

-T.C
KONYA TEKNİ ÜNİVERSİTESİ BİLGİSAYAR MÜHENDİSLİĞİ
BİLİŞİM TEKNOLOJİLERİ UYGULAMASI
(DÖNEM PROJESİ) ARA RAPOR FORMU

Öğrencinin Adı- Soyadı	Dilemre Ülkü
Numarası:	171213055
Danışmanı Adı Soyadı:	Mesut Güündüz
Projenin Konusu: Konya'daki Tarım Eğilimleri ve Verimliliği	

DÖNEM İÇİ YAPILAN ÇALIŞMALAR

Konya Büyükşehir Belediyesi'ne ait açık veri platformu olan acikveri.konya.bel.tr sitesinden, veri seti kategorileri altında bulunan tarım kategorisi altındaki 18 veri seti indirilmiştir. Veri setlerine ait genel bilgiler aşağıdaki tabloda gösterilmiştir.

Veri seti ismi	Açıklama	Veri tipi
İlçe/Mahalle Bazlı Doğrudan Satın Alınabilecek Ürünler	Doğrudan satın alınabilecek ürünler	text
İlçe/Mahalle Bazlı Sulak Alan Miktarı	Sulak alan miktarı	nümerik
İlçe/Mahalle Bazlı Meşhur Ürünler	Meşhur ürünler	text
İlçe/Mahalle Bazlı En Çok Yetiştirilen Ürünler	En çok yetiştirilen ürünler	text
İlçe/Mahalle Bazlı Traktör Sayısı	Toplam traktör sayısı	text
İlçe/Mahalle Bazlı Bireysel Arazi Ortalaması	Bireysel arazi ortalaması	text
İlçe/Mahalle Bazlı Alternatif Yeni Ürün Çeşitleri Üretimi	Alternatif yeni ürün çeşitleri üretimi	text
İlçe/Mahalle Bazlı Endemik Bitkiler	Endemik (sadece bu bölgede yetişen bitkiler)	text
İlçe/Mahalle Bazlı Talep Edilen Tarımsal Eğitimler	Talep edilen tarımsal eğitimler	text
İlçe/Mahalle Bazlı Tıbbi ve Aromatik Bitki Yetiştiriciliği	Tıbbi ve aromatik bitki yetiştiriciliği	text
İlçe/Mahalle Bazlı Arıcılık Durum Bilgisi	Mahallede arıcılığın yapılp yapılmadığı bilgisi	text
İlçe/Mahalle Bazlı Beslenen Hayvan Türleri	Beslenen hayvan türleri bilgisi	text
İlçe/Mahalle Bazlı Sulama Suyu Kaynağı Yönetimi	Sulama suyu kaynağı yönetimi	text
İlçe/Mahalle Bazlı Hayvan Beslemede Kullanılan Ürünler	Hayvan beslemede kullanılan ürünler	text
İlçe/Mahalle Bazlı Sulanabilir Alanda Yetiştirilen Ürünler	Sulanabilir alanda yetiştirilen ürünler	text
İlçe/Mahalle Bazlı Sulama Suyu Kaynağı	Tarımsal sulama suyu kaynağının nasıl olduğunun bilgisi	text
İlçe/Mahalle Bazlı İyi Tarım Uygulamaları	İyi tarım uygulamalarının bulunup bulunmadığı bilgisi	text
İlçe/Mahalle Bazlı Tarımsal Sulama Suyu Durum Bilgisi	Tarımsal sulama suyu olup olmadığı bilgisi	text

Tablo 1: Veri Sözlüğü

Veriler 2016 yılına aittir. Mevcut veri setlerinin her birinde İLÇE ADI ve MAHALLE ADI bilgisi text veri tipinde bulunmaktadır. Veriler csv formatında indirilmiştir. Glob ve OS kütüphaneleri ile dosya içerisinde bulunan tüm CSV dosyaları okunup Pandas kütüphanesi ile tüm veri setleri ayrı bir DataFrame veri yapısında, DataFrame'ler ise bir listesinde tutulmuştur. CSV dosyalarında Türkçe karakterleri okuyabilmek için ISO-8859-9 standardı kullanılmış, ayrıca olarak “;”, belirtilmiş, bazı CSV dosyalarında bir elamanda virgülle ayrılmış birden çok metin bulunduğu için satır sayısı verilmiş, satır sonlandırıcısı olarak “\n” belirtilmiştir. Ardından liste içinde tutulan DataFrame'ler Reduce fonksiyonu ve Pandas Merge fonksiyonu ile tüm veri setleri inner join parametresi ile birleştirilmiştir.

