



Kural Tabanlı Sınıflandırma ile Potansiyel Müşteri Getirisi Hesaplama

Hazırlayanlar

ELİF BEGÜM ÇETİNKAYA - 02205076054

MEHMET EMİN ŞAHİN - 02190201039

ÇAĞRI MİRAC ÇELİK - 02200201014

OĞUZHAN SOLMAZ - 02190201064

EREN UZUN - 02205076052

Proje Linki

<https://github.com/EMIN200097/VeriMadenciliği>

İş Problemi

Bir oyun şirketi müşterilerinin bazı özelliklerini kullanarak seviye tabanlı yeni müşteri tanımları oluşturmak ve bu yeni müşteri tanımlarına göre segmentler oluşturup bu segmentlere göre yeni gelebilecek müşterilerin şirkete ortalama ne kadar kazandırabileceğini tahmin etmek istemektedir.

Örneğin:

Türkiye'den IOS kullanıcısı olan 25 yaşındaki bir erkek kullanıcının ortalama ne kadar kazandırabileceği belirlenmek isteniyor.



Veri Seti Hikayesi

Persona.csv veri seti uluslararası bir oyun şirketinin sattığı ürünlerin fiyatlarını ve bu ürünleri satın alan kullanıcıların bazı demografik bilgilerini barındırmaktadır. Veri seti her satış işleminde oluşan kayıtlardan meydana gelmektedir. Bunun anlamı tablo tekilleştirilmemiştir. Diğer bir ifade ile belirli özelliklere sahip bir kullanıcı birden fazla alışveriş yapmış olabilir.

PRICE	SOURCE	SEX	COUNTRY	AGE
39	android	male	bra	17
39	android	male	bra	17
49	android	male	bra	17
29	android	male	tur	17
49	android	male	tur	17





Değişkenler

persona.csv

PRICE – Müşterinin harcama tutarı

SOURCE – Müşterinin bağlandığı cihaz türü

SEX – Müşterinin cinsiyeti

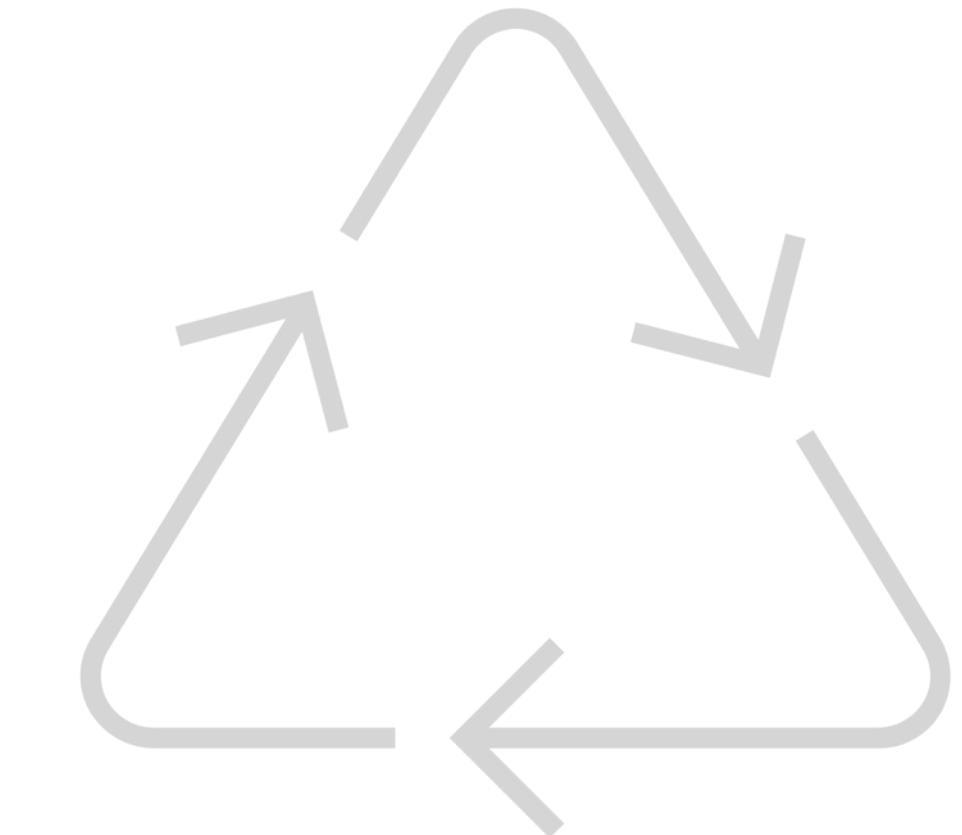
COUNTRY – Müşterinin ülkesi

AGE – Müşterinin yaşı

Kural Tabanlı Sınıflandırma

Uygulama Öncesi Veri Seti :

PRICE	SOURCE	SEX	COUNTRY	AGE
39	android	male	bra	17
39	android	male	bra	17
49	android	male	bra	17
29	android	male	tur	17
49	android	male	tur	17



Hedeflenen çıktı :

customers_level_based	PRICE	SEGMENT
BRA_ANDROID_FEMALE_0_18	35.6453	B
BRA_ANDROID_FEMALE_19_23	34.0773	C
BRA_ANDROID_FEMALE_24_30	33.8639	C
BRA_ANDROID_FEMALE_31_40	34.8983	B
BRA_ANDROID_FEMALE_41_66	36.7371	A



Proje Adımları

Adım 1 : ‘persona.csv’ dosyasını okutunuz ve veri seti ile ilgili genel bilgileri gösteriniz.

