



Öğretim Görevlisi: Dr. Öğr. Üyesi Nur Kuban Torun

Dersin Adı: Veri Madenciliği

Öğrenci Adı: Ahmet İZMİR

Öğrenci Numarası: 25174486726

Uzay Nedir?

Uzay veya feza, Dünya'nın ötesinde ve gök cisimleri arasında var olan, sonsuz olduğu düşünülen fakat sonsuz olduğu konusunda kesin yargılara varılamayan genişliktir.

Uzay ve Veri Madenciliği

Ülkemizde ve dünyada uzaya dair hızla artan bu ilgiyi ve bu ilginin sonucunda oluşan gelişmeleri, insanların bu gelişmelere karşı bakış açısını gözlemlemekle kalmayıp bunu analizler içeren dokümanlar haline getirmek bizlere fayda sağlayacaktır. Bu çalışma ile amaç güncel uzay programları ve çalışmalarının R programlama dili üzerinden duygu analizini ve metin madenciliğini yapmaktır.

Veri Madenciliği

Veri madenciliği, büyük ölçekli veriler arasında faydalı bilgiye ulaşma, bilgiyi madenleme işidir. Büyük veri yığınları arasında gelecekle ilgili tahminlerde bulunabilmemizi sağlayacak bağlantıların bilgisayar programı kullanarak aranması işlemidir.

Metin Madenciliği

Metin madenciliği yapısal olmayan ve düzensiz haldeki metin yığınlarından önceden bilinmeyen, potansiyel olarak kullanışlı, yapısal ve düzenli veri elde etme sürecidir. Elde edilen veri ile veri ile analizler yapılır.

Duygu Analizi

Duygu analizi bir konu hakkındaki bir çalışmanın belirli bir konuya yönelik olumlu, olumsuz ve nötr olarak değerlendirilmesi için metin parçasıyla belirtilen görüşleri algoritmalar aracılığı ile hesaplayıp tanımlama ve sınıflandırma süreçlerine denir. Günümüzde sosyal medyanın yaygınlığı nedeniyle duygu analizinde kullanacağımız verileri sosyal medyadan bulabiliyoruz.

Bu çalışmada Uzay ile ilgili Twitter'da paylaşılan tweetler arasından Türkiye'den paylaşılan 30 adet paylaşımın duygu analizi gerçekleştirildi. Bu çalışmada 30 adet tweetten durak kelimeler, sayılar çıkartılıp düzenlendikten sonra 36 pozitif kelime ve 10 negatif kelime hesaplandı ve duygu skoru 26 olarak hesaplandı.

	A	B	C	D	E	F	G	H	I	J	K	L
	Tweetler	Pozitif	Nötr	Negatif		Pozitif Kelimeler	Negatif Kelimeler	P-N	TP	TN	Duygu Skoru	Toplam Kelime
1	gördüğünüz şey meteor veya yıldız kayması değil görüntüdeki ışıklar elon muskin spacex firmasına ait enkazlar			2			değil-enkazlar	-2	36	-10	26	61
2	güneş sistemi ile ilgili güzel bir program olmuş emeğinize sağlık	2				güzel-sağlık		2	0,3436747	-0,9569423		
3	james webb kindi yörünge ateşlemesini tamamladı teleskop planlanan ateşlemeden ikincisini başarıyla	3				tamamladı-başarıyla-gerçekleştirdi		3				
4												

	A	B	C	D	E	F	G	H
5	bursa da evinde yaptığı teleskoplarla uzayı keşfediyor	1				keşfediyor		1
6	planlanan ateşlemeden ikincisini başarıyla gerçekleştirdi	2				başarıyla-gerçekleştirildi		2
7	herhangi bir aksilik olmazsa yarın aralık tarihinde james webb uzay teleskopu tarihi yolculuğuna çıkacak			1			aksilik	-1
8	uzay yolculuğunda kritik adımlardan biri başarıyla kalkanları açıldı	2				başarıyla-açıldı		2

	A	B	C	D	E	F	G	H	I
19	süpernova ve dünyadaki yaşam arasında olağanüstü bir bağlantı keşfedildi	3				süpernova-olağanüst-keşfedildi		3	
20	uzayda da süper güç olacağız	2				süper-güç		2	
21	insan yapımı mit araştırmacıları bilim kurgu filmlerindeki ufo lara benzer bir aracı tasarladı	2				bilim-tasarladı		2	
22	bilim insanları yok oluş bilmeceğini çözdü	2		1		bilim-çözdü	yok oluş	1	
23	canlı renge sahip kumaş kalem ile farklı tişörtler boyamaya hazır mısın belki bir gezegen belki astronot bi ay	1				canlı		1	

