SELÇUK ÜNİVERSİTESİ
SOSYAL BİLİMLER MESLEK YÜKSEKOKULU DERGİSİ (e-ISSN: 2564-7458)

Cilt 22 ● Sayı 2 ● Kasım 2019
SS. 589 - 600 / Araştırma Makalesi

GDELT Kullanarak Toplumsal Huzursuzlukların Tahmin Edilmesi: Tunus Örneği

Using GDELT Estimation of Social Unrest: The Tunisia Example

*Sadullah ÇELİK**

ÖZ

Günümüzde, toplumsal huzursuzluklar (protestolar, grevler, çatışmalar ve işgal olayları) birçok ülkenin sınırlarının ve siyasal yapılarının şekillenmesinde ve değişmesinde etkin bir rol almaktadır. Gerek demokrasilerde ve gerek otoriter rejimlerde toplumsal huzursuzlukların proaktif olarak ele alınması hükümet ve politika yapıcılar için büyük öneme sahiptir. Bugün geliştirilen GDELT projesi sayesinde artık toplumsal olaylar gerçek zamanlı izlenebilmekte ve bu sayede ülkelerin gelecekte yaşaması muhtemel süreçler tahmin edilebilmektedir. Bu çalışmada, ülkelerin istikrarsızlıkları ile ilgili huzursuzluk olaylarını tespit etmek için hesaplamalı bir yaklaşım kullanılmıştır. Bunun için tarihteki belli bir zaman aralığın da (30 günlük bir pencerede) benzer kalıpları (desenleri) tespit etmek için Google BigQuery'nin Pearson korelasyon özelliği kullanılmıştır. Çalışmada, Tunus'un 25 Temmuz 2013 sorasındaki 30 günlük süreçte yaşananlar, GDELT'ten elde edilen çeşitli olay verileri kullanılarak tahmin edilmiştir. Tunus'ta gerçekte yaşananlar tahmin sonuçları ile karşılaştırıldığında r=0.725 gibi yüksek bir korelasyon katsayısı bulunmuştur. Elde edilen bu korelasyon katsayısı Tunus için yapılan tahminlerin doğruluğunun güvenilir olduğunu göstermektedir.

ANAHTAR KELİMELER

GDELT, BigQuery, Pearson Korelasyonu, Huzursuzluk Olayları, Tahmin

ABSTRACT

Today, social unrest (protests, strikes, conflicts and occupation events) plays an active role in shaping and changing the borders and political structures of many countries. The proactive handling of social unrest, both in democracies and in authoritarian regimes, is of great importance for government and policy-makers. Thanks to the GDELT project developed today, social events can now be monitored in real time, thus predicting the future processes of countries. In this study, a computational approach is used to detect the incidents of unrest related to the instability of countries. To do this, Google BigQuery's Pearson correlation feature was used to identify similar patterns (patterns) for a specific time period in a date (in a 30-day window). In the study, what happened during the 30-day period of 25 July 2013 in Tunisia was estimated by using various event data obtained from GDELT. A high correlation coefficient of r = 0.725 was found when compared to the actual results in Tunisia. This correlation coefficient shows that the estimations made for Tunisia are reliable.

KEYWORDS

GDELT, BigQuery, Pearson Correlation, Events of Unrest, Prediction

Makale Geliş Tarihi / Submission Date 19.07.2019	*Makale Kabul Tarihi / Date of Acceptance* 26.08.2019
Atıf	Çelik, S. (2019). GDELT Kullanarak Toplumsal Huzursuzlukların Tahmin Edilmesi: Tunus Örneği. *Selçuk Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Dergisi*, 22 (2), 589-600.

* Arş. Gör. Dr. Aydın Adnan Menderes Üniversitesi, Nazilli İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, ssadullah.celik@gmail.com, ORCID: 0000-0001-5468-475X

GİRİŞ

Son çeyrek yüzyıldaki çalışmaların arkasındaki itici güçlerden biri, davranışlarımızı temel alan kalıplardır. Bu kalıplar, bizi insan yapan anlatıların ve duyguların küresel toplumu anlamamıza yardımcı olmaktadır. Ayrıca bu kalıplar büyük veri setlerini ve bilgi işlem platformlarını nasıl kullanabileceğimiz konusunda bizlere yol göstermektedir. Bilgi işlem platformları, geçmişimizi kataloglamakta ve görselleştirmekte, tarihin kalıplarını bulmakta ve tüm anlatılanları ve hikayeleri bir araya getirmektedir. Dijital ve sayısallaştırılmış tarihimiz kitlesel hesaplama gücü ile birleştirildiğinde ve yönetilerek yönlendirildiğinde büyük hayranlık uyandırmaktadır. Bilim insanları; kim olduğumuza, nerede olduğumuza ve nereye gittiğimize dair soruların veriye dayalı cevaplarının insanoğlunun geleceğinin keşfedilmesini sağlayabilir mi? Sorusuna cevap aramışlardır.