30 ilçeye ait 362 mahallenin verileri bulunmaktadır. 3 veri (traktör sayısı, bireysel arazi ortalaması, sulak alan miktarı) dışında kalan 15 veri kategorik verilerdir. Kategorik veri içeren 10 sütunun elemanlarında virgülle ayrılmış birden çok kategorik veri bulunmaktadır. Veri setinde tekrar eden satırla karşılaşılmamıştır.

Elde edilen veri setindeki bazı verilere ait örnekler aşağıdaki tabloda gösterilmiştir.

İLÇE ADI	MAHALLE ADI	ARICILIK\r	HAYVAN BESLEMEDE KULLANILAN ÜRÜNLER\r	TRAKTÖR SAYISI\r
Yunak	SEVİNÇ	HAYIR\r	ARPA, ÇAVDAR, SAMAN, YONCA, HAZIR YEM\r	75 Adet\r
Yunak	SIRAM	EVET\r	ARPA, ÇAVDAR, SAMAN, YONCA, HAZIR YEM\r	55 Adet\r
Yunak	YEŞİLOBA	HAYIR\r	YULAF, ARPA, SİLAJLIK MISIR, HAZIR YEM\r	35 Adet\r
Yunak	YEŞİLYAYLA	EVET\r	BUĞDAY, KÜSPE, ARPA, SAMAN, HAZIR YEM\r	20 Adet\r

Tablo 2: Ham veri

Veri ön işleme adımları şu şekildedir:

1. Veride bulunan ‘\r’ler kaldırılmıştır.
2. TRAKTÖR SAYISI sütunundaki ‘ Adet’ kısmı kaldırılmıştır.
3. BİREYSEL ARAZİ ORTALAMASI sütunundaki ‘ Dekar’ kısmı kaldırılmıştır.
4. SULAK ALAN MİKTARI sütununda bulunan ‘%’ kısmı kaldırılmıştır.
5. Boş veriler (‘ ’) numpy kütüphanesindeki nan verisi ile değiştirilmiştir.
6. Satır isimlerinde bulunan boşluklar ‘ _ ’ ile değiştirilmiştir.

Bu işlemler sonucu veri setindeki bazı verilere ait örnekler aşağıdaki tabloda gösterilmiştir.

İLÇE_A DI	MAHALLE_ ADI	ARICILIK	HAYVAN_BESLEMEDE_KULLANIL AN_ÜRÜNLER	TRAKTÖR _SAYISI
Yunak	SEVİNÇ	HAYIR	ARPA, ÇAVDAR, SAMAN, YONCA, HAZIR YEM	75
Yunak	SIRAM	EVET	ARPA, ÇAVDAR, SAMAN, YONCA, HAZIR YEM	55
Yunak	YEŞİLOBA	HAYIR	YULAF, ARPA, SİLAJLIK MISIR, HAZIR YEM	35
Yunak	YEŞİLYAYL A	EVET	BUĞDAY, KÜSPE, ARPA, SAMAN, HAZIR YEM	20

Tablo 3: Bazı ön işlemlerden sonra ham veri

Veriler incelendiği zaman toplam 4 satırda NaN veri ile karşılaşılmıştır. Bu satırlara ait bazı veriler aşağıdaki tabloda gösterilmiştir.

İLÇE_A DI	MAHALLE_ADI	ARICILIK	İYİ TARIM_UYGULA MALARI	TARIMSAL_SULAMA_SU YU
Selçuklu	BAŞARAKAVAK	Evet	NAN	NaN
Selçuklu	hasan	NaN	NaN	NaN
Selçuklu	Mehmet Fatih	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN

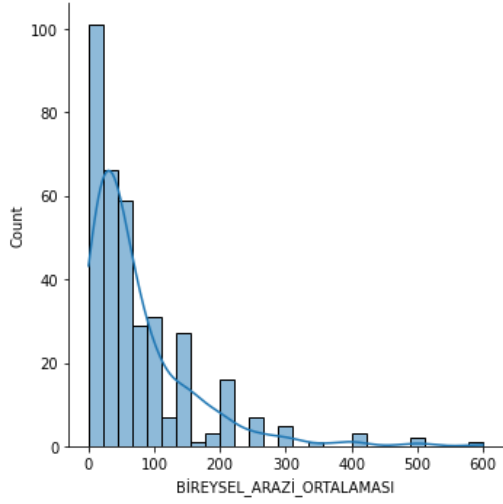
Tablo 4: NaN veri bulunan satırlar

Bu veriler veri setinin büyüklüğü göz önüne alınarak veri setinden atılmıştır. Bu işlem sonrası veri setindeki satır sayısı 369 olmuştur.

