

Veri Madenciliği ile Ulusal Basında Deprem Haberlerinin Analizi

Deniz BALCI
Dr.öğr.üyesi Levent TERLEMEZ

İstatistik Bölümü
Eskişehir Teknik Üniversitesi

29 Mayıs 2024



- ➊ Giriş
- ➋ Teorik Bölüm
- ➌ Uygulama Bölümü
- ➍ Sonuçlar
- ➎ Referanslar

1 Giriş

2 Teorik Bölüm

3 Uygulama Bölümü

4 Sonuçlar

5 Referanslar

Özet

Günümüzde hızla gelişen sosyal medya platformları, haber kaynakları ve gazeteler gibi çeşitli iletişim kanalları, haberlerin ve olayların hızla yayılmasını sağlayarak önemli bir rol oynamaktadır. Ancak, bu kanallar sadece haber verme amacıyla değil, aynı zamanda yönlendirme ve gündeme müdahale aracı olarak da kullanılabilmektedir. Bu seminer çalışmasında, diğer çalışmalardan farklı olarak, daha genel bir bakış açısı oluşturmak amacıyla veri madenciliği yöntemleri kullanılarak internet üzerindeki "deprem haberlerine" ilişkin ulusal medyadaki haberlerin içeriği karşılaştırmalı olarak incelenmiştir.

1 Giriş

2 Teorik Bölüm

Haber ve Medya

Veri Madenciliği Nedir Süreci?

Veri Madenciliği Yöntemleri

3 Uygulama Bölümü

4 Sonuçlar

5 Referanslar

1 Giriş

2 Teorik Bölüm

Haber ve Medya

Veri Madenciliği Nedir Süreci?

Veri Madenciliği Yöntemleri

3 Uygulama Bölümü

4 Sonuçlar

5 Referanslar

Haber ve Medya Etkisi

Medya, internetin insanlara ulaşmadaki gücünden ve hızından yararlanmak amacıyla, internet ortamına girmiştir. Gazeteler, teknolojik gelişmelerdeki hızlı ilerlemenin sonucunda geleneksel gazete yayımlamanın yanı sıra, okura yeni bir yöntemle ulaşabilmenin fırsatını elde etmişlerdir. İnternet gazeteleri, kamuoyunu doğru bilgilendirmekle yükümlü olmasına rağmen, medya-siyaset-sermaye ilişkisi, kamuoyuna aktarılan bilginin doğruluğunu şüpheli hale getirmiştir. [2]

Medyanın toplum üzerinde etkisi

Genelde insanların medya tarafından verilen her şeyden etkilendikleri herkesçe kabul gören bir gerçektir. Medyada verilenler geleneksel yapıya zarar verdiği gibi bütün toplumlarda da büyük değişikliklere neden olmaktadır. Özellikle gelişmekte olan ülkelerin bu tek yönlü haber ve bilgi akışından etkilendikleri kesindir.[3]

1 Giriş

2 Teorik Bölüm

Haber ve Medya

Veri Madenciliği Nedir Süreci?

Veri Madenciliği Yöntemleri

3 Uygulama Bölümü

4 Sonuçlar

5 Referanslar

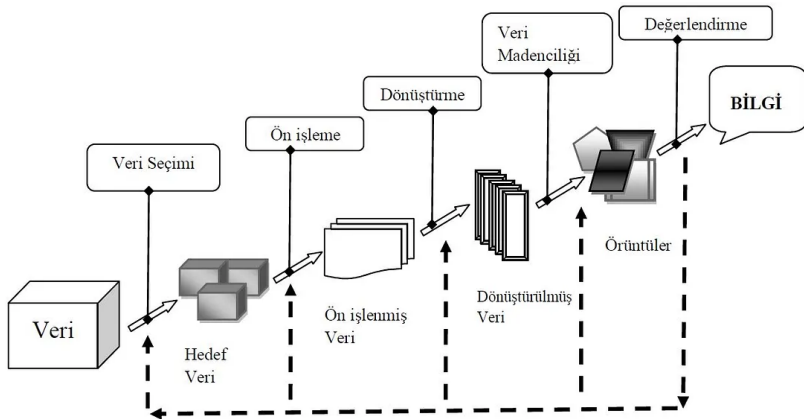
Veri Madenciliği Nedir

Veri tabanları veya dosyalarda bulunan verilerin belirli istatistiksel yöntemlerden yararlanılarak, kullanılabilir hale getirilmesi işlemine veri madenciliği denir. [5]

Bir çok veri madenciliği sürecinin kendine has adımları olsa da veri madenciliği sürecinden izlediği adımlar genellikle şöyledir[6]

- 1 Problemin tanımlanması
- 2 Verilerin hazırlanması
- 3 Modelin kurulması ve değerlendirilmesi
- 4 Modelin kullanılması
- 5 Modelin izlenmesi

Veri Madenciliği Süreci



Şekil 1: Veri Madenciliği Süreci

Şeklinde açıklayabiliriz.