Sivil huzursuzluk hem demokrasilerde hem de otoriter rejimlerde sık görülen bir durumdur. Bu olayları tespit etmek ve tahmin etmek sosyal bilimciler ve politika belirleyiciler için kilit öneme sahiptir. Çünkü bu olaylar önemli toplumsal ve kültürel değişikliklere yol açabilmektedir. Siyaset bilimciler için, sivil huzursuzluklar, vatandaşların görüşlerini ve tercihlerini otoritedeki kişilere yansıtma konusunda yardımcı olduğu için önemlidir. Bu nedenle, gelecekteki zamanın söz konusu haberlerde ve sosyal medyada tespit edilmesi, protesto tahmin sistemi geliştirmenin doğrudan bir yoludur. Bu gizli istikrarsızlıkların ortaya çıkmadan önce tahmin edilmesi birçok alan için önemlidir. Turizm endüstrisi için bu huzursuzlukların önceden tahmin edilmesi, seyahat uyarıları yapılmasını sağlayabilir. Hükümetler için bir protesto tahmin sistemi, vatandaşların şikâyetlerine öncelik verilmesini sağlayabilir. Kolluk kuvvetleri için gösterilere hazırlıklı olma konusunda yardımcı olabilir. Sosyal bilimciler için protesto tahminleri, vatandaşların kendilerini nasıl ifade ettikleri konusunda bir fikir verebilir. Son olarak, protestolar birden fazla endüstri üzerinde (özellikle dünya çapında tedarik zinciri yönetimine dayananlar) zayıflatıcı bir etkiye sahip olabilir. Bu nedenle bir protesto tahmin sistemi, alternatif seyahat ve nakliye rotalarının planlanmasına yardımcı olabilir (Muthiah et al., 2015). Tüm bunalar, birçok sosyal ve veri bilimcisini, bu olaylarda yer alan kalıpları ortaya koymaya ve gelecekteki gizli protestoların tahminine daha fazla odaklanmaya yönelendirmiştir (Kallus, 2014, Cadena et al., 2015, Korkmaz, 2015).

Washington'daki Georgetown Üniversitesi'nde veri bilimci olan Kalev Leetaru tarihteki büyük veri kalıplarının incelenerek geleceğin tahmin edilebileceğine inanmaktadır. Leetaru, büyük veri tabanlarını sorgulamak için tasarlanan Google BigQuery adlı bir aracı kullanarak, 1979 yılına kadar uzanan bir olay veritabanına ulaşmayı başarmıştır (Cellan-Jones, 2014). Bu olay veritabanı, son 40 yılda 100'den fazla dilde sayısız kaynaktaki olayların medya raporlarını derleyen GDELT (Golbal Database of Events, Language, and Tone- Küresel Olay Veri Tabanı, Dil ve Ton)'dir (Leetaru, 2014). Leetaru, GDELT'te toplanan çeyrek milyar veri kaydını Google'ın altyapısında bulunan BigQuery'i kullanarak birkaç saniyede analiz edebilmektedir. Leetaru, yaptığı bazı analizlerde yıllar içerisinde kendisini tekrarlayan karmaşık olay kalıplarına ulaşmıştır. Leetaru, Mısır, Ukrayna ve Lübnan da yaşanan son olaylara bakarak bu olaylara benzeyen kalıpları bulmaya çalışmıştır. Modern medyanın kısa dikkat süresi göz önüne alındığında, bu yöntem tarihi inceleyen ilkel ve güvenilmez bir yol gibi gözükse de, Leetaru, bunun altında yatan modellerin tahmin edebileceğimizden daha karmaşık olduğunu söylemektedir.

Leetaru, sözkonusu kalıplara ilişkin şunları söylemektedir: "Örneğin; Mısır'da yaşanan olayların son iki ayını alıyorsanız ve geçmişte en benzer dönemler için her ülkenin geçmişini araştırıyorsanız, o zaman tüm bu dönemlerden sonra neler olduğuna bakarsanız. Bu size daha sonra Mısır da ne olacağı konusunda oldukça iyi bir tahmin verir." Leetaru, tarihçiler için bu tür bir hesaplama aracının, mesleki uzmanlıklarına yönelik bir tehdit olmadığı ve tarihçilerin bunu başka bir teknik olarak görmeleri gerektiğini söylemektedir.

Şimdiye kadar dünyadaki huzursuzlukları öngörmede GDELT'i kullanan az sayıda çalışma vardır. Alikhani (2014), Qiao ve Wang (2015) ve Keneshloo ve arkadaşları (2014) çalışmalarında GDELT'i kullanarak regresyon yöntemi ile öngörü yapmışlardır. Yonamine (2013) çalışmasında, GDELT'i kullanarak zaman serisi tahmini yapmıştır. Qiao ve Chen (2016) Tayland'ın istikrarsızlığı ile ilgili göstergeleri tahmin etmek için Gizli Markov Modelleri temelli bir yöntem oluşturmak için GDELT'i kullanmışlardır. Qiao ve Wang (2015) GDELT'i kullanarak ptortesto olaylarını tespit etmede ve öngörmede hesaplamalı bir yaklaşım kullanmışlardır.

Bu makale Leetaru'ın (2014), "Psikotarihte Doğru: Google BigQuery ile Dünya Tarihinin Yapılarını Keşfetme" adlı çalışmasından esinlenerek yapılmıştır. Leetaru (2014) çalışmasında, Mısır devriminin ikinci günü; 27 Ocak 2011 ile Ukrayna Cumhurbaşkanı Viktor Yanukoviç'in ülkesinden kaçtığı; 22 Şubat 2014 tarihleri referans noktası olarak almıştır. Yapılan analizler sonucunda Mısır ve Ukrayna'ın ikinci 60 günlük dönemleri için korelasyon katsayıları sırasıyla; r=0,48 ve r=0.49 tahmin edilmiştir. Elde edilen bu sonuçlarda

şu anda bir ülkede yaşanan benzer tarihi dönemleri sorgulayarak, geçmişte her bir dönem sonrasında yaşananlara gelecekte hedef ülkenin ölçülebilir bir tahmini olarak bakılabileceği sonucuna varılmıştır.

