Употреба на машинно обучение за засичане на депресия

Проектът е изготвен от мен, Стоян Куцаров. Използва се dataset от Kaggle, който показва различни социално-икономически фактори и тяхната връзка с депресията. След това потребителят бива питан въпроси, които моделът използва за да предположи дали един човек е депресиран или не.

Идеята за проекта ми дойде след миналия петък. Видях, че темата за засичане на Алцхаймер и деменция беше използвана, затова реших да направя нещо различно. Първото друго психично заболяване, за което ми хрумна беше депресията и следователно реших да направя приложение, което спомага за засичането и. Макар и то никога да не види бял ден смятам че се справя адекватно със заданието си.

За интерфейса бе използван tkinter, обема на кода се състои в това, че просто има много критерии и трябва да се обходи всяка една от тях. Колкото до самата логика на приложението е обичайното за тренирането на един модел. Numpy преобразува входни данни в използваем формат, pandas обработва dataset-а, seaborn се използва за проверка на приложението, спомага за създаването на confusion матрици, които правят грешките нагледни. sklearn се ползва за моделът, който приложението използва.

Първо от датасета се премахват ненужните колони, това са тези, които съм сметнал, че не оказват влияние върху крайния резултат или просто не важат за повечето от нас. След това се премахват празните стойности, като се запълват със средното за колоната. От този нов датасет създаваме друг, които се преработва и се използва за обучението на модела. Минава се през няколко итерации на обучението, докато се стигне до стойности, които се смятат за приемливи. В края на главния файл е функция, която предвижда стойността, която търсим. Отиваме в интерфейса, където се създава прозорец с фон, label, combobox и бутон. Потребителят отговаря на всички въпроси, след което отговорите му се вкарват в масив, обработват се и по тях моделът прави предположение за присъствието или отсъствието на депресия в този човек.

Макар че в момента проектът е функционален, както винаги има неща, които искат допипване. Винаги може да се стремим към по-висока точност по предположенията, както и да се добави допълнителна функционалност. Може да се обсъди идеята за по-голям датасет с повече въпроси и предлагащ по-точни данни.