Problemin tanımlanması Veri madenciliği çalışmalarında başarılı olmanın en önemli şartı, projenin hangi işletme amacı için yapılacağını ve elde edilecek sonuçların başarı düzeylerinin nasıl ölçüleceğinin tanımlanmasıdır. [6]

Verilerin hazırlanması Modelin kurulması aşamasında ortaya çıkacak sorunlar, bu aşamaya sık sık geri dönülmesine ve verilerin yeniden düzenlenmesine sebep olacaktır. Bu durum verilerin hazırlanması ve modelin kurulması aşamaları için, bir analistin veri keşfi sürecinin toplamı içerisinde enerji ve zamanının %50 – %85 harcamasına neden olmaktadır. [6]

Modelin kurulması ve değerlendirilmesi Tanımlanan problem için en uygun modelin bulunabilmesi, olabildiğince çok sayıda modelin kurularak denenmesi ile mümkündür. Bu nedenle veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varılıncaya kadar yinelenen bir süreçtir (Zaki Mohammed, 2003) [6]

Modelin Kullanılması ve Güncellenmesi Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak kullanılabilir. Kurulan modeller risk analizi, kredi değerlendirme, dolandırıcılık tespiti gibi işletme uygulamalarında doğrudan kullanılabilirliği gibi, promosyon planlaması simülasyonuna entegre edilebilir veya tahmin edilen envanter düzeyleri yeniden sipariş noktasının altına düştüğünde, otomatik olarak sipariş verilmesini sağlayacak bir uygulamanın içine gömülebilir. Zaman içerisinde bütün sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde ortaya çıkan değişiklikler, kurulan modellerin sürekli olarak izlenmesini ve gerekiyorsa yeniden düzenlenmesini gerektirecektir. [7]

Modelin izlenmesi Zaman içerisinde bütün sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde ortaya çıkan değişiklikler, kurulan modellerin sürekli olarak izlenmesini ve yeniden düzenlenmesini gerektirecektir. Tahmin edilen ve değişkenler arasındaki farklılığı gösteren grafikler model sonuçlarının izlenmesinde kullanılan yararlı bir yöntemdir [6]

1 Giriş

2 Teorik Bölüm

Haber ve Medya

Veri Madenciliği Nedir Süreci?

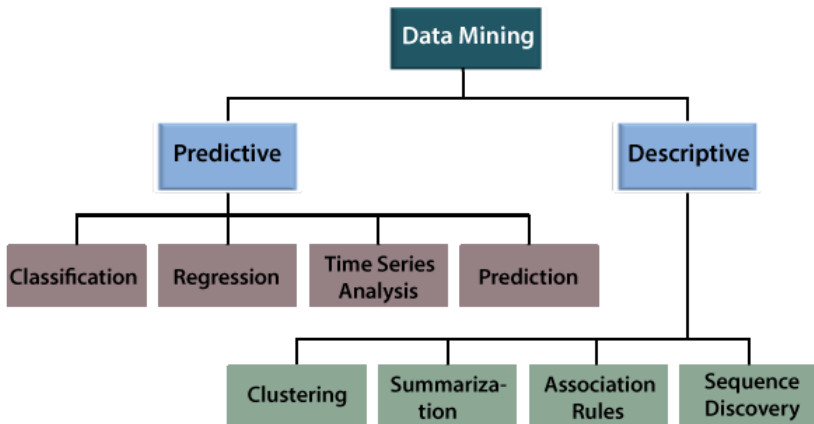
Veri Madenciliği Yöntemleri

3 Uygulama Bölümü

4 Sonuçlar

5 Referanslar

Başlıca Veri Madenciliği Yöntemleri



1 Giriş

2 Teorik Bölüm

3 Uygulama Bölümü

Kullanılan Araçlar Ve Kütüphaneler
Verilerin Elde edilmesi ve Temizlenmesi
Modelin kurulması ve değerlendirilmesi

