozwiązaniem jest kalkulator ryzyka cukrzycy typu 2 dostępny na platformie telemedycznej HaloDoctor. Narzędzie to oparte jest na międzynarodowej skali FINDRISC (Finnish Diabetes Risk

Score) i służy do oszacowania ryzyka zachorowania w perspektywie najbliższych 10 lat. Formularz składa się z ośmiu kluczowych pytań, które łączą parametry ilościowe (BMI, obwód talii) z wywiadem dotyczącym stylu życia (aktywność fizyczna, spożycie warzyw/owoców) oraz historią medyczną (nadciśnienie, hiperglikemia, obciążenie genetyczne).

Rys. 2.2. Kalkulator ryzyka HaloDoctor

Do głównych zalet tego rozwiązania należą:

- Integracja z usługami medycznymi:** Jako część platformy telemedycznej, narzędzie często sugeruje bezpośrednią konsultację lekarską w przypadku wysokiego wyniku, lub w celu prawdziwej diagnozy.
- Przejrzysta skala punktowa:** Wynik jest sumowany w zakresie 0-24 punktów, co pozwala na łatwe przypisanie pacjenta do jednej z czterech grup ryzyka.

Minusy kalkulatora to:

- Subiektywność danych:** Pytania o dietę (np. "Czy jadasz warzywa codziennie?") opierają się na deklaracji użytkownika, co może prowadzić do zafałszowania wyniku.
- Bariera wejścia:** Wymóg dokładnego zmierzenia obwodu w pasie może być problematyczny dla użytkownika korzystającego z aplikacji "w biegu".
- Sztywny algorytm:** Podobnie jak w innych kalkulatorach opartych na FINDRISC, wagę poszczególnych pytań są stałe i nie dostosowują się dynamicznie do specyficznych kombinacji cech pacjenta.

3. Analiza Wymagań

3.1. Wymagania funkcjonalne i niefunkcjonalne

3.1.1. Wymagania funkcjonalne

- Wprowadzanie danych diagnostycznych i metabolicznych (wiek, BMI, ciśnienie, cholesterol, poziom aktywności...) poprzez formularz.
- Wyświetlanie wyniku predykcji ryzyka wystąpienia cukrzycy (klasyfikacja: zdrowy, stan przedcukrzowy, cukrzyca) w formie graficznej.
- Generowanie i wyświetlanie inteligentnych zaleceń zdrowotnych ("Smart Advisor") dopasowanych do wyniku użytkownika.
- Przeglądanie historii wykonanych predykcji oraz wprowadzonych logów.
- Możliwość wykonania szybkiej symulacji ryzyka dla użytkownika niezalogowanego.
- Możliwość utworzenia nowego konta użytkownika.
- Możliwość zalogowania się w aplikacji.
- Możliwość usunięcia błędnych wpisów z dziennika zdrowia.
- Możliwość edycji profilu użytkownika (dane stałe, np. wzrost do obliczeń BMI).

3.1.2. Wymagania niefunkcjonalne

- Wymagany stały dostęp do Internetu w celu komunikacji z API predykcyjnym.
- Czas odpowiedzi modelu predykcyjnego poniżej 1 sekundy.
- Skuteczność modelu uczenia maszynowego na poziomie minimum 80%.
- Bezpieczne przechowywanie haseł oraz loginów użytkowników w bazie danych.

3.2. Diagram przypadków użycia

3.2.1. Przypadek użycia: Szybka symulacja ryzyka

Aktor: Gość

Opis: Wykonanie szybkiej analizy ryzyka cukrzycy bez konieczności logowania się.

Warunki wstępne: Użytkownik niezalogowany, otwarty ekran startowy aplikacji.

Przebieg:

1. Gość wybiera opcję „Oblicz ryzyko cukrzycy”.
2. System wyświetla formularz do diagnozy.
3. Gość uzupełnia formularz i zatwierduje go.
4. System przetwarza dane i wyświetla diagnozę.
5. System wyświetla interpretację wyniku.

3.2.2. Przypadek użycia: Rejestracja

Aktor: Gość

Opis: Utworzenie nowego konta użytkownika.

Warunki wstępne: Użytkownik nie posiada konta lub nie jest zalogowany.

Przebieg:

1. Gość wybiera opcję rejestracji.
2. System wyświetla formularz do rejestracji.
3. Gość podaje wymagane dane i zatwierdza wybór.
4. System weryfikuje poprawność danych i tworzy konto w bazie.
5. Następuje automatyczne logowanie użytkownika.

3.2.3. Przypadek użycia: Logowanie

Aktor: Gość

Opis: Uwierzytelnienie użytkownika w systemie.

Warunki wstępne: Użytkownik posiada konto, ale nie jest zalogowany.

