

ÖZET

Bu bitirme projesi gerçek hayattan bir problemi temel alarak oluşturuldu. Büyük bir seyahat acentesinin sitesinde, sitede otel arayan birine otel sıralaması sadece SEO ve provider özellikleri üzerinden oteller sıralanıyordu ve sitede daha iyi bir sıralama yaptırmak için veri tabanında ki veriyi işe yarar biçimde analiz etme ihtiyacı ortaya çıktı. Yani en fazla sayıda iyi yorum alan otelin üst sıralarda çıkması, kişinin site içinde sürecinin hızlanması anlamına geliyordu.

Veri bilimi, karmaşık problemleri çözer yani elimizde ki veriyi değerli bilgiye çevirir. Son zamanlar da veri biliminde kullanılan programlama dili çoğu zaman Python oluyor. Python dili hazır birçok veri bilimi kütüphanesi bulunduruyor.

Projemize başlarken öncelikle kaggle sitesinden bulduğumuz bir seyahat acentesinin veri setini, python dilinin hazır kütüphanesi olan pandas ile kullanılır hale getirdik. Elimizdeki tüm farklı otellerin bir listesini çıkarttık.

Ardından yorumların içinden bulmamız gereken kelimeleri belirlememiz gerekiyordu, bu kelimelerin en doğru şekilde belirlenmesi için tüm yorumları “veri.txt” dosyasına atıp bu dosyanın içinde kullanılan tüm kelimelerin sayıları ile birlikte collection yapısıyla bastırdık. Seçtiğimiz kelimeleri sayılarına göre belirledik, örneğin “good” kelimesini seçtik çünkü yorum sütununda en çok kullanılan yedinci kelime olduğunu farkettilik ve bu şekilde tüm pozitif ve negatif kelimeleri örnek teşkil edecek şekilde belirlemiş olduk.

Veri setimizde bulunan “Review_Text” sütununu yani yorumların bulunduğu sütunu ayrı olarak dictionary tipinde bir listeye atama yaptık. RegEx modülünü kullanarak belirlediğimiz örnek kelimeleri tüm veri setinde arattık. Burada aldığımız sonuçlar hangi otele ait yorumda kaç tane pozitif ve negatif kelime geçtiğini tespit ettik ve bunları dictionary yapısına atadık.

Son adımımız: pozitif ve negatif yorumların bulunduğu dictionary yapılarının key ve value değerlerini kullanarak pozitif kelime sayısından negatif kelime sayılarını çıkartıp en fazla sayıda iyi yorum alan otel sıralamasını gerçekleştirmiş olduk.