

# Türkçe Kural Tabanlı Resmi Doküman Tipi Tespiti

Bekir BAKAR, Filiz AKSOY, Apdullah YAYIK, Sevcan İÇÖZ, Vedat AYBAR  
& Prof. Dr. Tunga GÜNGÖR

Konuşmacı: Bekir BAKAR

[bekir.bakar@mobildev.com](mailto:bekir.bakar@mobildev.com)



## Mobildev Kimdir?

## Türkiye’de Kişisel Verilerin Korunmasına Yönelik Düzenlemeler

- Yaptırımlar ve Yükümlülükler

## Datamin: Kişisel Verilerin Kategorilendirilmesi Çözümü

## Doküman Tipleri ve Oluşturulma Yöntemleri

## Doküman Sınıflandırma

- Bilgi Çıkarımı
- Etki Değeri Tanımlanması
- Benzeşme Oranı Tespiti
- Karar Verme

## Deneysel Sonuçlar

- Doğrudan Dijital Ortamda Oluşturulan Dokümanlar
- Sonradan Dijital Ortama Aktarılan Dokümanlar





Mobildev, müşterilerinin satış ve pazarlama ihtiyaçlarına yönelik katma değerli servisler geliştiren, Türkiye'nin önde gelen Servis Sağlayıcı şirketleri ile işbirliği yaparak, portföyüne 18 yılda 23.000'den fazla kurumu katmış bir **yerinde AR-GE** teknoloji şirkettir.



**18+**

2002 yılından bu yana iletişim sektöründe edinilen 18 yıldan fazla tecrübe



**23.000 +**

Her sektörden ve her büyüklükteki 23.000'den fazla müşteri



**Kapsama Alanı**

Türkiye'de bulunan tüm GSM operatörlerinin resmi iş ortağı



**400 M+**

Tüm mesajlaşma ağından geçen aylık 400 milyondan fazla trafik



**Ar-Ge**

100'den fazla farklı platformda, çözümler sunan ürünler yerinde Ar-Ge merkezidir.

# Türkiye’de Kişisel Verilerin Korunmasına Yönelik Düzenlemeler



6698 sayılı *Kişisel Verilerin Korunması Kanunu* (2016) ile birlikte;

## Veri Sorumlusunun Yükümlükleri

- İlgili kişiden gelen taleplerin karşılanması (bilgi hakkı, silme/değiştirme talebi vb. ve GDPR’da bulunan unutulma hakkı)
- Verilerin işleme amaçları sonlandığında gerekli aksiyonların alınması,
- Şikayet halinde 30 gün içerisinde ilgili kişiye cevap verilebilmesi,

## Veri Sorumluluğunun Yerine Getirilmemesi Durumunda Uygulanabilecek Cezai Yaptırımlar

- Kişisel verilerin hukuka aykırı kaydedilmesi (1-3 yıl hapis cezası)
- Hukuka aykırı olarak paylaşılması, yayılması (2-4 yıl hapis cezası)
- Geçerlilik süresi geçmiş olmasına rağmen verilerin silinmemiş olması; (1-2 yıla kadar hapis cezası)
- Veri güvenliği yükümlülüklerinin yerine getirilmemesi (15.000-1 mio TL arası para cezası) vb.



- KVKK ile birlikte gelen yasal sorumluluklara Kurumların uyum gerçekleştirebilmesi için kişisel veri yönetimi sağlanması amacıyla,
- Kanun'da tanımlanmış hali ile kimliği belirli veya belirlenebilir gerçek kişiye ilişkin her türlü bilgiyi
  - Tanımlayıcı (Adı, soyadı, doğum tarihi, vb.)
  - Devlet Kayıtları (TCKN, ehliyet , SGK kayıtları, vb.),
  - Demografik Bilgiler (Cinsiyet, yaş, eğitim vb.),
- Yapısal veya yapısal olmayan ortamlarda kişisel verilerin tespit edilmesi,
  - Yapısal Ortamlar : MsSQL, PostgreSql , vb. veri tabanları
  - Yapısal Olmayan Ortamlar: MS Exchange, Word, pdf , vb.
- Kişisel verilerin ilişkilendirilmesi,