Przebieg:

1. Gość wybiera opcję Logowania.
2. System wyświetla formularz logowania.
3. Gość wprowadza login i hasło.
4. System weryfikuje zgodność danych z bazą bezpiecznie przechowywanych haseł.
- 5a. Dane są poprawne – użytkownik uzyskuje dostęp do aplikacji jako użytkownik zalogowany.
- 5b. Dane są błędne – system wyświetla komunikat o błędzie.
- 6b. Powrót do kroku 3.

3.2.4. Przypadek użycia: Symulacja ryzyka oraz porady

Aktor: Użytkownik Zalogowany

Opis: Wykonanie analizy ryzyka cukrzycy z trwałym zapisem wyniku w historii użytkownika.

Warunki wstępne: Użytkownik zalogowany w aplikacji.

Przebieg:

1. Użytkownik wybiera opcję „Oblicz ryzyko cukrzycy”.
2. System wyświetla formularz do diagnozy.
3. Użytkownik uzupełnia formularz i zatwierdza go.
4. System przetwarza dane i wyświetla diagnozę wraz z poradami.
5. System zapisuje dane w bazie danych.

3.2.5. Przypadek użycia: Przeglądanie historii predykcji

Aktor: Użytkownik Zalogowany

Opis: Przeglądanie zapisanych wyników badań.

Warunki wstępne: Użytkownik zalogowany, posiada w bazie co najmniej jeden zapisany wynik.

Przebieg:

1. Użytkownik wybiera moduł historii.
2. System pobiera historię badań zdrowotnych z bazy danych.
3. System wyświetla listę wykonanych predykcji.
4. Użytkownik może wybrać konkretny wpis, aby zobaczyć jego szczegóły.

4. Modele Uczenia Maszynowego

4.1. Przegląd Modeli

W celu zapewnienia wysokiej wiarygodności predykcji, system wykorzystuje zespół trzech różnych modeli klasyfikacyjnych działających równolegle. Pozwala to na porównanie wyników i zwiększenie pewności diagnozy.

Wykorzystane algorytmy to:

1. **Logistic Regression:** Model liniowy, służący jako punkt odniesienia (baseline). Charakteryzuje się wysoką interpretowalnością. W eksperymentach osiągnął dokładność 64.33% przy najkrótszym czasie treningu (0.27s).
2. **Random Forest:** Zespół drzew decyzyjnych. Model ten jest odporny na overfitting. Osiągnął dokładność 79.45% w czasie treningu wynoszącym 25.41s.
3. **Gradient Boosting:** Zaawansowana metoda ensemble. Zapewniała najwyższą ogólną dokładność (84.9%), jednak odbyło się to kosztem najdłuższego czasu treningu (75.34s).

Modele zostały skonfigurowane do klasyfikacji wieloklasowej (0 - brak cukrzycy, 1 - stan przedcukrzycowy, 2 - cukrzyca) lub binarnej, w zależności od konfiguracji treningowej. W obecnej wersji system wspiera klasyfikację w 3 klasach.

4.2. Dane Treningowe i Preprocessing

Modele zostały wytrenowane na zbiorze danych pochodzący z CDC BRFSS (Behavioral Risk Factor Surveillance System), zawierającym ok. 250,000 rekordów [1].

4.2.1. Cechy wejściowe

Każdy model przyjmuje na wejściu wektor 18 cech opisujących stan zdrowia i styl życia pacjenta.

Tabela 4.1. Cechy wejściowe modelu

YY

Kategoria	Cecha	Opis
Demografia	Sex, Age	Płeć i wiek (kategorie)
Badania	BMI, HighBP, HighChol	Wskaźnik masy ciała, nadciśnienie, wysoki cholesterol
Zdrowie	GenHlth, PhysHlth, MentHlth	Ogólna ocena zdrowia, dni złego samopoczucia fiz./psych.
Choroby	Stroke, HeartDiseaseorAttack	Przebyty udar, choroby serca
Nawyki	Smoker, Alcohol, Fruits, Veggies	Palenie, alkohol, dieta
Inne	PhysActivity, DiffWalk	Aktywność fizyczna, trudności w chodzeniu
Opieka	AnyHealthcare, NoDocbcCost	Dostęp do opieki, koszt wizyt

4.2.2. Preprocessing

Przed podaniem danych do modeli zastosowano następujące kroki przetwarzania wstępnego:

1. **Czyszczenie danych:** Usunięcie brakujących wartości i duplikatów.
2. **Skalowanie:** Zastosowano StandardScaler do normalizacji cech numerycznych (np. BMI), aby sprowadzić je do wspólnej skali (średnia 0, odchylenie standardowe 1).

3. **Balansowanie klas:** Zbiór treningowy był niebalansowany (znacznie więcej osób zdrowych). Zastosowano parametr `class_weight='balanced'` w modelach oraz techniki oversamplingu (podczas eksperymentów) w celu wyrównania szans dla klas mniejszościowych (cukrzyca).