Bu çalışmanın amacı, dünya haber medyasını takip eden GDELT veritabanını kullanarak ülkelerin benzer tarihi kalıplarını yakalamak ve benzer süreçlerden sonraki sürecin ülkelerin geleceğini tahmin etmede kullanılıp kullanılamayacağını araştırmaktır. Bu çerçevede, Tunus'un tarihinde önemli protestoların yaşandığı 25 Temmuz 2013 tarihi referans noktası alınmıştır. Öncelikle, Google BigQuery'nin korelasyon özelliğinden yararlanarak Tunus'un 25 Temmuz 2013'ten önceki 30 gününe en çok benzeyen dünya tarihindeki kalıplar ülkeler bazında araştırılmıştır. Daha sonra, Tunus'un 25 Temmuz 2013'ten önceki 30 gün ile yüksek korelasyonlu çıkan ülkelerin Tunus'un 25 Temmuz 2013'ten sonraki 30 günlük süreçte de (Tunus ile) korelasyonlu olup olmadığı araştırılmıştır. Son olarak geçmişte birbirini takip eden 30 gün içerisinde en çok benzerlik gösteren tarihi dönemlerin ortalaması alınarak Tunus'ta 25 Temmuz 2013'ten sonraki 30 günde yaşanabilecekler için olası bir tahmin olarak kullanılıp kullanılamayacağı araştırılmıştır.

Çalışmanın birinci bölümünde, Google BigQuery'nin yapısı hakkında genel bilgiler verilmiştir. İkinci bölümde, 100'den fazla dilde her ülkenin hemen her köşesinden dünyanın yayın, baskı ve web haberlerini izleyen GDELT veri seti ile Pearson korelasyon katsayısı hakkında genel bilgiler verilmiştir. Üçüncü bölümde, elde edilen bulgular yorumlanarak sunulmuştur. Sonuç bölümde ise, elde edilen sonuçlar sıralanarak, yorumlanmış ve tartışılmıştır.

1.BIGQUERY NEDİR?

Büyük veri kümelerinin saklanması ve sorgulanması, doğru donanım ve altyapı olmadan zaman alıcı ve masraflıdır. BigQuery, Google altyapısının işlem gücünü kullanarak süper hızlı SQL sorguları sağlayarak bu sorunu çözen bir kurumsal veri deposudur (Hoffa, 2017). BigQuery, kullanıcıların, çok terabaytlı veri kümelerine saniyeler içinde hızlı, SQL sorguları yapmalarına imkan sağlamaktadır (Google Cloud Platform Blog, 2014). BigQuery'deki sorgular bir REST API aracılığıyla gönderilir ve gelişmiş sorgular, kullanıcı tanımlı JavaScript işlevleriyle genişletilebilen standart SQL veritabanında yapılmaktadır. Her gün müşteriler tarafından kullanılabilen yüzlerce terabaytlık yeni veri (toplu iş ve akış) sorgulanmak üzere BigQuery'e yüklenmektedir. Binlerce işlemci tek bir sorgulamaya katılabilir ve veri endeksleme veya bölümlemeye gerek kalmadan hızlı sonuçlar elde edilebilmektedir (Leetaru and Hoffa, 2015). BigQuery ölçeklenebilme ve kolay kullanılabilme özelliği sayesinde veriler hakkında gerçek zamanlı bilgiler sunmaktadır. Ayrıca GDELT'in BigQuery'deki kullanılabilirliği sayesinde küresel insan toplumu ve dünya hakkında gerçek zamanlı görüşlere erişmekte mümkündür.

Google'ın BigQuery veritabanı, GDELT gibi veri kümeleri için özel olarak tasarlanmıştır. BigQuery GDELT veri kümesinde gerçek zamana yakın anlık sorgulama yapmaktadır (GDELT, 2019). GDELT'in gerçek zamanlı sorgulamalarından çıkan desenler, önemlerini, uygunluklarını ve altında yatan itici güçleri belirlemek için tarihi arşivler bağlamında önemlidir (Leetaru and Hoffa, 2015). Kırılma olaylarının analizini etkinleştirmek için gerçek zamanlı güncellemelerin hemen kullanılabilir olması gerekir. Bu nedenle, BigQuery'nin akış eklemeleri yoluyla desteklediği hem gerçek zamanlı hem de geçmiş veri depolarında birleşik sorgulamaya izin veren bir ortama ihtiyaç duyulur (Leetaru and Hoffa, 2015). BigQuery'de GDELT'e sahip olmanın en önemli yanı, karmaşık sorgulama, verilerin çıkarılması ve gerçek dünya analizlerinin veritabanında yapılmasına izin vermesidir. Son 40 yılda, dünyadaki en önemli çatışma etkileşimlerini aylara göre hesaplamak veya bir grup ülke arasındaki farklı ilişki sınıfları arasında çapraz sekanslı korelasyonlu sorgular BigQuery'nin içinde çalıştırılabilir ve birkaç saniye içerisinde sonuçlara ulaşılmaktadır.

