O zbiorze danych

Coroczne badanie CDC Z 2020r. i stanowi główną część systemu nadzoru nad czynnikami ryzyka behawioralnego (BRFSS), który przeprowadza coroczne ankiety telefoniczne w celu gromadzenia danych na temat stanu zdrowia mieszkańców USA.

400 tys. Dorosłych

401 958 wierszy

Jak opisano powyżej, pierwotny zbiór danych składający się z prawie 300 zmiennych został zredukowany do zaledwie około 20 zmiennych. Oprócz klasycznej EDA, ten zbiór danych można wykorzystać do zastosowania szeregu metod uczenia maszynowego, w szczególności modeli klasyfikatorów (regresja logistyczna, SVM, losowy las itp.). Zmienną „Choroby serca” należy traktować binarnie („Tak” – respondent miał chorobę serca; „Nie” – respondent nie miał choroby serca). Należy jednak pamiętać, że klasy nie są zrównoważone, dlatego nie zaleca się klasycznego podejścia do stosowania modelu. Ustalenie wag/podpróbkowanie powinno dać znacznie lepsze wyniki. Na podstawie zbioru danych skonstruowałem model regresji logistycznej i umieściłem go w aplikacji, która może Cię zainspirować: <https://share.streamlit.io/kamilpytlak/heart-condition-checker/main/app.py>. Czy możesz wskazać, które zmienne mają istotny wpływ na prawdopodobieństwo wystąpienia chorób serca?

Według [CDC](https://www.cdc.gov/heartdisease/risk_factors.htm) choroby serca są jedną z głównych przyczyn śmierci ludzi większości ras w USA (Afroamerykanie, Indianie amerykańscy i rdzenni mieszkańcy Alaski oraz ludzie biali). Około połowa Amerykanów (47%) ma co najmniej 1 z 3 kluczowych czynników ryzyka chorób serca: wysokie ciśnienie krwi, wysoki poziom cholesterolu i palenie. Innymi kluczowymi wskaźnikami są cukrzyca, otyłość (wysokie BMI), brak wystarczającej aktywności fizycznej lub spożywanie zbyt dużej ilości alkoholu. Wykrywanie i zapobieganie czynnikom, które mają największy wpływ na choroby serca, jest bardzo ważne w opiece zdrowotnej. Z kolei postęp obliczeniowy umożliwia zastosowanie metod uczenia maszynowego do wykrywania „wzorców” na podstawie danych, które mogą przewidzieć stan pacjenta