4.3. Szczegółowa Ewaluacja i Problem Niebalansowania

Modele zostały zweryfikowane na zbiorze testowym liczącym 50 736 próbek. Kluczowym wyzwaniem w tym zbiorze danych jest silne niebalansowanie klas – zdecydowana większość próbek (ok. 84%) to osoby zdrowe (klasa 0).

4.3.1. Pułapka Dokładności (Accuracy Paradox)

Analizując wyniki, można zauważać, że **Gradient Boosting** osiągnął najwyższą dokładność (84.9%). Jest to jednak wynik mylący. Model ten zoptymalizował się pod klasę większościową, niemal całkowicie ignorując klasy rzadkie, które są najważniejsze z medycznego punktu widzenia.

- Recall dla stanu przedcukrzycowego (klasa 1) w Gradient Boosting wyniósł 0.00 (model w ogóle jej nie wykrywa).
- Recall dla cukrzyca (klasa 2) wyniósł jedynie 0.18.

Dlatego w systemach medycznych ważniejsza od ogólnej dokładności jest zdolność do detekcji choroby (Czułość/Recall).

Logistic Regression, mimo najniższej ogólnej dokładności (64.33%), wykazał się relatywnie najlepszą zdolnością do wykrywania trudnych przypadków (Recall dla klasy 2 wynosi 0.59, a dla klasy 1 wynosi 0.30). Oznacza to, że częściej generuje fałszywe alarmy (niższa precyzja), ale rzadziej przeoczy osobę chorą.

Random Forest stanowi kompromis, oferując lepszą precyzję niż regresja i znacznie lepszą wykrywalność chorób niż Gradient Boosting.

Tabela 4.3. Porównanie zdolności detekcji Cukrzycy (Klasa 2) - klasy rzadkiej

Model	Precision	Recall (Czułość)	F1-Score
Logistic Regression	0.35	0.59	0.44
Random Forest	0.38	0.48	0.43
Gradient Boosting	0.56	0.18	0.27

Podane czasy (0.27s vs 75.34s) odnoszą się do procesu **trenowania** modelu, a nie predykcji. Czas predykcji dla pojedynczego pacjenta we wszystkich modelach jest liczony w milisekundach i nie stanowi wąskiego gardła. Zastosowanie trzech modeli (Ensemble) w aplikacji pozwala zniwelować słabości pojedynczych algorytmów.

4.4. Implementacja Predykcji

Proces predykcji w systemie przebiega następująco (listing 4.1):

Listing 4.1. Funkcja predykcji w backendzie

```

1 def predict_diabetes_risk(data):
2     # 1. Konwersja danych do DataFrame
3     input_df = pd.DataFrame([data])
4

```

```
5     # 2. Skalowanie danych (uzycie zapisanego scalera)
6     input_scaled = scaler.transform(input_df)
7
8     # 3. Iteracja przez modele
9     results = {}
10    for name, model in models.items():
11        # Predykcja klasy i prawdopodobienstw
12        prediction = model.predict(input_scaled) [0]
13        probs = model.predict_proba(input_scaled) [0]
14
15        results[name] = {
16            'prediction': int(prediction),
17            'probabilities': probs.tolist()
18        }
19
20    return results
```

5. Wyjaśnialność i Integracja AI

5.1. Wyjaśnialność Modeli (XAI) z SHAP

Modele uczenia maszynowego często działają jak "czarne skrzynki". Aby zwiększyć zaufanie użytkownika do diagnozy, system implementuje metody XAI (Explainable AI), konkretnie bibliotekę SHAP (SHapley Additive exPlanations).

Dla każdej predykcji system oblicza wartości SHAP, które pokazują, jak każda cecha wpłynęła na wynik końcowy (zwiększając lub zmniejszając ryzyko).

5.1.1. Implementacja SHAP

Analiza odbywa się w czasie rzeczywistym po wykonaniu predykcji:

Listing 5.1. Generowanie wyjaśnień SHAP

```
1 def get_shap_explanation(model, input_scaled_df):
2     # Utworzenie explainera dla modelu drzewiastego
3     explainer = shap.TreeExplainer(model)
4     shap_values = explainer.shap_values(input_scaled_df)
5
6     # Przetwarzanie wyników...
7     # Identifikacja czynników ryzyka (wpływ dodatni)
8     # Identifikacja czynników ochronnych (wpływ ujemny)
9
10    return risk_factors, protective_factors
```

Wyniki analizy SHAP są przekazywane do modelu językowego w celu wygenerowania bardziej precyzyjnych i spersonalizowanych zaleceń. Dzięki temu porada zdrowotna uwzględnia konkretne czynniki, które zaważyły na wyniku, takie jak BMI czy nadciśnienie.

5.2. Generatywna AI - Google Gemini

Oprócz twardych danych liczbowych, system oferuje "ludzką" poradę generowaną przez duży model językowy (LLM). Projekt wykorzystuje API Google Gemini.

