**1. Tytuł Projektu**

**"Skrypt do predykcji danych demograficznych oparty o dane historyczne GUS"**

**2. Opis Projektu**

* **Streszczenie**: Projekt zakłada przygotowanie prostego modelu przewidywania populacji Polski lub regionu w oparciu o dane historyczne i niezaawansowane metody statystyczne (przykładowo autoregresja). Korzyścią wyjściowych danych jest otrzymanie danych demograficznych przyszłych wielkościach zmiennych losowych w określonym przyszłym momencie.
* **Tło i uzasadnienie**: Projekt jest potrzebny ze względu na rosnące zainteresowanie analizą demograficzną oraz potrzebę prognozowania populacji w różnych kontekstach (np. planowanie rozwoju miast, polityka społeczna).

**3. Cele Projektu**

* **Główny cel**: Stworzenie modelu przewidującego populację Polski lub wybranego regionu.
* **Cele szczegółowe**:
  + Zebranie i przetworzenie danych demograficznych z GUS.
  + Wybór odpowiednich metod statystycznych do modelowania populacji.
  + Implementacja modelu w języku Python z wykorzystaniem biblioteki Pandas.
  + Wizualizacja danych za pomocą biblioteki Seaborn i Matplotlib.

**4. Analiza Wymagań**

* **Wymagania funkcjonalne**:
  + System powinien umożliwiać wczytywanie danych demograficznych.
  + Model powinien przewidywać populację na podstawie dostępnych danych.
* **Wymagania niefunkcjonalne**:
  + Wydajność: Model powinien działać w akceptowalnym czasie.
  + Bezpieczeństwo: Dane demograficzne powinny być przechowywane i przetwarzane w sposób bezpieczny.
  + Interfejs użytkownika: Prezentacja wyników w czytelny sposób.

**5. Zakres Funkcjonalności**

* **Lista funkcjonalności**:
  + Wczytywanie danych demograficznych z plików źródłowych.
  + Przetwarzanie danych (czyszczenie, transformacja).
  + Wybór i implementacja modelu statystycznego.
  + Prezentacja wyników za pomocą Jupyter Notebook, tabel i wykresów

**6. Architektura i Technologie**

* **Architektura systemu**:
  + Interaktywny skrypt w Jupyter Notebook
* **Technologie**:
  + Python jako główny język programowania.
  + Biblioteka Pandas do przetwarzania danych.
  + Biblioteka Seaborn i Matplotlib do wizualizacji danych.
  + Biblioteka XGBoost do tworzenia modelu XGRegresor.
  + Biblioteka Sklearn do tworzenia i obsługi modelu drzew decyzyjnych.
  + Notatniki Jupyter do prezentacji wyników.

**7. Podział na Zadania i Harmonogram**

* **Lista zadań**:
  + Wczytanie danych z GUS.
  + Przetworzenie danych (czyszczenie, transformacja).
  + Wybór i implementacja modelu.
  + Wizualizacja wyników w Jupyter Notebook.
* **Harmonogram**:
  + Etap 1: Wczytanie danych (tydzień 1)
  + Etap 2: Przetwarzanie danych (tydzień 2-3).
  + Etap 3: Implementacja modelu (tydzień 4-5).
  + Etap 4: Wizualizacja danych – stworzenie tabel i wykresów (tydzień 6-7)
  + Etap 5: Redukcja danych
  + Etap 6: Stworzenie korelogramów
  + Etap 7: Rozpoczęcie pracy nad ewaluacją
  + Etap 8: Stworzenie wykresu ewaluacji
  + Etap 9: Dodanie interaktywności do skryptu
  + Etap 10: Aktualizacja modelu i poprawa ewaluacji
  + Etap 11: Stworzenie modelu XGBoost
  + Etap 12: Poprawa interaktywności skryptu
  + Etap 13: Dodanie możliwości wyboru wyglądu wykresów
  + Etap 14: Dodanie opisów i komentarzy
  + Etap 15: Dodanie dokumentacji
  + Etap 16: Poprawienie README

Obraz zawierający tekst, zrzut ekranu, Czcionka, numer

Opis wygenerowany automatycznie

**8. Przydział Zadań**

* **Tabela przydziału zadań**:

|  |  |  |
| --- | --- | --- |
| Imię i nazwisko | Rola | Zakres czynności |
| Klaudia Kopeć | Dokumentalista, Analityk danych | Wczytanie danych z GUS i przetworzenie danych |
| Dominik Moskal | Grafik, Analityk danych | Wizualizacja danych (stworzenie wykresów i tabel z wynikami) |
| Michał Żychowski | Programista, Analityk danych | Implementacja modelu |

**9. Ryzyka i Strategie Zarządzania Ryzykiem**

* **Identyfikacja ryzyk**:
  + Brak dostępu do aktualnych danych demograficznych.
  + Trudności w implementacji modelu.
  + Ograniczenia czasowe.
* **Strategie minimalizacji ryzyka**:
  + Regularne aktualizacje danych z GUS.
  + Wybór prostego modelu, który nie wymaga zaawansowanych technik.
  + Realistyczne oszacowanie czasu potrzebnego na realizację projektu.

**10. Dokumentacja i Raportowanie**

* **Zasady dokumentowania**:
  + Postępy w projekcie będą dokumentowane w Jupyter Notebook.
  + Wykorzystanie systemu kontroli wersji (Git + Github)