	A	B	C	D	E	F	G	H
29	bilim insanları uzayda zor koşullar insanları yamyamlığa yöneltebilir			2			zor-yamyamlığa	2
30	dünyanın en değerli değersiz parası	1		1		değerli	değersiz	0
31	en azından nasa birşeyler yapmaya çalışıyorsa biz türkler daha ekonomik sorunu çözmüş değiliz	1		2		çalışıyor	çözmüş-değiliz	-1
32								
33								
34	TOPLAM POZİTİF=	42						26
35								
36								
37	TOPLAM NEGATİF=	19						
38								

İçerik Analizi

Bu içerik analizinde 25.12.2021 – 15.01.2022 tarihleri arasında Uzak hakkında Twitter’da paylaşılan tweetleri kullandım. Kendi dilimizde yani Türkçe yaptığım bu analizde ülkemizde son zamanlarda artan uzak ve havacılık ilgisiyle beraber güncel ve her gün çekilecek tweet sayısının artarak gideceğini düşündüm. Bu düşünce ile beraber çektiğim tweetlerde her gün artarak ilerleyen bir grafik yakaladım. Ancak tweetler artarak yükselmesine rağmen istediğim tweet sayısına ulaşamadım. Bu nedenle tekrar eden kelimelerde düşük seviyeler elde ettim.

5 Kere Tekrar Eden Kelimeler

```
> findFreqTerms(tdm, lowfreq = 5)
```

[1] "adım"	"uzay"	"bir"	"için"	"nasa"	"uzaya"	"ilk"	"."
[9] "elonmusk"	"space"	"spacex"	"astronomi"	"fotoğrafı"	"günün"	"gezegen"	"ile"
[17] "yeni"	"yaşam"	"uzayda"	"büyük"	"dünyanın"	"havacılık"	"teknoloji"	"alındı"
[25] "grizua"	"ulusalkanal"	"uydu"	"veri"	"zonguldak"	"cep"	"türkiye"	"mars"
[33] "roket"	"fotoğraf"	"the"	"araçların"	"biran"	"bunun"	"gerekıyor"	"girmesi"
[41] "hayatımza"	"kazadan"	"kurtulmaya"	"mywayturkey"	"otonom"	"refleksi"	"önce"	"insan"
[49] "işte"	"bütünleşme"	"canlı"	"dikkat"	"dünyamız"	"edin"	"göreceğiniz"	"lütfe"
[57] "organizma"	"patla..."	"saat"	"tek"	"volkanının"	"yeraltı"	"yönünde"	"▲"
[65] "için"	"evrenin"	"sondakika"	"güneş"	"uzay..."	"marsta"	"değil"	"tarih"
[73] "ilk"	"bugün"	"ocak"	"bilim"	"evren"	"fizik..."	"okumak"	"tamamını"
[81] "yazının"	"@"	"..."	"sadece"	"alanında"	"avorhankircali"	"fırlatma"	"geliştirmeye"
[89] "gerek"	"gerekse"	"gözümüz"	"sistemleri"	"yönelik"	"çalışmalarım..."	"science"	"ana"
[97] "♥"	"gökbatan"	"timesofturkiye"	"yolculuğu"	"türkiye'nin"	"bin"	"program"	"yılında"
[105] "adı"	"türkiyenin"	"uydusu"	"türkiye'nin"	"fırlatıldı"	"her"	"teknofest"	"☐"
[113] "grizu"	"endoğan"	"kanala"	"konuştu"	"takım"	"ulusal"	"öğrencileri"	"fırlatılıyor"
[121] "magazinsortie"							

Çıkan kelimelere baktığımızda kelime fazlalığından kaynaklanan veri analizi zorluğu dikkat çekiyor. Nasa, SpaceX, Teknofest gibi olgulara olan ilginin fazlalığı dikkat çekiyor. Bu da insanların güncel, teknolojik uzak gelişmeleri ile ilgilendiğinin büyük bir örneği.