5.2.1. Prompt Engineering

Kluczem do uzyskania wartościowych porad jest odpowiednio skonstruowany prompt. Aplikacja dynamicznie buduje zapytanie do modelu, zawierające:

1. Dane pacjenta (wiek, płeć, wyniki badań).
2. Wynik predykcji (ryzyko cukrzycy).
3. Kontekst roli ("Jesteś asystentem medycznym...").
4. Ograniczenia ("Nie stawiaj ostatecznej diagnozy, sugeruj konsultację lekarską").

Listing 5.2. Integracja z Gemini API

```
1 def generate_llm_advice(data, prediction_result):
```

```
2     prompt = f"""
3     Jako ekspert medyczny, przeanalizuj nastepujacy przypadek:
4     Pacjent: Kobieta, 45 lat, BMI 28.
5     Wynik modelu ML: Wysokie ryzyko cukrzycy (85%).
6
7     Podaj 3 konkretne kroki, ktore pacjent moze podjac,
8     aby zmniejszyc ryzyko. Uzywaj empatycznego jezyka.
9     """
10
11    response = model.generate_content(prompt)
12    return response.text
```

Dzięki temu użytkownik otrzymuje nie tylko suchy wynik "Ryzyko: Wysokie", ale także spersonalizowany plan działania (np. "ZwiększM aktywność fizyczną do 30 min dziennie, skonsultuj poziom cukru z lekarzem POZ").

6. Implementacja i API

6.1. Backend (Flask)

Backend aplikacji pełni rolę huba integrującego bazę danych, modele ML oraz zewnętrzne API. Został zrealizowany w mikro-frameworku Flask, co zapewnia lekkość i łatwość rozbudowy.

6.1.1. Struktura API

Główne endpointy API to:

- POST /api/predict - Główny endpoint. Przyjmuje JSON z danymi ankiety, zwraca wyniki ze wszystkich 3 modeli, analizę SHAP oraz poradę AI.
- POST /api/auth/register - Rejestracja nowego użytkownika.
- POST /api/auth/login - Logowanie (zwraca token JWT access + refresh).
- GET /api/user/history - Pobranie historii predykcji zalogowanego użytkownika.

6.1.2. Model Danych (SQLAlchemy)

Baza danych przechowuje informacje o użytkownikach oraz historię ich badań. Wykorzystano ORM SQLAlchemy.

Relacje:

- User (1) \longleftrightarrow (N) PredictionResult
- User (1) \longleftrightarrow (1) UserProfile (dane demograficzne)

6.2. Frontend (React)

Warstwa prezentacji została zbudowana jako Single Page Application. Wykorzystuje:

- **React Router:** Do obsługi nawigacji bez przeładowywania strony.
- **Axios:** Do komunikacji z API.
- **Context API:** Do zarządzania stanem sesji użytkownika (autentykacja).

Proces badania składa się z 3 kroków (formularz wieloetapowy): 1. Podstawowe dane (Wiek, Płeć). 2. Parametry zdrowotne (BMI, Nadciśnienie). 3. Styl życia (Dieta, Używki).

Poniżej przedstawiono fragment komponentu formularza obsługującego wysyłkę danych:

