

Career Navigator | Anna Wolańska, Arkadiusz Jaciubek

Raport stanu prac (24.05.2024)

Link do repozytorium:

https://github.com/ajaciubek/issi_proj

Opis projektu

Projekt "Career Navigator" to aplikacja oparta na FastAPI, która umożliwia przewidywanie ról zawodowych na podstawie podanych umiejętności. Wykorzystuje modele uczenia maszynowego do analizy danych i generowania prognoz. Projekt jest skonfigurowany do pracy w środowisku Python 3.12 i integruje się z narzędziami takimi jak DVC, AWS S3 oraz pre-commit.

Struktura projektu

Kluczowe foldery i pliki:

- api: Zawiera logikę aplikacji FastAPI.
 - app.py: Główna aplikacja FastAPI.
 - model.py: Definicje modeli danych dla żądań i odpowiedzi.
- ml: Moduły związane z uczeniem maszynowym.
 - predictor.py: Funkcja predict_job_role do przewidywania ról zawodowych.
 - settings.py: Konfiguracja ścieżek do modeli i innych zasobów.
- data: Przechowuje dane i modele.
 - programming_skills.json: Lista umiejętności programistycznych.
 - Pliki .dvc (np. job_model.keras.dvc): Śledzenie modeli i danych za pomocą DVC.
- config: Pliki konfiguracyjne.
 - .ml-env: Zmienna środowiskowa z konfiguracją modeli.
- test: Testy aplikacji.
 - test_app.http: Testy HTTP dla endpointów FastAPI.

Funkcjonalności

1. *Endpointy API:*
 - GET /: Zwraca komunikat powitalny. W następnym etapie zostanie przygotowany do zwracania aplikacji frontendowej.
 - POST /predict: Przyjmuje listę umiejętności i limit wyników, zwraca przewidywane role zawodowe z prawdopodobieństwami.
2. *Uczenie maszynowe:*
 - Model TensorFlow do przewidywania ról zawodowych.
 - SentenceTransformer do enkodowania tekstu.
 - LabelEncoder do dekodowania etykiet ról zawodowych.
3. *Zarządzanie danymi:*

- DVC do śledzenia wersji danych i modeli.
- AWS S3 jako zdalne repozytorium danych.

Postęp prac

1. *Kod aplikacji:*
 - Główne funkcjonalności API są zaimplementowane i działają.
 - Moduły uczenia maszynowego są skonfigurowane i zintegrowane z aplikacją.
2. *Dane i modele:*
 - Modele i dane są śledzone za pomocą DVC.
 - Pliki .dvc wskazują na poprawną konfigurację zdalnego repozytorium S3.
3. *Testy:*
 - Testy HTTP dla endpointów są przygotowane w test_app.http.
4. *Konfiguracja:*
 - Plik pyproject.toml definiuje zależności projektu.
 - Pre-commit hooks są skonfigurowane do analizy kodu (Ruff, Xenon).
5. *Dokumentacja:*
 - Plik README.md zawiera podstawowe instrukcje uruchamiania projektu.


Wymagania systemowe

- *Python:* Wersja 3.12.
- *Narzędzia:* Git, AWS CLI, DVC, Uvicorn.

Prace w trakcie

1. *Backend:*
 - Uzgodnienie komunikacji za pośrednictwem API i ostatecznej wersji endpointów.
2. *Dane i modele:*
 - Nowa baza danych oparta na agregatorze ofert pracy Jooble została oczyszczona (po oczyszczeniu dataset zawiera 100k rekordów).
 - Rekordom zostały przypisane etykiety za pomocą LLM.
 - Umiejętności są podzielone na kilka kategorii, każda ma przypisanych kilka pozycji zawodowych.
 - Został przygotowany nowy model, ale konieczny jest fine-tuning dla poprawy rezultatów, eksperymentujemy z pre-processingiem w celu poprawy wyników.
3. *Frontend:*
 - Makiety serwisu zostały zaprojektowane przy użyciu programu Figma. Są dostępne pod linkiem.

- Została rozpoczęta praca nad frontendową częścią aplikacji.



Find your path

Python

Polars

Pandas


pyTorch

PHP

Perl

Show directions

Możliwość wprowadzenia umiejętności (wybór z listy rozwijanej).



Find your path


Show directions

python X

api X


pandas X

Hey! Could you tell me a bit more about your skills?
That way, I can recommend roles that are a better match for what you're great at.



Find your path

It looks like your skills (python, api, pandas, machine learning) could fit into a few different areas. Which one do you think you'd feel most comfortable in?



Backend







Data science

Databases

Pozyskiwanie informacji od użytkownika w sposób przypominający rozmowę z doradcą zawodowym.

Perfect match!

Here are some roles that match your skills best. Click to see details, and give a thumbs up to the ones you like—it helps me learn and improve future matches!

Machine Learning Engineer (96% match)	 
<ul style="list-style-type: none">✓ Python✓ Machine Learning✓ Pandas✓ pyTorch✓ API✗ TensorFlow	86% of jobs ads features that skill
Python Developer (81% match)	 
MLOps Engineer (67% match)	 

Przedstawienie najlepszych dopasowań wraz z brakującymi umiejętnościami oraz ich popularnością w ofertach.

Następne kroki

1. *Rozszerzenie testów:*
 - Dodanie testów jednostkowych i integracyjnych.
2. *Optymalizacja modeli:*
 - Sprawdzenie wydajności modelu TensorFlow i SentenceTransformer.
3. *Rozbudowa dokumentacji:*
 - Szczegółowe instrukcje konfiguracji i wdrożenia.
4. *Integracja CI/CD:*
 - Rozszerzenie workflow GitHub Actions o testy i wdrożenie.