2.VERİ SETİ

Bu çalışmada dünya haber medyasını takip eden GDELT veri kümesindeki çatışma verileri kullanılmıştır. Bu veriler dünya medyasında ülkelerle ilgili yapılan çatışma haberlerinin frekansıdır. Buradaki çatışma olayları; askeri seferberlikler, yardım ya da diplomatik ilişkilerin durdurulması/azaltılması, ambargo, boykot ve yaptırımlar, protestolar, sokağa çıkma yasağı ve kitlesel gözaltına alınma gibi zorlama ve fiziksel saldırılardan oluşmaktadır.

Google tarafından desteklenen GDELT projesi, dünyanın dört bir yanından yayın, baskı ve çevrimiçi haber kaynaklarından çıkan yüz milyonlarca olay kaydından oluşmaktadır (Van der Aalst et al., 2011). GDELT, yerel olarak ele alınan olaylar hakkında bilgi toplayan 440 milyondan fazla girişe sahiptir. GDELT, insan toplumunun en geniş, en kapsamlı ve en yüksek çözünürlüklü açık kaynaklı veritabanıdır. Bu veritabanı, 1979'dan günümüze kadar ulusal ve uluslararası haber kaynaklarından 100'den fazla dilde çeviri yapmaktadır. GDELT, dünyanın her ülkesinde insani toplumsal boyutta davranış ve inançların bir kataloğunu oluşturarak, gezegendeki her kişiyi, organizasyonu, yeri, sayısı, temayı, haber kaynağını ve olayı, etrafta olup bitenleri

yakalayan tek bir ağa bağlamaktadır. GDELT sayesinde her gün dünya içeriğinin ne olduğu, kimlerin etkileşim içerisinde olduğu ve dünyanın nasıl hissettiğini anlamak mümkündür (Fallahi, 2017).

3.METODOLOJİ

İki değişken arasındaki ilişkinin gücü genellikle r ile gösterilen Pearson Momentler Çarpımı korelasyon katsayısı (Pearson Product Moment correlation coefficient) ile ifade edilmektedir. Pearson korelasyon katsayıları -1.0 ile +1.0 arasında değişmektedir. Eğer korelasyon katsayısı; r=-1.0 ise, ölçülen tüm noktalar negatif eğime sahip bir doğru (çizgi) üzerine düşerek mükemmel bir korelasyonu temsil etmektedir. Eğer r=0.0 ise, değişkenler arasında kesinlikle doğrusal bir ilişki yoktur. Eğer r=+1.0 ise, pozitif bir eğime sahip olan bir doğru (çizgi) üzerindeki noktalar mükemmel bir korelasyonu temsil etmektedir. X ve Y iki değişken olmak üzere; bu iki değişken arasındaki basit doğrusal korelasyon katsayısı (1) denklemindeki formül ile hesaplanılır:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} \tag{1}$$

Denklem (1)'deki SS_{xy}, X ve Y'nin ortalamalardan sapmalarının çarpımları toplamını SS_{xx} ve SS_{yy} ise sırasıyla X'lerin ve Y'lerin ortalamalardan sapmalarının kareler toplamını göstermektedir. SS_{xy}, SS_{xx} ve SS_{yy} (2) denklemindeki gibi hesaplanılır.

$$\begin{aligned} SS_{xy} &= \sum(X-\bar{X})(Y-\bar{Y}) = \sum xy \\ SS_{xx} &= \sum(X-\bar{X})^2 = \sum x^2 \\ SS_{yy} &= \sum(Y-\bar{Y})^2 = \sum y^2 \end{aligned} \tag{2}$$

Denklem (1)'deki r ortalamalar cinsinden elde edilen gözlem değerleri kullanılarak (3) denklemindeki gibi hesaplanılır:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} \tag{3}$$

Hesaplanmış olan basit korelasyon katsayısının anlamlılığını test etmede seçilen anlamlılık düzeyinde (4) denklemindeki gibi test edilir (Gürsakal, 2013: 351).

$$\begin{aligned} H_0&: \rho = 0 \\ H_1&: \rho \neq 0 \end{aligned} \tag{4}$$

Sıfır hipotezi, anakütle korelasyon katsayısının sıfıra eşit olduğunu, alternatif hipotez ise, anakütle korelasyon katsayısının sıfırdan farklı olduğunu belirtmektedir (Gürsakal, 2013: 351).

Pearson korelasyon katsayısının istatistiksel olarak anlamlı olup olmadığı, $t - testi$ ile test edilmektedir. Karar aşamasında bir t_h istatistiği (5) denklemindeki gibi hesaplanır (Köseoğlu ve Yamak, 2001:275).

$$t_h = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{5}$$

Eğer hesaplanan t_h istatistiğin mutlak değeri $(n-2)$ serbestlik derecesi ve seçilen anlamlılık düzeyine karşılık (çift yönlü t tablo) gelen kritik tablo değerinden büyük, yan; $|t_h| > t_{n-2,\alpha}$ ise sıfır hipotezi (H_0) red edilir. Yani anakütle korelasyon katsayısının sıfırdan farklı olduğu sonucuna varılır. Diğer taraftan eğer; $|t_h| < t_{n-2,\alpha}$ ise, anakütle korelasyon katsayısının sıfır olduğu sonucuna varılır (Köseoğlu ve Yamak, 2001:275)

4.BULGULAR

BigQuery'de GDELT'e sahip olmanın en önemli yönlerinden biri, çeyrek milyar verinin gerçek zamanlı sorgulanması ve özetlenmesidir. Ayrıca, BigQuery'nin korelasyon desteği, GDELT'i küresel insan toplumunun temelini oluşturan temel kalıpları aramak için kullanma imkânı sunmaktadır. Bu çalışma da; geleceği tahmin etmek için geçmişteki benzer dönemlere bakıla bilir mi? Sorusuna yanıt aranmıştır. Bu amaçla çalışma da; 25 Temmuz 2013'te Tunus'ta katledilen Halkçı Cephe Milletvekili'nin öldürülmesinden ve hükümetin devrilmesinden sonra ülkede ne olacağını tahmin etmek mümkün müydü? Sorusuna cevap bulmaya çalışılmıştır. Yani, esasında çalışmada, geleceği tahmin etmek için geçmişteki benzer tarihi kalıplara (desenlere) bakılabilir mi? Sorusuna yanıt aranmıştır.