Listing 6.1. Obsługa formularza w React

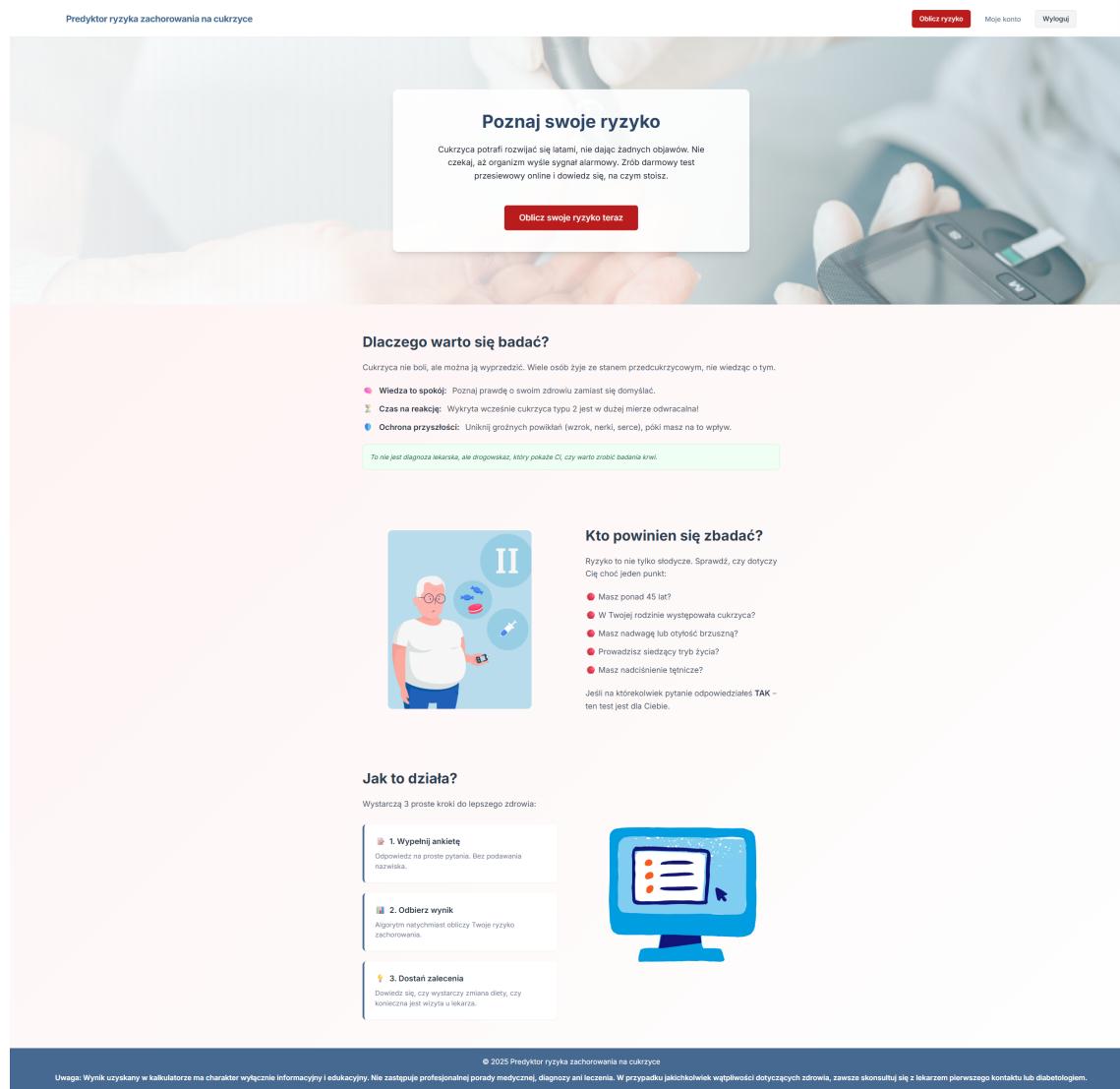
```
1 const handleSubmit = async (formData) => {
2     try {
3         const response = await api.post('/predict', formData);
4         setResult(response.data);
5         navigate('/results');
6     } catch (error) {
7         console.error("Blad predykcji:", error);
8     }
9 };
```

6.2.1. Widoki Aplikacji

Interfejs użytkownika został zaprojektowany z myślą o przejrzystości i łatwości obsługi. Poniżej przedstawiono kluczowe ekran aplikacji.

6.2.1.1. Strona Główna i Formularz

Strona główna wita użytkownika i wyjaśnia cel aplikacji. Głównym elementem jest wieloetapowy formularz (Wizard), który prowadzi pacjenta przez proces wprowadzania danych.



Rys. 6.1. Interfejs strony głównej

Kroki formularza zostały zaprojektowane tak, aby były intuicyjne i nie przytłaczały ilością pytań na jednym ekranie.

Predyktator ryzyka zachorowania na cukrzycę

Moje konto Wyloguj

1 TO SIE STAN ZDROWIA **2** NAVYKI **3**

Podstawowe informacje
Wypełnij podstawowe informacje.

Jaka jest Twoja płeć?
 Kobieta Mężczyzna

Wybierz przedział wiekowy

18-24	25-29
30-34	35-39
40-44	45-49
50-54	55-59
60-64	65-69
70-74	75-79
ponad 80	

Podaj swój wzrost w cm
Wpisz wartość cm

Jaka jest Twoja waga?
Wpisz wartość kg

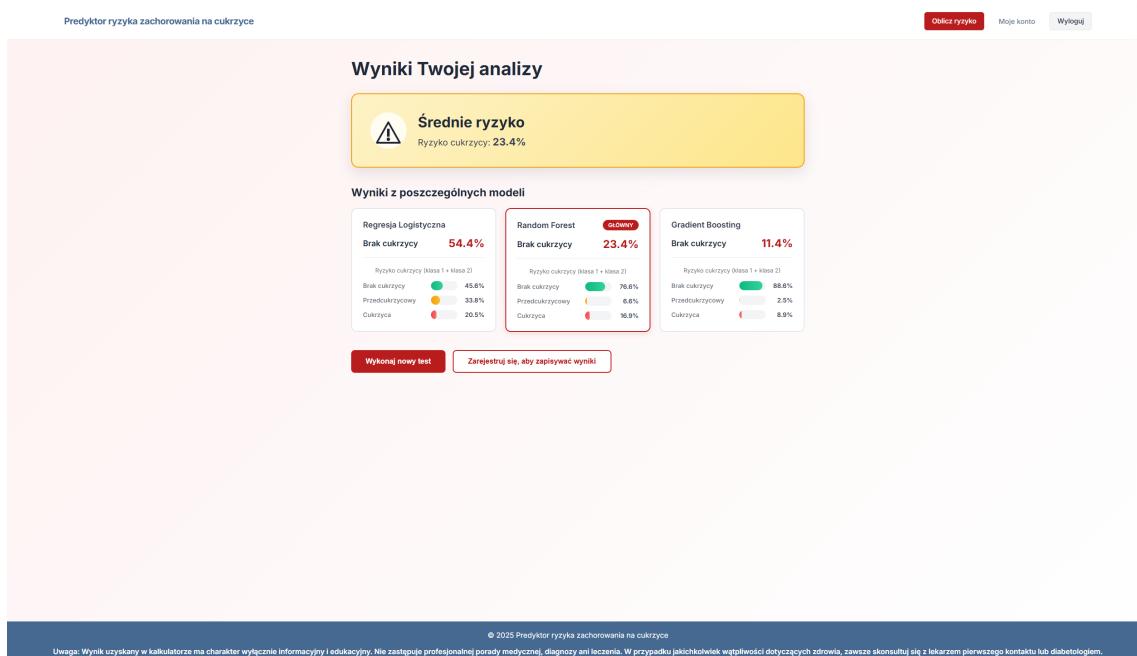
© 2025 Predyktator ryzyka zachorowania na cukrzycę