# Doküman Tipleri ve Oluşturulma Yöntemleri

## Sabit Biçimli Doküman

- Eski / Yeni Kimlik Belgesi (D-1 / D-2),
- Eski / Yeni Sürücü Belgesi (D-3 / D-4),
- Adli Sicil Belgesi (D-5),
- Öğrenci Belgesi (D-6) ,
- Nüfus Kayıt Belgesi (D-7),
- Araç Ruhsatı (D-8) ,
- Mezun Belgesi (D-9) ,
- Yerleşim Yeri ve Adres Bilgileri Belgesi (D-10),
- Askerlik Durum Belgesi (D-11)

## Oluşturulma Yöntemleri

- Doğrudan Dijital Ortamda Oluşturulmuş Dokümanlar
- Kamera veya Tarayıcı Aracılığı ile Sonradan Dijital Ortama Aktarılmış Dokümanlar

# Doküman Sınıfandırma: Bilgi Çıkarımı

---

- Doğrudan dijital ortamda oluşturulmuş dokümanlarda (doc, pdf vb.)
  - Bilgi çıkarımı esnasında herhangi bir bilgi kaybı olmamaktadır
  - Bu işlem ilgili dokümandaki bilginin metin olarak alınması şeklinde yapılmıştır
- Kamera veya tarayıcı aracılığı ile sonradan dijital ortama aktarılmış dokümanlar (resimler) için
  - Görüntü iyileştirme
  - Optik karakter tanıma yapılarak görselin metne dönüştürülmesi



# Doküman Sınıflandırma : Etki Değeri Tanımlaması

## Etki Değeri Tanımı

### Kısıtlar

- Ters Frekans değeri arttıkça etki değeri azalmalı
- Alan adı sayısı arttıkça etki değeri azalmalı

$$\text{Etki Değeri} = \frac{\text{Doküman Sayısı}}{(\text{Ters Frekans Değeri}) \times (\text{Toplam Alan Adı Sayısı})}$$

TABLO I: Örnek Etki Değerleri

Doküman Tipi	Alan Sayısı	Alan Adı	Ters Frekans	Etki Değeri
D-2	8	TC Kimlik No	2	$\frac{2}{2 \times 8} = 0.125$
		Geçerlilik Tarihi	1	$\frac{2}{1 \times 8} = 0.250$
D-10	10	TC Kimlik No	2	$\frac{2}{2 \times 10} = 0.100$
		Diploma No	1	$\frac{2}{1 \times 10} = 0.200$



D-2

MEZUN BELGESİ

T.C. Kimlik No	:	
Adı Soyadı	:	
Baba Adı	:	
Anne Adı	:	
Doğum Tarihi	:	
Program	:	YILDIZ TEKNİK ÜNİVERSİTESİ/SOSYAL BİLİMLER ENSTİTÜSÜ/İKTİSAT (YL) (TEZLİ)
Diploma No	:	
Diploma Notu	:	3.12 / 4
Mezuniyet Tarihi	:	23.07.2019
Durum	:	MEZUNİYET

D-10

# Doküman Sınıflandırma: Benzeşme Oranı Tespiti

---

**Algoritma 3:** Benzeşme Oranı Tespiti

---

**Girdi:** Dok, ED,  $E = [e_1, e_2, \dots, e_m]$ ,  $D = [d_1, d_2, \dots, d_m]$

**Çıktı:**  $B = [b_1, b_2, \dots, b_n]_{ilk\_atama=0}$ ,  $T = [t_1, t_2, \dots, t_n]_{ilk\_atama=0}$

```
for each  $d_i$  do
  for each  $d_{ij}$  do
    If KONTROL(RE( $d_{ij}$ ), Dok, ED))
      |  $b_i \leftarrow b_i + e_{ij}$ 
    end
     $t_i \leftarrow t_i + e_{ij}$ 
  end
end
end
```

---

**dok:** Hedef doküman

**E:** Etki değeri

**ED:** En az düzeltme mesafesi

- Girilen dokümanın her bir sınıflara benzeme oranı tespiti
- Düzenli ifadeler ve en az düzeltme mesafesi ile tespit edilen alan isimlerinin etki değerlerinin kümülatif toplamı
- **KONTROL Fonksiyonu**
  - Düzenli ifade
  - Metin
  - En az düzeltme mesafesi