4 Sonuçlar

5 Referanslar

1 Giriş

2 Teorik Bölüm

3 Uygulama Bölümü

Kullanılan Araçlar Ve Kütüphaneler

Verilerin Elde edilmesi ve Temizlenmesi

Modelin kurulması ve değerlendirilmesi

4 Sonuçlar

5 Referanslar

Kullanılan Araçlar

- 1 **Jupyter Notebook:**Jupyter Notebook, veri bilimi, makine öğrenmesi ve daha birçok alanda verimli bir geliştirme ortamı sağlayan güçlü bir araçtır.
- 2 **Pip:**Kullandığımız kütüphanelerin ve paketlerin yönetimini sağlayan sistemdir.
- 3 **Custom Search JSON API:**Programlanabilir arama motorundan programlı bir şekilde arama sonuçları almak ve görüntülemek için web siteleri ve uygulamalar geliştirmenizi sağlar. [8]
- 4 **NumPy:**NumPy (Numerical Python), bilimsel hesaplama için geniş kapsamlı destek sağlayan açık kaynaklı bir Python kütüphanesidir.
- 5 **Natural Language Toolkit (NLTK):**Natural Language Toolkit (NLTK), Python programlama dili için açık kaynaklı bir doğal dil işleme (NLP) kütüphanesidir.

Kütüphaneler:

- ❶ **BeautifulSoup:** Python tabanlı web scraping kütüphanelerinden birisidir. İçinde bir HTML parser bulundurur ve bu parser metodları üzerinden HTML içeriğindeki spesifik konumlara erişmemizi sağlar. Bu kütüphane sayesinde HTML ve XML dosyalarını rahatlıkla inceleyebiliriz.
- ❷ **Requests:** Requests kütüphanesi, Python'da HTTP (web istekleri) yapmamızı sağlayan kütüphanedir. Verisini almak istediğimiz web sitelere erişim için bu kütüphaneyi kullanacağız.
- ❸ **matplotlib:** Matplotlib, NumPy için görselleştirme eklentisidir.
- ❹ **pandas:** Pandas, veri manipülasyonu ve veri analizi için yazılmış bir kütüphanedir.
- ❺ **wordcloud:** Kelime bulutu oluşturmak için kullanılan bir kütüphanedir.

1 Giriş

2 Teorik Bölüm

3 Uygulama Bölümü

Kullanılan Araçlar Ve Kütüphaneler

Verilerin Elde edilmesi ve Temizlenmesi

Modelin kurulması ve değerlendirilmesi

4 Sonuçlar

5 Referanslar

Verilerin Elde edilmesi

Veriler, Google arama motorundaki ilk 30 web sitesinden BeautifulSoup adlı kütüphane kullanılarak elde edilmiştir. Bu web sitelerinden alınan ilk paragraflar bir TXT dosyasına kaydedilmektedir. Kısıtlı veri ve araçlarla çalışmamızın sebebi, çoğu kullanıcının arama motorlarında genellikle ilk sayfayı baz almasıdır.

Not: İnternet reklam firması Chitika'nın araştırma biriminin yaptığı araştırmadan elde edilen veriler de, Google arama sonuç sayfalarından ilkinin tıklanma oranı %94 iken, ikinci sayfanın oranı sadece %6'dır."(Jensen, 2011) [11].

- 1 Google Custom Search JSON API üzerinden bağlantılar alınır.
- 2 Elde edilen bağlantılar kontrol edilir (blacklist) ve paragraflar elde edilir.
- 3 Bu veriler temizliğe gönderilir.

Verilerin Temizlenmesi

Elde edilen veriler aşağıdaki adımlarla temizlenir.

- 1 Alıntılar ve fazla boşluklar silinir.
- 2 Özel karakterler kaldırılır.
- 3 Bütün kelimeler küçük harfe çevrilir ve tokenize edilir.
- 4 Bazı köke sahip kelimeler ayrıştırılır(deprem,depremler gibi)
- 5 Temizlenen veriler txt dosyasına aktarılır.

1 Giriş

2 Teorik Bölüm

3 Uygulama Bölümü

Kullanılan Araçlar Ve Kütüphaneler
Verilerin Elde edilmesi ve Temizlenmesi
Modelin kurulması ve değerlendirilmesi

4 Sonuçlar

5 Referanslar

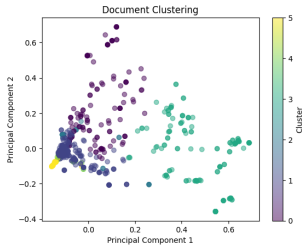
Modelin kurulması ve değerlendirilmesi

Temizlediğimiz veriler için Temel olarak 3 farklı yöntemi kullanmayı tercih ediyoruz.