Uwaga: Wynik uzyskany w kalkulatorze ma charakter wyłącznie informacyjny i edukacyjny. Nie zastępuje profesjonalnej porady medycznej, diagnozy ani leczenia. W przypadku jakichkolwiek wątpliwości dotyczących zdrowia, zawsze skonsultuj się z lekarzem pierwszego kontaktu lub diabetologiem.

Rys. 6.2. Widok kroków formularza diagnostycznego

6.2.1.2. Prezentacja Wyników

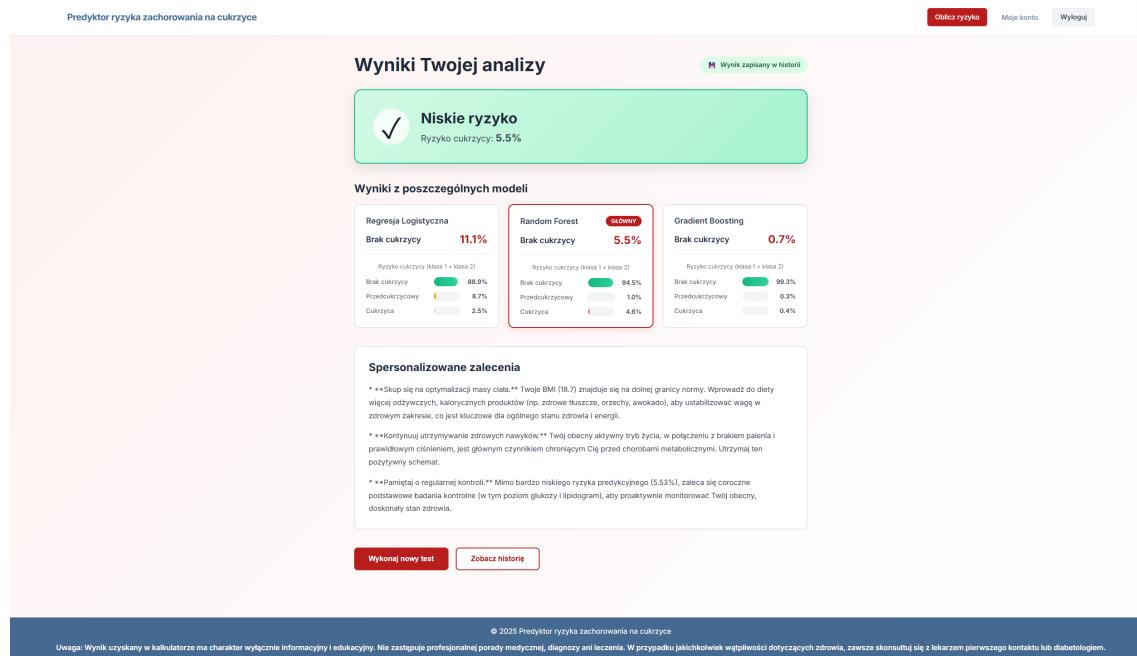
Po przetworzeniu danych, użytkownik otrzymuje czytelny wynik w formie graficznej (wskaźnik ryzyka) oraz tekstuowej. Kolorystyka (zielony/żółty/czerwony) natychmiastowo informuje o poziomie zagrożenia.