# Doküman Sınıflandırma: Karar Verme

---

**Algoritma 4:** Karar Verme

---

**Girdi:**  $\mathbf{B}$ ,  $\mathbf{T}$ ,  $s$

**Çıktı:**  $k$

**If**  $\max(\mathbf{B}) \geq s \times \mathbf{T}(\arg \max(\mathbf{B}))$

  |  $k \leftarrow \arg \max(\mathbf{B})$

**else**

  |  $k \leftarrow \text{diğer}$

**end**

---

**s:** Eşik değeri (0,25)

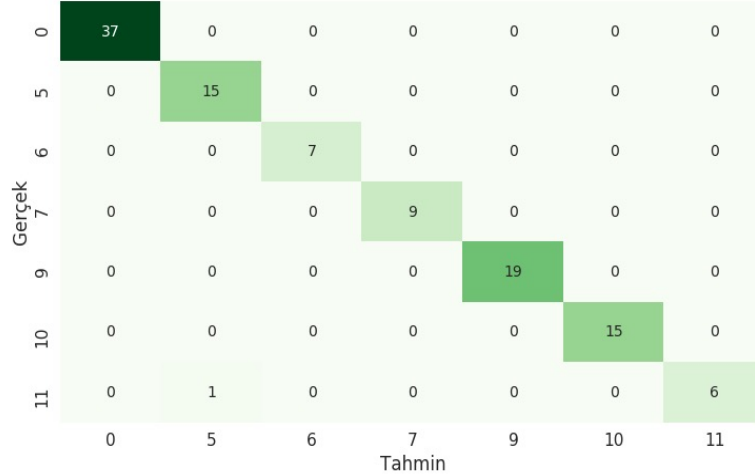
**max(B):** En çok benzeşme

**T:** Doküman tipleri için tam benzeşme oranı

**k:** Dokümana atanan sınıf

- Eşik değeri, her dokümanın tam eşleşme durumunda elde edilecek benzeşme oranına göre kabul sınır değeri çarpanıdır.
- Benzeşme değeri ilgili doküman türüne ait tam benzeşme değerine oranlar belirli bir sınırı aşıyor ise sınıf ataması yapılır.
- Yanlızca, benzeşme oranı en yüksek olan doküman için belirlenen eşik değerden yüksek ise sınıf ataması yapılır.

# Deneysel Sonuçlar



Doğrudan dijital ortamda oluşturulanlar 108 adet doküman için doğrulama matrisi

## Doküman Tipleri

- Eski / Yeni Kimlik Belgesi (D-1 / D-2),
- Eski / Yeni Sürücü Belgesi (D-3 / D-4),
- Adli Sicil Belgesi (D-5),
- Öğrenci Belgesi (D-6) ,
- Nüfus Kayıt Belgesi (D-7),
- Araç Ruhsatı (D-8) ,
- Mezun Belgesi (D-9) ,
- Yerleşim Yeri ve Adres Bilgileri Belgesi (D-10),
- Askerlik Durum Belgesi (D-11)

# Deneysel Sonuçlar

0	58	0	0	0	0	0	0	0	0	0	0	0
1	67	3	1	0	0	0	0	0	0	0	0	0
2	52	0	17	0	0	0	0	0	0	0	0	0
3	19	0	0	4	0	0	0	0	0	0	0	0
4	7	0	0	0	9	0	0	0	0	0	0	0
5	6	0	0	0	0	4	0	0	0	0	0	0
6	8	0	0	0	0	0	3	0	0	0	0	0
7	9	0	0	0	0	0	0	0	0	0	0	0
8	18	0	0	0	0	0	0	0	0	0	0	0
9	4	0	0	0	0	0	1	0	0	9	0	0
10	10	0	0	0	0	0	0	0	0	0	93	0
11	11	0	0	0	0	1	0	0	0	0	0	5
	0	1	2	3	4	5	6	7	8	9	10	11

Kamera veya tarayıcı ile digital ortama aktarılan 417 adet doküman için doğrulama matrisi

## Doküman Tipleri

- Eski / Yeni Kimlik Belgesi (D-1 / D-2),
- Eski / Yeni Sürücü Belgesi (D-3 / D-4),
- Adli Sicil Belgesi (D-5),
- Öğrenci Belgesi (D-6) ,
- Nüfus Kayıt Belgesi (D-7),
- Araç Ruhsatı (D-8) ,
- Mezun Belgesi (D-9) ,
- Yerleşim Yeri ve Adres Bilgileri Belgesi (D-10),
- Askerlik Durum Belgesi (D-11)

# Teşekkürler

Bu çalışma Mobildev A.Ş. tarafından Tübitak destekli olarak geliştirilen *Datamin: Kişisel Verilerin Kategorilendirilmesi Çözümü* kapsamında gerçekleştirilmiştir.

*Datamin* hakkında detaylı bilgi almak için



*Mobildev* hakkında detaylı bilgi almak için