- 1 Wordcloud-Frekans analizi
- 2 K means clustering TF-IDF vectorization
- 3 N-gram

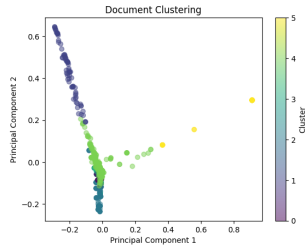
26 / 35

K means clustering TF-IDF vectorization



Şekil 4: Deprem haberleri

- $c = 6$ Silhouette Score: 0.0708...
- $c = 20$ Silhouette Score: 0.1060..



Şekil 5: Hatay Depremi

- $c = 6$ Silhouette Score: 0.084...
- $c = 20$ Silhouette Score: 0.153...

N-GRAM

```
[('mayıs', 'afad', 'kandilli', 'deprem', 'listesi'), 23), (('deprem', 'mi', 'oldu', 'son', 'depremler'), 19),
(('mi', 'oldu', 'son', 'depremler', 'bugün'), 19), (('oldu', 'son', 'depremler', 'bugün', 'istanbulda'), 18),
(('son', 'depremler', 'bugün', 'istanbulda', 'deprem'), 16), (('depremler', 'bugün', 'istanbulda', 'deprem', 'm
i'), 16), (('bugün', 'istanbulda', 'deprem', 'mi', 'oldu'), 16), (('mayıs', 'ankarada', 'izmirde', 'deprem', 'm
i'), 16), (('ankarada', 'izmirde', 'deprem', 'mi', 'oldu'), 16), (('deprem', 'mi', 'oldu', 'son', 'dakika'), 1
5), (('mayıs', 'dün', 'gece', 'deprem', 'mi'), 15), (('dün', 'gece', 'deprem', 'mi', 'oldu'), 15), (('uzmanı',
'prof', 'dr', 'naci', 'görür'), 15), (('izin', 'alınmadan', 'kaynak', 'gösterilerek', 'dahi'), 13), (('alınmada
n', 'kaynak', 'gösterilerek', 'dahi', 'iktibas'), 13), (('kaynak', 'gösterilerek', 'dahi', 'iktibas', 'edileme
z'), 13), (('arasında', 'ilk', 'sıralarda', 'yer', 'alıyor'), 13), (('kanal', 'd', 'nin', 'üç', 'sezondur'), 1
3), (('meydana', 'geldi', 'afet', 'acil', 'durum'), 12), (('geldi', 'afet', 'acil', 'durum', 'yönetim'), 12),
(('afet', 'acil', 'durum', 'yönetim', 'başkanlığı'), 12), (('acil', 'durum', 'yönetim', 'başkanlığı', 'afad'), 1
2), (('durum', 'yönetim', 'başkanlığı', 'afad', 'son'), 12), (('yönetim', 'başkanlığı', 'afad', 'son', 'depreml
e'), 12), (('başkanlığı', 'afad', 'son', 'depremler', 'listesine'), 12), (('afad', 'son', 'depremler', 'listesin
e', 'göre'), 12), (('depremleri', 'anlık', 'olarak', 'haberimizden', 'takip'), 12), (('anlık', 'olarak', 'haber
imizden', 'takip', 'edebilirsiniz'), 12), (('küçük', 'olan', 'depremler', 'iste', 'son'), 12), (('olan', 'depreml
er', 'iste', 'son', 'depremler'), 12), (('depremler', 'iste', 'son', 'depremler', 'listesinde'), 12), (('iste',
'son', 'depremler', 'listesinde', 'görüntülenmektedir'), 12), (('köşe', 'yazıları', 'magazinden', 'siyasete', 's
pordan'), 12), (('yazıları', 'magazinden', 'siyasete', 'spordan', 'seyahate'), 12), (('magazinden', 'siyasete',
'spordan', 'seyahate', 'bütün'), 12), (('siyasete', 'spordan', 'seyahate', 'bütün', 'konuların'), 12), (('sporda
```