Öncelikle aşağıdaki algoritma ile 25 Temmuz 2013'ten önceki bir ay boyunca Tunus'ta yaşanan huzursuzluk yoğunluğunun bir zaman çizelgesini oluşturmuştur. Daha sonra, dünyadaki her ülkeyi 1979

yılından günümüze kadar 30 günlük bir zaman penceresinde izleyerek Tunus'un son 30 günü ile karşılaştırmıştır. Bu karşılaştırma bir Pearson korelasyonu kullanılarak yapılmıştır. Karşılaştırma şu şekilde yapılmaktadır: Afganistan'dan başlayarak 1 Ocak 1979 – 31 Ocak 1979 arasındaki huzursuzluğun zaman çizelgesi oluşturulmakta ve Tunus'un 26 Haziran 2013 - 25 Temmuz 2013 arasındaki 30 günlük dönem ile karşılaştırılmaktadır. Ardından Afganistan 2 Ocak 1979'dan 1 Şubat 1979'a kadar tekrar Tunus ile karşılaştırılmaktadır. Böylece, Afganistan'ın GDELT'in yakaladığı son 40 yıl boyunca açılan pencerede, Tunus tarihinin son 30 gününe en çok benzeyen dünya tarihindeki geçmiş dönemleri belirlenmektedir. Bu süreç tüm ülkeler için benzer şekilde tekrarlanmaktadır. Son olarak algoritma, geçmişte birbirini takip eden bir ay içinde en çok benzer olan tarihi dönemleri ve ortalamasını almakta ve bunu önümüzdeki bir ay boyunca Tunus'ta yaşanabilecekler için olası bir tahmin olarak sunmaktadır. Aşağıdaki kodtaki ülke, tarih ve periyot değiştirilerek farklı ülkeler için benzer analizler yapılabilir. Algoritmadaki periyot en az 30 en fazla 255 günlük seçilebilmektedir.

```
SELECT
  STRFTIME_UTC_USEC(a.ending_at, "%Y-%m-%d") ending_at1,
  STRFTIME_UTC_USEC(b.ending_at-30*86400000000, "%Y-%m-%d") starting_at2,
  STRFTIME_UTC_USEC(b.ending_at, "%Y-%m-%d") ending_at2,
  a.country, b.country, CORR(a.c, b.c) corr, COUNT(*) c
FROM (
  SELECT country, date+i*86400000000 ending_at, c, i
  FROM [gdelt-bq:sample_views.country_date_matconf_numarts] a
  CROSS JOIN (SELECT i FROM [fh-bigquery:public_dump.numbers_255] WHERE i < 30) b
) b
JOIN (
  SELECT country, date+i*86400000000 ending_at, c, i
  FROM [gdelt-bq:sample_views.country_date_matconf_numarts] a
  CROSS JOIN (SELECT i FROM [fh-bigquery:public_dump.numbers_255] WHERE i < 30) b
  WHERE country='Tunisia'
  AND date+i*86400000000 = PARSE_UTC_USEC('2013-07-25')
) a
ON a.i=b.i
WHERE a.ending_at != b.ending_at
GROUP EACH BY ending_at1, ending_at2, starting_at2, a.country, b.country
HAVING (c = 30 AND ABS(corr) > 0.254)
ORDER BY corr DESC
```

Kaynak: **Hoffa, 2018.**

Bu yaklaşım dünya tarihinin sınırlı ve ölçülebilir geçici kalıplar kümesi içerisinde kendini tekrar ettiğini varsaymaktadır. GDELT'in inanılmaz ölçeği ilk kez böyle bir hipotezi sınama imkanı sunmaktadır. Ancak, hesaplanması gereken çok fazla korelasyon (yaklaşık 2,5 milyon) olduğundan bunu yapmak çok masraflıdır. BigQuery'yi kullanarak birkaç satır SQL kodu ile Tunuslu Halkçı Cephe Milletvekili Muhammed El-Brahimi'nin suikaste kuban gittiği ve sokak protestolarının başladığı, 25 Temmuz 2013'ten sonraki ikinci günde Tunus'a en çok benzeyen dönemleri bulmak için dünya tarihi taranmıştır. Bunun için BigQuery de 25 Temmuz 2013'ten önceki (2013-06-05 – 2013-07-25) 30 günde Tunus'taki çatışma olaylarının günlük haber

frekanslarına bakılmıştır. Burada 30 günlük bir yuvarlanan pencere ve Pearson korelasyonu kullanılarak, dünyadaki her ülkenin son 40 yılındaki 30 günlük periyotlar Tunus'un son 30 günü (2013-06-05 – 2013-07-25) ile karşılaştırılmış ve en yüksek korelasyonulu çıkan 4 periyot (dönem) Tablo 1'deki gibi bulunmuştur.