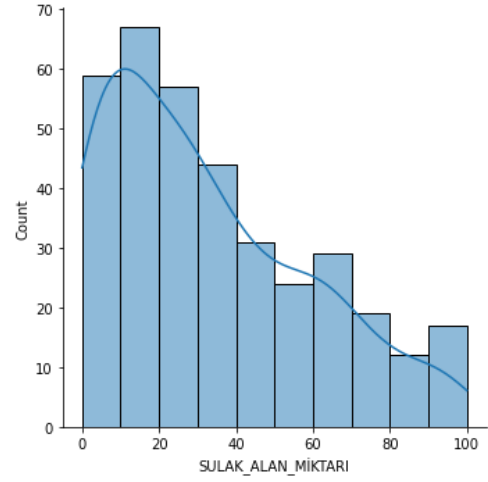
Her bir elemanı virgülle ayrılmış sütunlar encodig yapılmıştır. Encoding işlemi sırasında sütun isimlendirmeleri “eski sütun adı_kategorik veri ismi” şeklinde yapılmış olup, mevcut mahalle kategorik değişkeni barındırıyorsa 1, barındırmıyorsa 0 değerini alacak şekilde değer ataması yapılmıştır. ARICILIK, TARIMSAL SULAMA SUYU ve İYİ TARIM UYGULAMALARI satırlarındaki EVET, VAR değerleri 1, HAYIR, YOK değerleri 0 ile değiştirilmiştir.

Encoding işlemi sonrası veri setindeki satır sayısı 20 satır iken 282 satır olmuştur. İLÇE ADI ve MAHALLE ADI veri string, bu veriler dışındaki tüm veriler integer veri tipindedir.

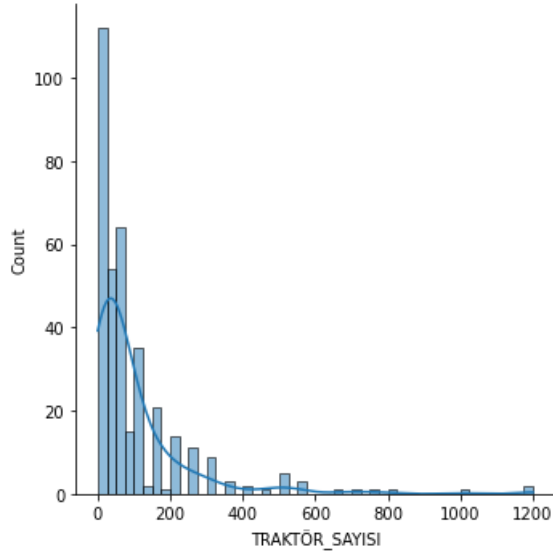
Yapılan veri görselleştirme çalışmaları sırasında veriler arasındaki ilişkiler incelenmiştir. Kullanılan kütüphaneler Seaborn ve Matplotlib’tir.



