Jakub Misiło

Kamil Nędza

Michał Oliwa

**Heart Disease Analysis – realizacja projektu**

Realizacja projektu została rozpoczęta 4 czerwca 2022 r. Pierwszym zadaniem naszego zespołu było rozdzielenie zadań – tak aby każdy członek miał określony cel do którego dążył. Podział był następujący:

- Jakub Misiło – realizacja zadania predykcji na podstawie zgromadzonych danych, stworzenie i zarządzanie zdalnym repozytorium w serwisie GitHub, śledzenie progresu oraz stworzenie dziennika projektu.

- Kamil Nędza – stworzenie odpowiednich wykresów, wizualizacja danych. Kompleksowa analiza, w szczególności korelacja pomiędzy danymi parametrami, która pozwoliła na wybranie odpowiednich do przewidywań kolumn, stworzenie plakatu.

- Michał Oliwa – odpowiednie przetworzenie danych, przygotowanie kodu odpowiedzialnego za to. Przygotowanie danych w sposób uniwersalny – odpowiedni do analizy, oraz umożliwiający predykcje, stworzenie plakatu.

Kolejnym krokiem było rozpoczęcie prac nad obróbką danych. Braki danych zostały usunięte – nasz zbiór jest na tyle duży, że przypadki w których występowały braki mogły zostać wyrzucone. Dane które składały się z łańcuchów znaków musiały zostać zastąpione odpowiadającymi każdej klasie (możliwej do uzyskania wartości z kolumny) liczbie. Umożliwiło to sprawną analizę, oraz przygotowanie prostego modelu, potrafiącego przewidzieć szansę na wystąpienie u danej osoby choroby serca.

Następnie do pracy mogły przystąpić kolejne dwie osoby. Została wykonana szczegółowa analiza – za pomocą macierzy korelacji dowiedzieliśmy się jakie czynniki są najbardziej powiązane z występowaniem chorób serca. Do tego zostało wykonane wiele dodatkowych wykresów – abyśmy mogli jeszcze lepiej poznać złożoność danych oraz to jakie czynniki powinniśmy wykorzystać do predykcji. Oprócz macierzy korelacji wykorzystane zostały jeszcze: histogram, wykres słupkowy, czy też zaawansowany wykres kołowy!

Po tym kroku mógł zostać stworzony model, który na podstawie wprowadzonych danych od pacjenta będzie w stanie przewidzieć ryzyko wystąpienia u niego choroby serca. Zadanie wymagało zbalansowania naszych danych – przypadki w których nie wystąpiła żadna choroba serca przeważały nad przypadkami przeciwnymi – 286430:26584. Jak widać różnica jest ogromna. Losowo wybrane 26584 przypadki bez wystąpienia choroby serca zostały wybrane i sklejone w jedną ramkę danych razem z wszystkimi przypadkami z wystąpieniem problemów z sercem. Pozwoliło to wykonać miarodajne dostosowanie modelu do naszych danych. Zostało wykonane kilka przykładowych przewidywań, w celu sprawdzenie poprawności przewidywania.

Nasza praca nad projektem przebiegła bez komplikacji. Wszystkie założone przez nas punkty zostały wykonane.