Tablo 1: Tunus'un 2013-07-25 Öncesi 30 Günlük Dönemi İle En Yüksek Korelasyonlu İlk Dört Ülke

Sıra	İlişkili Ülke	İlişkili Dönem Başlangıç	İlişkili Dönem Bitiş	Gün sayısı	Korelasyon Katsayısı
1	İsveç	2017-03-08	2017-04-07	30	0.993**
2	Guyana	2017-02-05	2017-03-07	30	0.992**
3	Angola	2015-05-10	2015-06-09	30	0.992**
4	Sri Lanka	2019-03-22	2019-04-21	30	0.991**

** Korelasyon katsayısı 0.01 düzeyinde istatistiksel olarak anlamlı.

Analiz sonucunda son 40 yılda, Tunus'un 25 Temmuz 2013'ten önceki 30 gününe (2013-06-05 – 2013-07-25) en çok benzeyen, dünya tarihinin dört döneminin bir listesi Tablo 1'deki gibi bulunmuştur. Analiz sonucunda anlamlı ve pozitif korelasyonlu olan toplam 59.526 dönem bulunmuş ve en yüksek korelasyonlu ilk 4 dönem Tablo 1'deki gibidir.

Tablo 1'deki sonuçlardan BigQuery'nin hesapladığı korelasyon katsayılarının çok yüksek çıkması oldukça dikkat çekicidir. Korelasyon katsayıları dikkate alındığında Tunus'un 2013-07-25 tarihinden önceki 30 günde; İsveç (2017-03-08 – 2017-04-07), Guyana (2017-02-05 – 2017-03-07), Angola (2015-05-10 – 2015-06-09) ve Sri Lanka (2019-03-22 – 2019-04-21) ile çok benzer olayların yaşandığı görülmektedir. Tablo 1'deki korelasyon katsayılarına göre; farklı zaman ve mekanlarda ülkelerin çok benzer süreçleri yaşadıkları görülmektedir.

Tablo 1'deki ülkelerin 30 günlük zaman dilimlerinde hesaplanan korelasyon katsayılarının nedenini daha iyi anlamak için söz konusu ülkelerin bu dönemde yaşadığı olaylara bakmak gerekmektedir. Tablo 1'deki ülkelerin söz konusu dönemlerde yaşadıkları olaylar Tablo 2'de verilmiştir.

Tablo 2: İlişkili Dönemlerde Ülkelerin Yaşadığı Olaylar

Ülke	30 Günlük Dönem	Olay
İsveç	2017-03-08 ile 2017-04-07	İsveç'in başkenti Stockholm'ün merkezinde kamyonla düzenlenen saldırıda 4 kişi öldü, 15 kişi de yaralandı. İsveç Başbakanı Stefan Lofven, saldırının "terör eylemi" olduğunu söyledi.
Guyana	2017-02-05 ile 2017-03-07	Guyana da başlayan grevler nedeniyle hayat hayat felç oldu. Guyana da eczane ve acil servisleri dışında tüm ticari ve kurumsal hayat durdu. Ayrıca kara, hava ve deniz yolları da grevciler tarafından kapatıldı. Grev yapan gruplar, kar maskeleriyle tüm dükkan ve mağazaları gezerek kapatma çağrısında bulundu.
Angola	2015-05-10 ile 2015-06-09	Angola'nın başkenti Luanda'da gözaltına alınan 15 eylemcinin serbest bırakılmasını talep etmek için onlarca protestocu sokaklara döküldü. Luanda'da yapılan gösteriye şiddetle tepki gösterildi. Protestocular, Bağımsızlık Meydanı'na girdiklerinde polisler tarafından bastırıldı.
Sri Lanka	2019-03-22 ile 2019-04-21	Sri Lanka'nın başkenti Colombo, Negombo ve Batticaloa kentlerindeki otel ve kiliselere bombalı saldırı düzenlendi. Saldırılarda 207 kişi hayatını kaybetti, 500'den fazla kişi yaralandı. Ülkede sokağa çıkma yasağı ilan edildi.

Tablo 2'deki dört ülkenin 30 günlük dönemlerinde yaşadıkları olaylar verilmiştir. Tablo 1'deki yüksek korelasyon katsayılarının çıkmasının asıl sebebi Tablo 2'de ülkelerin yaşadığı olaylardır. Tablo 2 dikkate alındığında İsveç ve Sri Lanka'nın Tunus ile yüksek korelasyonlu çıkması bu ülkelerde yaşanan terör

saldırılarından kaynaklanmaktadır. Diğer taraftan Guyana ve Angola'nın Tunus ile yüksek korelasyonlu çıkmasında protestoların etkili olduğu görülmektedir.