Şekil 1: Bireysel Arazi Ortalamasının Histogramı

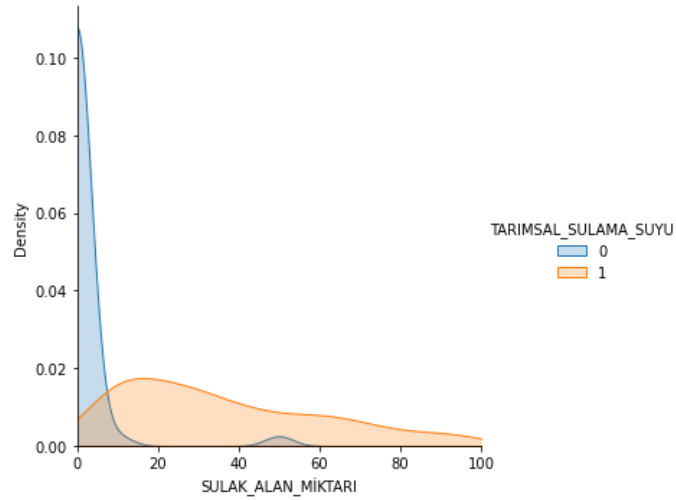


Şekil 2: Sulak Alan Miktarının Histogramı



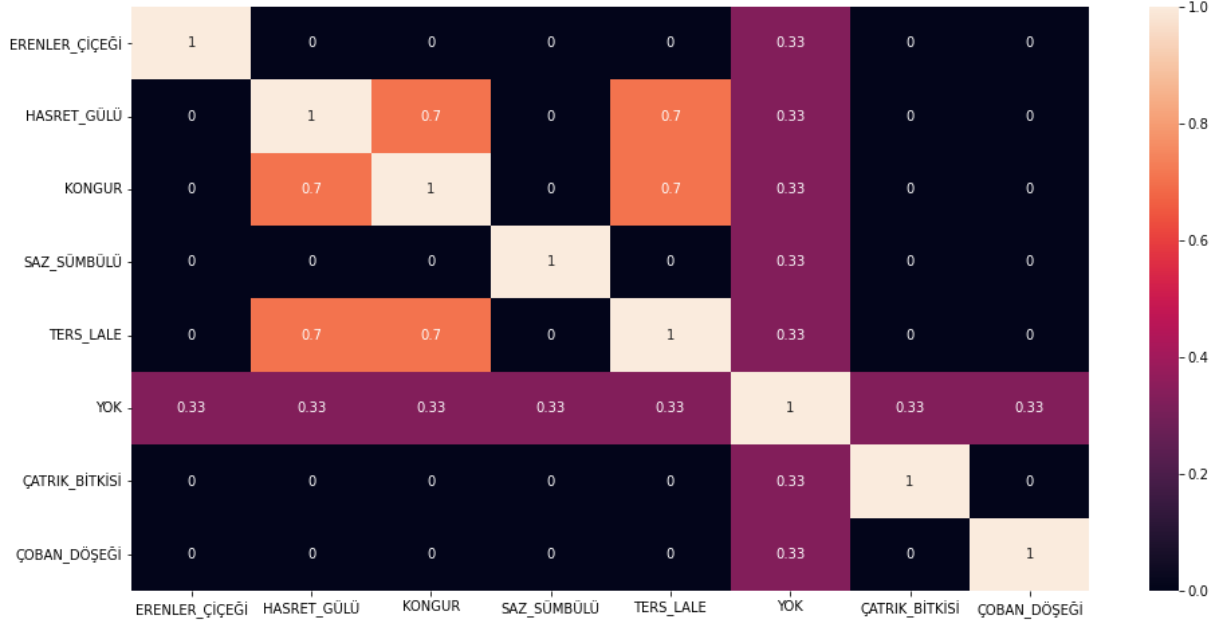
Şekil 3: Traktör Sayısı Histogramı

Yukarıdaki sayısal değişkenlerin histogram dağılımlarını incelediğimiz zaman pozitif çarpık normal dağılıma benzer bir dağılım gözlemlemekteyiz.



Şekil 4: Sulak alan miktarının tarımsal sulama suyunun olup olmamasına göre yoğunluk grafiği

Şekil 4'e göre Tarımsal sulama suyu bulunmayan mahallelerin neredeyse tamamı sulak alan miktarı %0 ile %20 olan mahallelerden oluşmaktadır.



Şekil 5: Belirli bir bölgede yetişen Endemik Bitkilerin birbirleri ile korelasyon ısı haritası

Korelasyon hesaplaması için phi korelasyon katsayısı ve PHIK kütüphanesi kullanılmıştır. Dörtlü (phi) korelasyon katsayısı, sınıflama ölçeğinde ölçülmüş iki kategorili süreksiz iki değişken arasındaki ilişkinin incelenmesinde kullanılır. Şekil 5'i incelediğimiz zaman ters lale, kongur ve hasret gülünün aynı ortamlarda yetiştiğini söyleyebiliriz.

Son olarak sayısal veriler içeren SULAK_ALAN_MİKTARI, TRAKTÖR_SAYISI, BİREYSEL_ARAZİ_ORTALAMASI sütunları min-max normalizasyon yöntemi ile normalleştirildikten sonra basit bir sinir ağı ile İYİ_TARIM_UYGULAMALARI tahmin edilmeye çalışılmıştır. Veriler Scikit-learn kütüphanesi ile %77'si (240) eğitim, %33'ü (129) test olacak şekilde iki veri setine ayrıldıktan sonra tüm veri setleri TensorFlow kütüphanesinin tensor veri tipine dönüştürülmüştür.

4 tamamen bağlı katmandan oluşan bir ağ oluşturulmuştur. Çıkış katmanını dışında tüm katmanlarda ReLu aktivasyon fonksiyonu kullanılmıştır. Sırası ile tüm katmanlardaki nöron sayıları şu şekildedir: 278, 140, 70, 1. Optimizasyon algoritması olarak Adam (Adaptif Momentum), kayıp fonksiyonu olarak Binary Crossentropy seçilmiştir.

Modelimiz 15 epoch'ta 5 batch ile eğitildiği zaman %90 doğruluğa ulaşılmıştır.