Şekil 6: Deprem haberleri

N-GRAM Hatay Depremi

```
[('mayıs', 'afad', 'kandilli', 'deprem', 'listesi'), 23), (('deprem', 'mi', 'oldu', 'son', 'depremler'), 19),
(('mi', 'oldu', 'son', 'depremler', 'bugün'), 19), (('oldu', 'son', 'depremler', 'bugün', 'istanbulda'), 18),
(('son', 'depremler', 'bugün', 'istanbulda', 'deprem'), 16), (('depremler', 'bugün', 'istanbulda', 'deprem', 'm
i'), 16), (('bugün', 'istanbulda', 'deprem', 'mi', 'oldu'), 16), (('mayıs', 'ankarada', 'izmirde', 'deprem', 'm
i'), 16), (('ankarada', 'izmirde', 'deprem', 'mi', 'oldu'), 16), (('deprem', 'mi', 'oldu', 'son', 'dakika'), 1
5), (('mayıs', 'dün', 'gece', 'deprem', 'mi'), 15), (('dün', 'gece', 'deprem', 'mi', 'oldu'), 15), (('uzmanı',
'prof', 'dr', 'naci', 'görür'), 15), (('izin', 'alınmadan', 'kaynak', 'gösterilerek', 'dahi'), 13), (('alınmada
n', 'kaynak', 'gösterilerek', 'dahi', 'iktibas'), 13), (('kaynak', 'gösterilerek', 'dahi', 'iktibas', 'edileme
z'), 13), (('arasında', 'ilk', 'sıralarda', 'yer', 'alıyor'), 13), (('kanal', 'd', 'nin', 'üç', 'sezondur'), 1
3), (('meydana', 'geldi', 'afet', 'acil', 'durum'), 12), (('geldi', 'afet', 'acil', 'durum', 'yönetim'), 12),
(('afet', 'acil', 'durum', 'yönetim', 'başkanlığı'), 12), (('acil', 'durum', 'yönetim', 'başkanlığı', 'afad'), 1
2), (('durum', 'yönetim', 'başkanlığı', 'afad', 'son'), 12), (('yönetim', 'başkanlığı', 'afad', 'son', 'depreml
e'), 12), (('başkanlığı', 'afad', 'son', 'depremler', 'listesine'), 12), (('afad', 'son', 'depremler', 'listesin
e', 'göre'), 12), (('depremleri', 'anlık', 'olarak', 'haberimizden', 'takip'), 12), (('anlık', 'olarak', 'haber
imizden', 'takip', 'edebilirsiniz'), 12), (('küçük', 'olan', 'depremler', 'iste', 'son'), 12), (('olan', 'depreml
er', 'iste', 'son', 'depremler'), 12), (('depremler', 'iste', 'son', 'depremler', 'listesinde'), 12), (('iste',
'son', 'depremler', 'listesinde', 'görüntülenmektedir'), 12), (('köşe', 'yazıları', 'magazinden', 'siyasete', 's
pordan'), 12), (('yazıları', 'magazinden', 'siyasete', 'spordan', 'seyahate'), 12), (('magazinden', 'siyasete',
'spordan', 'seyahate', 'bütün'), 12), (('siyasete', 'spordan', 'seyahate', 'bütün', 'konuların'), 12), (('sporda
```

Şekil 7: Hatay haberleri

- 1 Giriş
- 2 Teorik Bölüm
- 3 Uygulama Bölümü
- 4 Sonuçlar
- 5 Referanslar

Sonuçlar

Bu çalışmanın amacı, ulusal basında yer alan deprem haberlerinin analiz edilmesi ve bu haberlerde kullanılan dilin, öne çıkan kelimelerin ve temaların belirlenmesiydi. Çalışmamızda veri madenciliği yöntemleri kullanarak, deprem haberlerinin kapsamlı bir analizini gerçekleştirdik.

- 1 Giriş
- 2 Teorik Bölüm
- 3 Uygulama Bölümü
- 4 Sonuçlar
- 5 Referanslar**

- [1] <https://tr.wikipedia.org/wiki/>
- [2] <https://dergipark.org.tr/tr/download/article-file/66270> giriş bölümü
- [3] <https://dergipark.org.tr/tr/download/article-file/212342> sonuç bölümü
- [4] <http://nek.istanbul.edu.tr:4444/ekos/TEZ/32994.pdf>
6.bölüm internet gazeteciliğinin medya içinde yarattığı etkiler
- [5] https://ab.org.tr/ab11/kitap/tekerek_AB11.pdf
- [6] <https://medium.com/veri-madencili%C4%9Fi/veri-madencili%C4%9Fi-s%C3%BCreci-11243e7966fc>
- [7] <https://dergipark.org.tr/tr/download/article-file/32076>

- [8] <https://developers.google.com/custom-search/v1/overview?hl=tr>
- [9] Türkiye’de Ulusal ve Yerel Gazetelerin Habercilik Anlayışları: Deprem Haberlerinin Karşılaştırmalı Analizi-
<https://dergipark.org.tr/tr/pub/josc/issue/19013/200756>
- [10] <https://dergipark.org.tr/tr/download/article-file/3509650>
- [11] KAMUSAL ALAN OLARAK İNTERNET VE GELENEKSEL MEDYANIN KARŞILAŞTIRILMASI: ARAMA MOTORLARI ÖRNEĞİNDE KAMUSAL ALANIN OLUŞUMU-Nurhan Yel

Dinlediğiniz için Teşekkürler