Peki, ülkelerin bu 30 günlük zaman dilimlerinde yaşadıkları bu benzerlikler rassal mı? Yoksa devam eden 30 günlük zaman diliminde bu güçlü ilişki devam ediyor mu? Bu sorulara cevap bulmak için Tunus'un 25 Temmuz 2013'ten sonraki 30 gününe (2013-07-25 - 2013-08-24) ve diğer ülkelerin ilişkili dönemin bitişinden sonraki 30 günleri arasındaki korelasyona bakmak gerekmektedir. Bunun için aşağıdaki kodla BigQuery de İsveç, Guyana, Angola ve Sri Lanka için ayrı ayrı analizler yapılarak bu ülkelerin Tunus ile 2013-07-25'den sonraki 30 günlük korelasyonlarına bakılmıştır.

```
SELECT country, date, c AS count
 FROM [gdelt-bq:sample_views.country_date_matconf_numarts] a
 CROSS JOIN (SELECT i FROM [fh-bigquery:public_dump.numbers_255] WHERE i < 30) b
 WHERE country='Sweden'
 AND date+i*86400000000 = PARSE_UTC_USEC('2017-05-07')
"""
```

Kaynak: **Hoffa, 2018.**

Yukarıdaki kodla yapılan analizler sonucunda Tablo 1'deki dört ülkenin ilişkili dönemlerden sonraki 30 gün (Tunus ile 2013-07-25 tarihinden sonraki 30 gün için) için hesaplanan korelasyon katsayıları Tablo 3'teki gibidir.

Tablo 3: Tunus'un 2013-07-25 Sonrası 30 Günlük Dönemi İle Dört Ülkenin Korelasyonu

Sıra	İlişkili Ülke	İlişkili Dönem Başlangıç	İlişkili Dönem Bitiş	Gün sayısı	Korelasyon Katsayısı
1	İsveç	2017-04-08	2017-05-07	30	0.687**
2	Guyana	2017-03-08	2017-04-06	30	0.450*
3	Angola	2015-06-10	2015-07-09	30	0.512**
4	Sri Lanka	2019-04-22	2019-05-21	30	0.671**

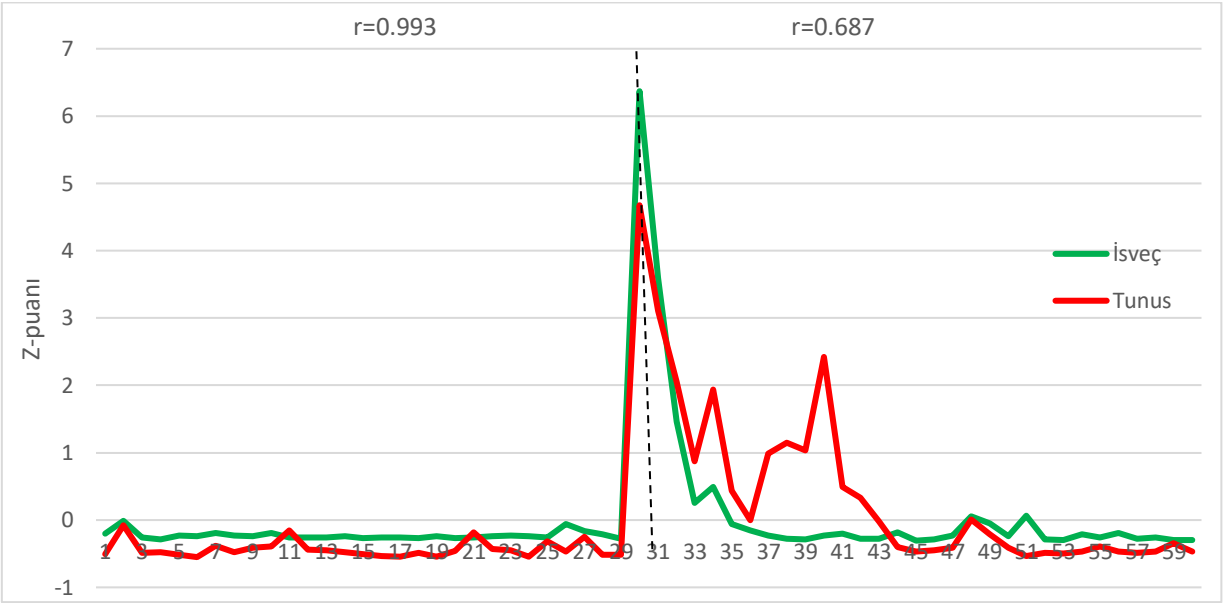
** Korelasyon katsayısı 0.01 düzeyinde istatistiksel olarak anlamlı.

* Korelasyon katsayısı 0.05 düzeyinde istatistiksel olarak anlamlı.