Rys. 6.3. Ekran wyników predykcji z wykresem ryzyka

6.2.1.3. Analiza Szczegółowa (SHAP i AI)

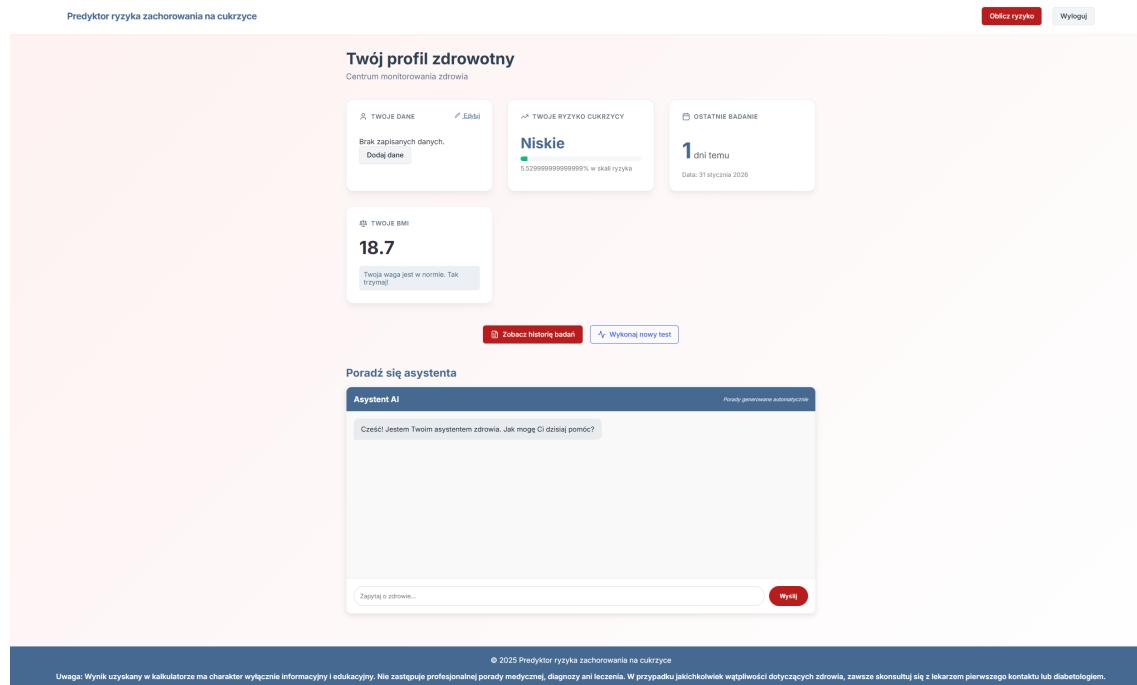
Dla bardziej docieklejnych użytkowników dostępna jest sekcja szczegółowa, zawierająca spersonalizowaną poradę wygenerowaną przez asystenta AI.



Rys. 6.4. Szczegółowa analiza czynników ryzyka i porada AI

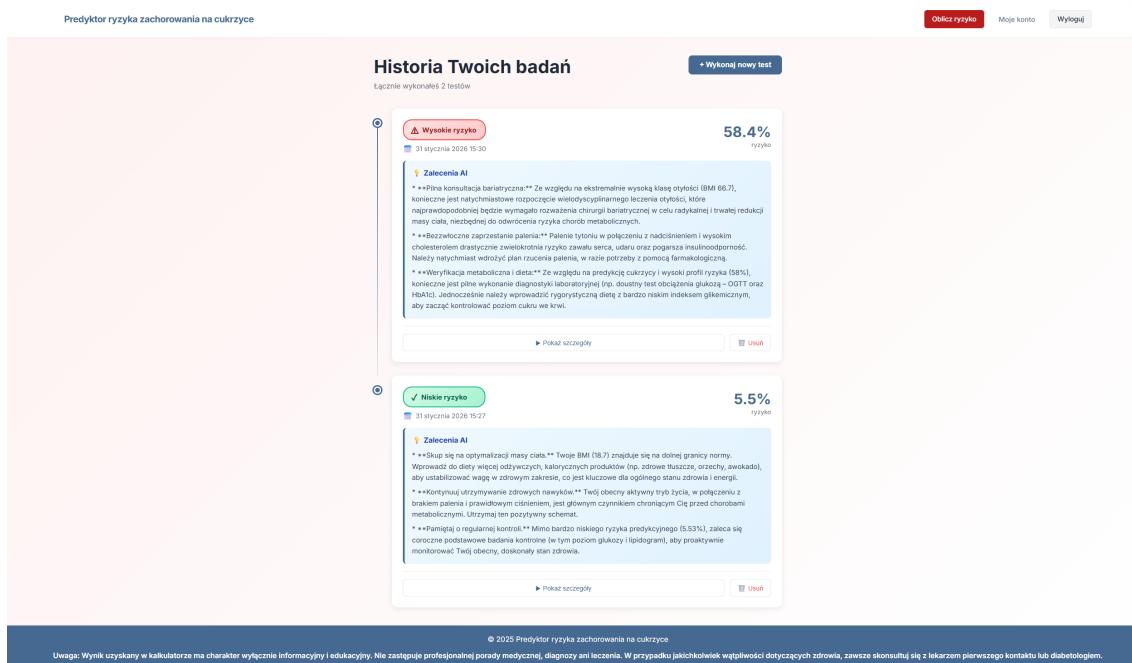
6.2.1.4. Panel Użytkownika i Historia

Zalogowani użytkownicy mają dostęp do panelu, w którym mogą zarządzać swoim profilem oraz przeglądać historię wykonanych badań.



