

01 Smart User Data Cleaner

1. Giriş

- Projenin amacı
 - Projenin amacı verilmiş ham ve karışık veriyi çeşitli Python string fonksiyonları kullanarak temizlemek ve ihtiyaca göre düzenlemektir.
- Veri temizlemenin neden önemli olduğu
 - Gerçek hayatta da gelen ham veri temizlenerek anlamlı parçalara ayrılır ve kullanıcı kayıt sistemi gibi sistemlerde temiz haliyle ihtiyaca göre kullanılır.
- String işlemlerinin veri ön işleme sürecindeki rolü
 - Kullanıcıdan aldığımız veri içinde ihtiyaca göre sayılar bulunsa bile string olarak gelir. Ancak bu data string işlemleri ile ihtiyaca göre ayrıldıktan sonra kullanılabilir hale getirilir.

2. Kullanılan Python Konuları

- Değişkenler (Variables) – Belirlediğimiz değerleri içinde tutar.
- String veri tipi – Bu değerlerden metin şeklinde olan verilerdir.
- String slicing – String verinin belli bir kısmını almak için kullandığımız string işlemi.
- String metodları (strip, lower, title, split, find) – strip sağdan soldan boşluk silmeye, lower tüm harfleri küçük hale getirmeye, title bütün harfleri büyük hale getirmeye, split stringi istenen yerden bölmeye yarar.
- int ve float dönüşümleri – int ve float gelen string verinin gerektiği yerlerde sayıya çevrilip kullanılması için vardır.

3. Veri Temizleme Süreci

- Ham verideki problemler
 - Gelen ham veride büyük küçük harf uyumsuzlukları, gereksiz boşluklar, int veya float olarak kullanılması gereken veriler vardı.
- Hangi adımda hangi problemin çözüldüğü
 - İlk başta strip fonksiyonu ile gereksiz boşluklardan kurtarıldı ve strip fonksiyonuna “;” parametresi verilerek gelen ham datanın bu parametreye göre parçalanması sağlandı.
 - Parçalanmış verinin sıfıncı indeksi olan ad soyad ilk olarak sağ ve soldaki gereksiz boşluklardan strip ile temizlendi, lower ile tüm harfler küçültülerek düzenli bir görünüm elde edildi, title ile kelimelerin ilk harfleri büyük hale getirildi.
 - Parçalanmış verinin birinci indeksi olan yaş, float işlemiyle kullanılabilecek hale dönüştürüldü ve 10 eklenerek 10 yıl sonraki yaş hesaplandı.
 - Parçalanmış verinin ikinci indeksi olan boy, yine float işlemiyle kullanılabilecek hale dönüştürülüp cm’ye dönüştürüldü.

- Yanlış girilmiş verilerin sisteme etkisi
- - Yanlış girilmiş veriler sisteme, uygulamay giriş yapılmasını engelleyebilir, örneğin @ içermeyen bir mail adresi.

4. Sonuç ve Değerlendirme

- Projenin çözdüğü problemler – Veri temizliği, verilerin okunulabilir hale getirilmesi
- Gerçek hayatta kullanım alanları – Daha geniş ve kapsamlı verilerin kullanılabilir hale getirilmek üzere birtakım string işlemlerine tabi tutulması
- Kazanılan Python becerileri – String fonksiyonları

Python and Data Science Bootcamp 2026 – Nisa Nur Tezcan