Soru 1 : Kaç unique SOURCE vardır?

Soru 2 : Kaç unique PRICE vardır?

Soru 3 : Hangi PRICE'dan kaçar tane satış gerçekleşmiş?

Soru 4 : Hangi ülkeden kaçar tane satış olmuş?

Soru 5 : Ülkelere göre satışlardan toplam ne kadar kazanılmış?

Soru 6 : SOURCE türlerine göre satış sayıları nedir?

Soru 7 : Ülkelere göre PRICE ortalamaları nedir?

Soru 8 : SOURCE'lara göre PRICE ortalamaları nedir?

Soru 9 : COUNTRY-SOURCE kırılımında PRICE ortalamaları nedir?

Adım 2: COUNTRY, SOURCE, SEX, AGE kırılımında ortalama kazanç hesaplanması

Elde edilen çıktı:

	COUNTRY	SOURCE	SEX	AGE	PRICE
0	bra	android	female	15	38.71
1				16	35.94
2				17	35.66
3				18	32.25
4				19	35.20

Adım 3 : Adım 2 elde edilen çıktıyı PRICE değerine göre sıralama

- Çıktıyı agg_df olarak kaydediyoruz.

Elde edilen çıktı:

	COUNTRY	SOURCE	SEX	AGE	PRICE
0	bra	android	male	46	59.0
1	usa	android	male	36	59.0
2	fra	android	female	24	59.0
3	usa	ios	male	32	54.0
4	deu	android	female	36	49.0

Adım 4: Indekste yer alan isimleri değişken ismine çevirme.

- Üçüncü adımın çıktısında yer alan PRICE dışındaki tüm değişkenler index isimleridir. Bu isimleri değişken isimlerine çevirdik.

Adım 5: Age değişkenini kategorik değişkene çevirme ve agg_df 'e ekleme

- Age sayısal değişkenini kategorik değişkene çevirme.
- Aralıkları oluşturma ve AGE_CAT olarak tanımlama
- Örneğin: '0_18', '19_23', '24_30', '31_40', '41_70'

Elde edilen çıktı:

COUNTRY	SOURCE	SEX	AGE	PRICE	AGE_CAT
bra	android	male	46	59.0	41_70
usa	android	male	36	59.0	31_40
fra	android	female	24	59.0	24_30
usa	ios	male	32	54.0	31_40
deu	android	female	36	49.0	31_40

Adım 6: Yeni seviye tabanlı müşterileri tanımlama

- Yeni seviye tabanlı müşterileri tanımlama ve veri setine değişken olarak ekleme
- Yeni eklenecek değişkenin adı: customers_level_based
- Önceki adımda elde ettiğimiz çıktıdaki gözlemleri bir araya getirerek customers_level_based değişkenini oluşturduk.

Bu tabloda bulunan gözlemler bir araya gelecek

COUNTRY	SOURCE	SEX	AGE	PRICE	AGE_CAT
bra	android	male	46	59.0	41_66
usa	android	male	36	59.0	31_40
fra	android	female	24	59.0	24_30
usa	android	male	32	54.0	31_40
deu	android	female	36	49.0	31_40

Elde edilmesi gereken çıktı

customers_level_based	PRICE
BRA_ANDROID_MALE_41_66	59.0
USA_ANDROID_MALE_31_40	59.0
FRA_ANDROID_FEMALE_24_30	59.0
USA_IOS_MALE_31_40	54.0
DEU_ANDROID_FEMALE_31_40	49.0

Adım 7: Yeni müşterileri (personaları) segmentlere ayırma.

- Yeni müşterileri (Örnek: USA_ANDROID_MALE_0_18) PRICE'a göre 4 segmente ayırdık (A, B, C, D).
- Segmentleri SEGMENT isimlendirmesi ile değişken olarak agg_df'e ekledik.
- Segmentleri betimleme (Segmentlere göre groupby yapıp price mean, max, sum'larını aldık).

Adım 8: Yeni gelen müşterileri sınıflandırıp, ne kadar gelir getirebileceklerini tahmin etme.

Soru 1 : 33 yaşında ANDROID kullanan bir Türk kadını hangi segmente aittir ve ortalama ne kadar gelir kazandırması beklenir?

Soru 2 : 35 yaşında IOS kullanan bir Fransız kadını hangi segmente aittir ve ortalama ne kadar gelir kazandırması beklenir?