

Medikal Maliyet Analizi

Veri Seti: <https://www.kaggle.com/datasets/mirichoi0218/insurance>

Bu projede, yukarıda verilen veri setini kullanarak uçtan uca bir veri bilimi uygulaması geliştirmeye çalışılacaktır. Projenin amacı, verilen değişkenlere göre bir kişinin sağlık sigortasının yaklaşık ne kadar masraflı olacağını tahmin etmektir. Projeyi oluştururken aşağıdaki talimatlara uymaya çalışınız ve projenin kendinize özgü olmasına dikkat ediniz.

1. Google Colaboratory Dosyası Açılması

- Projenizin .ipynb uzantılı olmasına özen gösteriniz.
- Projenizdeki detayları açıklayan yorum satırları olmasına dikkat ediniz.
- Proje teslim edilirken bu .ipynb uzantılı dosyasının hücreleri çalıştırılmış ve sonuçları görünecek şekilde teslim ediniz.

2. Gerekli Kütüphanelerin Eklenmesi

- Projede kullanılacak olan kütüphanelerin Colab ortamına ekleyiniz.
- Pandas, numpy, Seaborn, Matplotlib ve Sklearn kütüphanelerini ekleyiniz.

3. Keşifsel Veri Analizi Yapılması

- Veriyi inceleyerek, analiz ederek veriden anlamlı sonuçlar çıkarınız.
 - Bmi(Vücut Kitle İndeksi)'nin dağılımını inceleyiniz
 - “smoker” ile “charges” arasındaki ilişkiyi inceleyiniz
 - “smoker” (Sigara tüketen) ile “region”(Bölge) arasındaki ilişkiyi inceleyiniz.
 - “bmi” ile “sex”(Cinsiyet) arasındaki ilişkiyi inceleyiniz.
 - En çok “children” a sahip “region” ı bulunuz.
 - “Age” ile “bmi” arasındaki ilişkiyi inceleyiniz.
 - “bmi” ile “children” arasındaki ilişkiyi inceleyiniz.
 - “bmi” değişkeninde outlier var mıdır? İnceleyiniz.
 - “bmi” ile “charges” arasındaki ilişkiyi inceleyiniz.
 - “region”, “smoker” ve “bmi” arasındaki ilişkiyi bar plot kullanarak inceleyiniz.

- Veriyi incelerken veri görselleştirme tekniklerini olabildiğince kullanmaya çalışınız.
- Analizlerden çıkardığınız anlamları yorum satırı olarak ekleyiniz.

4. Veri Ön İşleme Yapılması

- Bu kısımda elinizde olan veriyi model eğitmek için hazır hale getiriniz.
- Kategorik değişkenleri düzenlemek için Label ve One-Hot Encoding tekniklerini kullanınız.
- Veri setinizi X_train, X_test, y_train, y_test olacak şekilde bölüştürünüz.
- Veri setini normalize ederek ölçekleyiniz.

5. Model Seçme

- Birkaç regresyon modeli seçiniz bunları ön işleme yapılan veri ile eğitiniz
- Seçilen modellerin performanslarını çapraz doğrulama kullanarak inceleyiniz.
- En iyi performans gösteren modeli seçiniz

6. Hiper-parametre Optimizasyonu

- Bir önceki adımda seçilen modelin hiper-parametrelerinin optimize ediniz.
- Grid Search ile parametreleri optimize ediniz.

7. Modeli Değerlendirme

- Regresyon modeli değerlendirme metriklerini kullanarak optimize edilmiş olan modelin değerlendirmesini yapınız. (Ör. Mean Squared Error, Mean Absolute Error vb.)

8. Projenin Teslim Edilmesi

- Proje için, bir kod dosyası ve proje dökümanı hazırlamanız gerekmektedir.
- Hazırladığınız bu dosyaları bir GitHub reposuna eklemenizi ve bu reponun linkini aşağıda linkte yer alan forma eklemeniz gerekmektedir.
- Proje takım halinde yapılacaktır. Oluşturulan takımlar 5-7 kişi arasında olmalıdır.
- Proje takımınız ile alakalı bilgileri bu form üzerinden iletebilirsiniz.
- Form Linki: <https://forms.gle/f9M1G62XxYt7JrYAA>
- Son Teslim tarihi: 29 Ağustos