Rys. 6.5. Panel użytkownika

Sekcja historii pozwala na śledzenie zmian ryzyka w czasie, co jest kluczowe dla prewencji długoterminowej.



Rys. 6.6. Widok historii predykcji

6.3. Podsumowanie

Stworzony system *DiabetesRiskPredictor* stanowi kompleksowe rozwiązanie demonstrujące praktyczne zastosowanie uczenia maszynowego w medycynie prewencyjnej. Połączenie klasycznych modeli klasyfikacyjnych z nowoczesną generatywną AI pozwala na dostarczenie użytkownikowi wartościowej i zrozumiałej informacji zwrotnej.

Bibliografia

- [1] Alex Teboul. Diabetes health indicators dataset, 2021. Accessed: 2026-01-26. URL: <https://www.kaggle.com/datasets/alextreboul/diabetes-health-indicators-dataset>.

Spis rysunków

1.1	Schemat architektury systemu DiabetesRiskPredictor	7
2.1	Formularz American Diabetes Association	9
2.2	Kalkulator ryzyka HaloDoctor	10
6.1	Interfejs strony głównej	20
6.2	Widok kroków formularza diagnostycznego	21
6.3	Ekran wyników predykcji z wykresem ryzyka	21
6.4	Szczegółowa analiza czynników ryzyka i porada AI	22
6.5	Panel użytkownika	22
6.6	Widok historii predykcji	23

Spis tabel

1.1	Technologie wykorzystane w projekcie	8
4.1	Cechy wejściowe modelu	14
4.3	Porównanie zdolności detekcji Cukrzycy (Klasa 2) - klasy rzadziej	15

Spis listingów

4.1	Funkcja predykcji w backendzie	15
5.1	Generowanie wyjaśnień SHAP	17
5.2	Integracja z Gemini API	17
6.1	Obsługa formularza w React	19

Streszczenie

Niniejszy projekt, *DiabetesRiskPredictor*, to nowoczesna aplikacja webowa służąca do przewidywania ryzyka wystąpienia cukrzycy typu 2. System integruje zaawansowane modele uczenia maszynowego (Logistic Regression, Random Forest, Gradient Boosting) z interaktywnym interfejsem użytkownika. Aplikacja nie tylko klasyfikuje ryzyko, ale również dostarcza spersonalizowanych wyjaśnień wyników dzięki wykorzystaniu metodologii SHAP (SHapley Additive exPlanations) oraz generatywnej sztucznej inteligencji (Google Gemini API).

Część serwerowa (Backend) została zaimplementowana w języku Python z użyciem frameworka Flask, natomiast część kliencka (Frontend) wykorzystuje bibliotekę React. Projekt kładzie duży nacisk na wyjaśnialność (XAI) oraz edukację użytkownika w zakresie profilaktyki zdrowotnej.

Załącznik nr 2 do Zarządzenia nr 228/2021 Rektora Uniwersytetu Rzeszowskiego z dnia 1 grudnia 2021 roku w sprawie ustalenia procedury antyplagiatowej w Uniwersytecie Rzeszowskim

OŚWIADCZENIE STUDENTA O SAMODZIELNOŚCI PRACY

.....Jakub Siłka

Karol Brudniak

Paweł Poweska

Imię (imiona) i nazwisko studenta

Wydział Nauk Ścisłych i Technicznych

.....Informatyka.....

Nazwa kierunku

.....131509, 125106, 131496.....

Numer albumu

1. Oświadczam, że moja praca projektowa pt.: System predykcji ryzyka cukrzycy typu 2 z wykorzystaniem uczenia maszynowego i AI

- 1) została przygotowana przeze mnie samodzielnie*,
2) nie narusza praw autorskich w rozumieniu ustawy z dnia 4 lutego 1994 roku o prawie autorskim i prawach pokrewnych (t.j. Dz.U. z 2021 r., poz. 1062) oraz dóbr osobistych chronionych prawem cywilnym,
3) nie zawiera danych i informacji, które uzyskałem/am w sposób niedozwolony,
4) nie była podstawą otrzymania oceny z innego przedmiotu na uczelni wyższej ani mnie, ani innej osobie.

2. Jednocześnie wyrażam zgodę/nie wyrażam zgody** na udostępnienie mojej pracy projektowej do celów naukowo–badawczych z poszanowaniem przepisów ustawy o prawie autorskim i prawach pokrewnych.

(miejscowość, data)

(czytelny podpis studenta)

* Uwzględniając merytoryczny wkład prowadzącego przedmiot

** – niepotrzebne skreślić