10 ve 15 Kere Tekrar Eden Kelimeler

```
> findFreqTerms(tdm, lowfreq = 10)
```

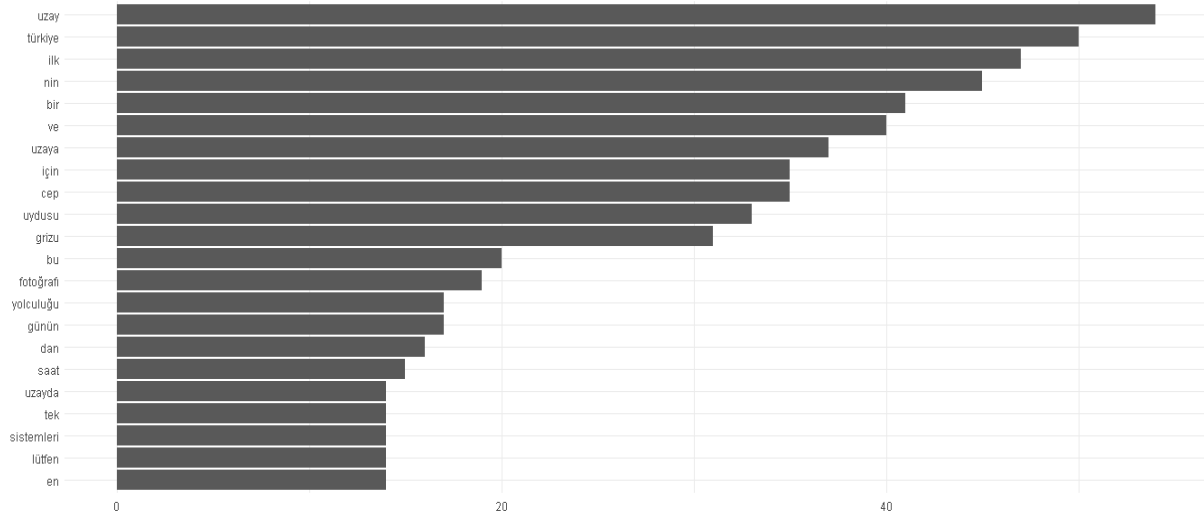
[1] "uzay"	"bir"	"için"	"nasa"	"uzaya"	"ilk"	"."	"space"
[9] "astronomi"	"fotoğrafı"	"günün"	"ile"	"uzayda"	"büyük"	"teknoloji"	"grizua"
[17] "uydu"	"cep"	"türkiye"	"mywayturkey"	"bütünleşme"	"canlı"	"dikkat"	"dünyamız"
[25] "edin"	"göreceğiniz"	"lütfe"	"organizma"	"patla..."	"saat"	"tek"	"volkanının"
[33] "yeraltı"	"yönünde"	"▲"	"için"	"uzay..."	"ilk"	"bilim"	"evren"
[41] "@"	"..."	"fırlatma"	"sistemleri"	"gökbatan"	"timesofturkiye"	"yolculuğu"	"türkiye'nin"
[49] "türkiyenin"	"uydusu"	"fırlatıldı"	"grizu"	"ulusal"			

```
> findFreqTerms(tdm, lowfreq = 15)
```

[1] "uzay"	"bir"	"için"	"nasa"	"uzaya"	"ilk"	"."	"space"
[9] "astronomi"	"fotoğrafı"	"günün"	"uzayda"	"grizua"	"uydu"	"cep"	"türkiye"
[17] "mywayturkey"	"canlı"	"saat"	"ilk"	"bilim"	"evren"	"..."	"gökbatan"
[25] "timesofturkiye"	"yolculuğu"	"türkiye'nin"	"türkiyenin"	"uydusu"			

10 ve 15 kere tekrar eden kelimelere baktığımızda daha çok millileşen olguların ön plana çıkmış olduğunu görüyoruz. Analizde sadece Türkçe veriler çekildiği için Türkiye’de uzaya olan ilginin ortaya çıkarılmış olması tekrarlanan kelime sayısını yüksek olarak almanın önemini gösteriyor. Timesofturkiye, gökvatan,mywayturkey gibi tweet başlıklarından bunu çok daha iyi anlayabiliyoruz

Tweetlerde en çok kullanılan kelimeler



Kaynakça:

https://tr.wikipedia.org/wiki/Verim_adencili%C4%9Fi

<http://www.metinmadenciligi.com/>