

Organizacja i infrastruktura projektu v 1.0.1

1. Opis projektu i produktu

Nazwa projektu/produktu: ActiveLabel - Platforma do oznaczania danych dla projektów uczenia maszynowego

Adresowany problem: Oznaczanie zbiorów danych na potrzeby trenowania modeli uczenia maszynowego jest czasochłonne i kosztowne. Wdrożenie metod active learning pozwoli na optymalizację procesu poprzez wybór przez model najbardziej istotnych danych do oznaczenia. Integracja wstępnej anotacji przez model dodatkowo przyspieszy procesy oznaczania danych.

Obszar zastosowania: Projekty uczenia maszynowego wymagające oznaczania danych.

Rynek: Startupy AI, zespoły badawczo-rozwojowe, koła naukowe uczenia maszynowego.

Interesariusze:

- Programiści ML
- Menadżerowie projektów MLowych
- DevOps Engineer
- Anotatorzy
- Właściciele danych
- Państwo
- Opiekun projektu

Użytkownicy i ich potrzeby: Potrzeby użytkowników:

- **Programiści ML:**
 - Dostęp do kompleksowych narzędzi integrujących modele AI w system anotacji danych
 - Łatwość ściągnięcia oznaczeń danych z systemu
 - Łatwość wgrywania nowych danych do systemu
- **Anotatorzy:**
 - Intuicyjny, przyjazny interfejs do efektywnego oznaczania danych
 - Automatyczne sugestie wynikające z analiz AI wspomagające proces anotacji
- **Menadżerowie projektów MLowych:**
 - Narzędzia do monitorowania postępów oznaczania danych
 - Podgląda na wydajność poszczególnych anotatorów
 - Minimalizacja kosztów oznaczania zarówno czasu roboczego ludzi jak i zużycia zasobów komputerowych
 - Bezpieczeństwo oznaczonych danych w razie awarii

Cel i zakres produktu:

- Rozwój oprogramowania umożliwiającego oznaczanie danych
- Wsparcie dla metod aktywnego uczenia
- Integracja z istniejącymi modelami AI
- Działanie w środowisku rozproszonym
- Wsparcie dla pracy zespołowej

Ograniczenia:

- Konieczność zapewnienia bezpieczeństwa danych
- Korzystanie z narzędzi otwartoźródłowych z licencjami umożliwiającymi modyfikację kodu

Inne współpracujące systemy:

- Modele uczenia maszynowego oparte na bibliotece PyTorch

Termin: 6 miesięcy

Główne etapy projektu:

1. Przegląd dostępnych systemów wspomagania etykietowania
2. Analiza wymagań
3. Projektowanie architektury systemu
4. Implementacja wstępnej wersji projektu
5. Integracja z systemami zewnętrznymi
6. Testowanie
7. Wdrożenie i monitorowanie.

2. Interesariusze i użytkownicy

- **Interesariusze:**

Klasyfikacja interesariuszy

I. Interesariusze wewnętrzni

- **Programiści ML:** Odpowiedzialni za rozwój, implementację oraz integrację modeli AI w systemie
- **Menadżerowie projektów MLowych:** Zarządzają harmonogramem, zasobami i monitorują postępy projektu
- **DevOps Engineer:** Zapewniają stabilności, bezpieczeństwa i skalowalności infrastruktury

II. Interesariusze zewnętrzni

- **Anotatorzy:** Użytkownicy systemu, odpowiedzialni za ręczne oznaczanie danych przy wspomaganiu automatycznych sugestii AI
- **Właściciele danych:** Podmioty udostępniające dane, ustalające zasady bezpieczeństwa i kontrolę nad informacjami
- **Państwo:** Zapewnienie bezpieczeństwa danych przez przechowywanie i udostępnianie ich zgodnie z obowiązującym prawem

III. Interesariusze nadzorujący projekt

- **Opiekun projektu:** Osoba sprawująca pieczę nad projektem, wymagająca przygotowania dokumentacji projektu w postaci pracy inżynierskiej
- **Użytkownicy końcowi:**
 - Programiści ML
 - Anotatorzy
 - Menadżerowie projektów MLowych.

3. Zespół

Członkowie zespołu i ich role:

- Jerzy Szyjut (mail) - Programista, kierownik projektu:
 - Umiejętności: Backend, devops i machine learning
 - Obszary odpowiedzialności: Utrzymywanie infrastruktury projektu, rozwój platformy, inżyniera wymagań
- Hubert Malinowski (mail) - Programista:
 - Umiejętności: Backend, machine learning i software testing
 - Obszary odpowiedzialności: Przegląd i wybór rozwiązań projektowych, rozwój i testowanie platformy
- dr inż. Tomasz Dziubich (mail) - Opiekun projektu:
 - Umiejętności: Machine learning, zarządzanie projektami B+R
 - Obszary odpowiedzialności: Nadzór pracy projektu, ekspertyza przy podejmowaniu decyzji strategicznych dla projektu

Forma pracy: Praca w stacjonarnie

4. Komunikacja w zespole i z interesariuszami

- **Spotkania wewnętrzne:** co dwa tygodnie, online, z członkami i opiekunem projektu
- **Spotkania z interesariuszami:** spotkania umawiane w razie potrzeby na platformie discord z członkami koła naukowego Gradient lub naukowcami zainteresowanymi naszym projektem
- **Środki komunikacja:** discord, stacjonarnie
- **Raportowanie postępów:** Github Projects

5. Współdzielenie dokumentów i kodu

- **Repozytorium kodu:** [GitHub](#)
- **Dokumentacja:** Github/Overleaf/Notion
- **Osoba odpowiedzialna za konfigurację i repozytorium:** Jerzy Szyjut
- **Osoba odpowiedzialna za porządek w dokumentacji:** Hubert Malinowski
- **Wersjonowanie:** Git

6. Narzędzia

- **Komunikacja:** Discord.
- **Zarządzanie projektem:** Github Projects
- **Repozytorium kodu:** GitHub
- **Tworzenie dokumentacji:** Github/Overleaf/Notion
- **Modelowanie:** draw.io.
- **Wytwarzanie i testowanie:** Python, Docker, PyTorch