Bu iş, lojistik regresyon uygulamanıza izin verecektir. İkili bir değişkeni (değerleri DOĞRU / YANLIŞ, 0/1 veya EVET / HAYIR) niceliksel bir açıklayıcı değişkenin bir fonksiyonu olarak analiz etmek için lojistik bir regresyon kullanabiliriz. Lojistik regresyon, sürekli değişkenleri tahmin etmekten değil, ikili değişkenleri tahmin etmekten oluşur.

Koruyucu hekimlik alanında çalışıyoruz. Bu nedenle işiniz, yaşam hijyeni (temizlik, aynı zamanda diyet, bir sporu veya fiziksel aktiviteyi teşvik etme, ergonomi ve çaba gösterme şekli, riskli davranışların önlenmesi vb.) Tavsiyelerde bulunmak ve hastalıkların tespitinde ve daha spesifik olarak kardiyovasküler risklerin önlenmesinde destek.

Fransa'da her yıl 300.000 ila 400.000 kardiyovasküler kaza meydana geliyor ve bunların üçte biri ölümcül. Kardiyovasküler risk nasıl daha iyi tahmin edilir? Birkaç risk faktörü tanımlanırsa, bu faktörler arasındaki etkileşimler nelerdir? Başta serebrovasküler kazalar (CVA) ve miyokardiyal enfarktüsler olmak üzere kardiyovasküler hastalıklar, Fransa'da ikinci önde gelen ölüm nedenidir. Kardiyovasküler risk faktörlerinin listesi maalesef uzun:

Dolayısıyla 12 risk faktörü, gerçek bir kardiyovasküler risk faktörleri ağı oluşturur.

Bu etkileşimlere dayanarak, araştırmacılar 4 risk faktörü grubunu vurgulayabildiler:

1. "Değiştirilemez faktörler" (cinsiyet, yaş ve aile öyküsü): Bunlar diğer faktörleri öngörür, ancak diğer faktörler tarafından tahmin edilemez.

2. "Yaşam tarzı faktörleri" (sigara içme, fiziksel hareketsizlik, alkolizm): birçok başka faktörü öngörürler (değiştirilemeyen faktörler hariç), ancak diğer faktörler tarafından pek tahmin edilmezler.

3. "Upstream klinik faktörler" (uyku bozuklukları, obezite, depresyon): Bunlar diğer birçok faktörü öngörür ve kendileri birçok faktör tarafından tahmin edilir.

4. "Aşağı akım klinik faktörler" (arteriyel hipertansiyon, dislipidemi, diyabet): çok az faktörü öngörürler, ancak bunun aksine birçok faktör tarafından tahmin edilirler.

Bu kardiyovasküler riskleri önleme yaklaşımınızda sizi en iyi şekilde desteklemek için, kardiyovasküler risklerin hızlı teşhisine olanak tanıyan bir araç geliştirmeye karar verdiniz. Bu araç, kardiyovasküler bir risk olup olmadığını tahmin etmek için bir makine öğrenimi algoritması (ikili sınıflandırma: ikili tahmin: 0 veya 1) uygulayacaktır.

Bu verilere dayanarak, risk altındaki öznelerin kimler olduğunu tahmin etmek için bir lojistik regresyon modeli oluşturun!

1. Lojistik regresyonu izleyin.

2. Aşağıdaki çalışma için bir jüpiter not defteri kullanın.

3. Matplotlib ve Seaborn kitaplıklarıyla verileri görselleştirin ve analiz edin.

4. Yukarıda sunulan örnek olayı Scikit-Learn kitaplığı ile çözün.

5. Yukarıda sunulan örnek olay incelemesini kendi python sınıfınızla çözün (Scikit-Learn kitaplığını kullanmadan).

6. Arthur 53 yaşında, sigara içen, atlet, 175 cm, 85 kg, normalin üzerinde kolesterol seviyesi ve normal glikoz seviyesi, ortalama sistolik kan basıncı ve buna karşılık gelen diyastolik kan basıncı olup olmadığını tahmin edin. veri setinin 3. çeyreğinin ortalaması (% 50-% 75) kardiyovasküler risk altında olan bir konudur.

7. Dizüstü bilgisayarınızı Github aracılığıyla erişilebilir hale getirin.

8. Github bağlantınızı paylaşın.