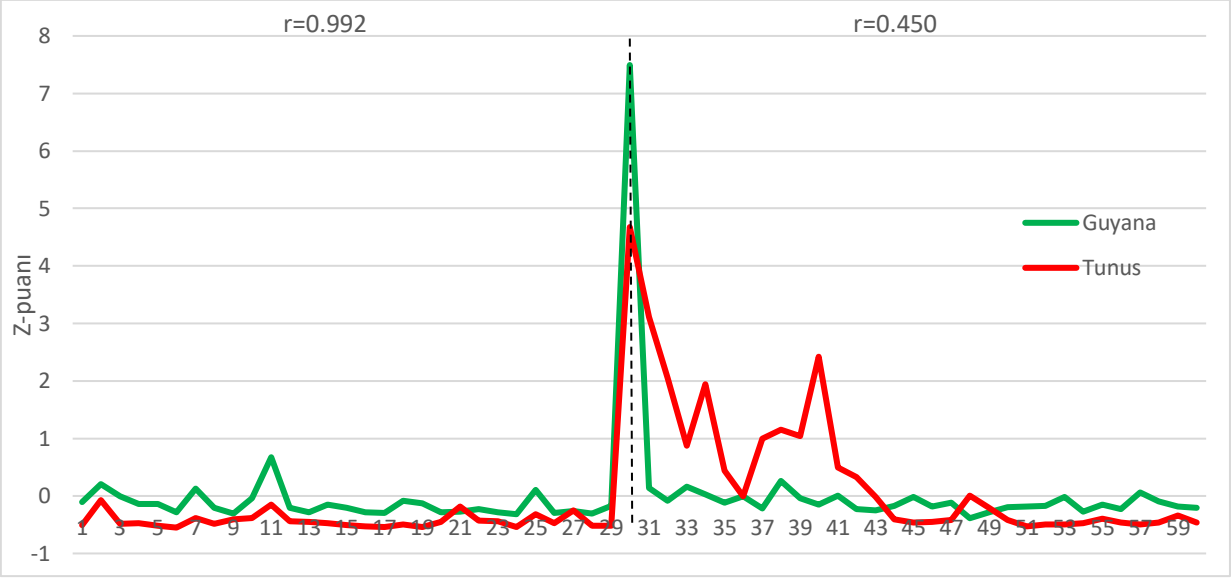
Tunus'un 2013-07-25 sonrasındaki 30 günlük dönemi için hesaplanan korelasyon katsayılarında az da olsa bir düşüş yaşanmıştır. Buna rağmen, İsveç ve Sri Lanka ile yüksek düzeyde korelasyon ilişkisinin hala devam etmesi oldukça dikkat çekicidir. Diğer taraftan Guyana ve Angola ile orta düzeyde bir korelasyon ilişkisinin devam ettiği görülmektedir. Tüm bunlar dikkate alındığında, belirli bir zaman diliminde benzer olayları yaşayan ülkelerin takip eden süreçte de benzer süreçleri yaşama ihtimallerinin oldukça yüksek olduğunu göstermektedir. Bu sonuçlar istatistiksel açıdan oldukça dikkat çekicidir.

Korelasyon katsayısının ilişkiyi yeterince temsil edip etmediğini anlamak için verilerin grafiğine bakmak gerekir. Elde edilen bulguları desteklemek için, ele alınan dört ülke ile Tunus'un (2013-07-25 - 2013-08-24) 60 günlük grafikleri çizilmiştir. Çizgi grafiklerinde birbirini takip eden 30 günlük periyotları ayırmak için kesikli siyah çizgi kullanılmıştır. Çizginin solunda, korelasyon için Bigquery'nin kullandığı 30 günlük periyot bulunurken sağda, bu süreyi izleyen 30 gün içinde, İsveç, Guyana, Angola ve Sri Lanka için, Tunus'ta yaşananlarla karşılaştırılmıştır. Her bir ülke için karşılaştırmaları kolaylaştırmak için ham veri sayıları (frekansları) "Z-puanları" (ortalamadan standart sapmaları) alınarak işlemler yapılmıştır.

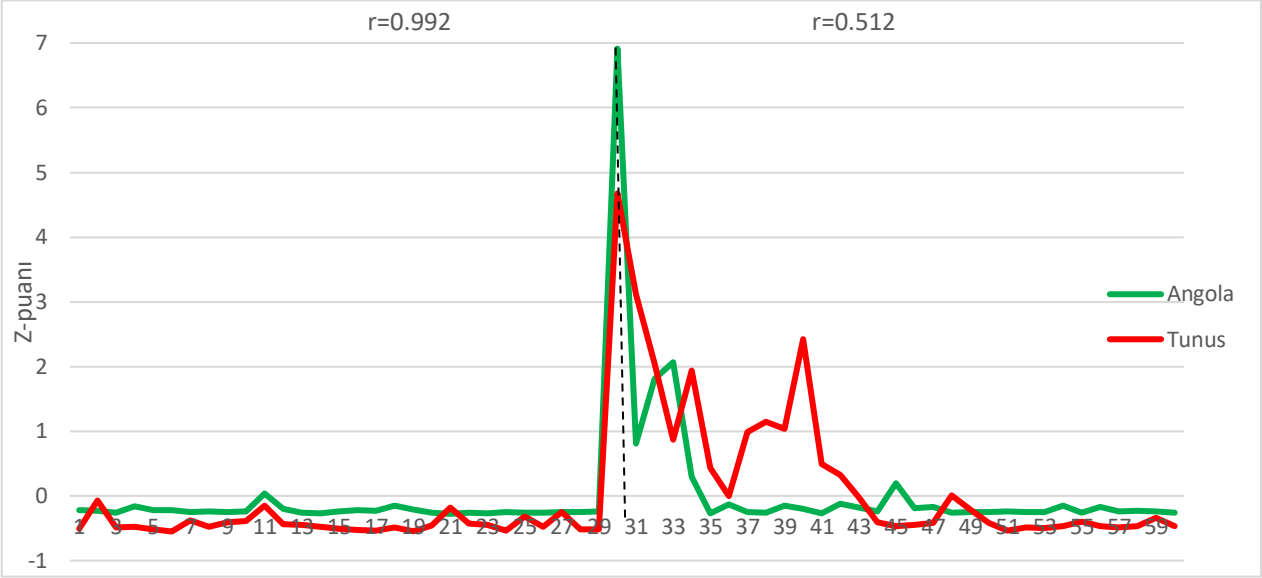
Şekil 1: İsveç 2017-03-08 – 2017-04-07 (siyah çizginin solu) ve 2017-04-08 – 2017-05-07 (siyah çizginin sağı), Tunus (2013-06-26 - 2013-07-25 ve 2013-07-26 – 2013-08-24) ile karşılaştırılması



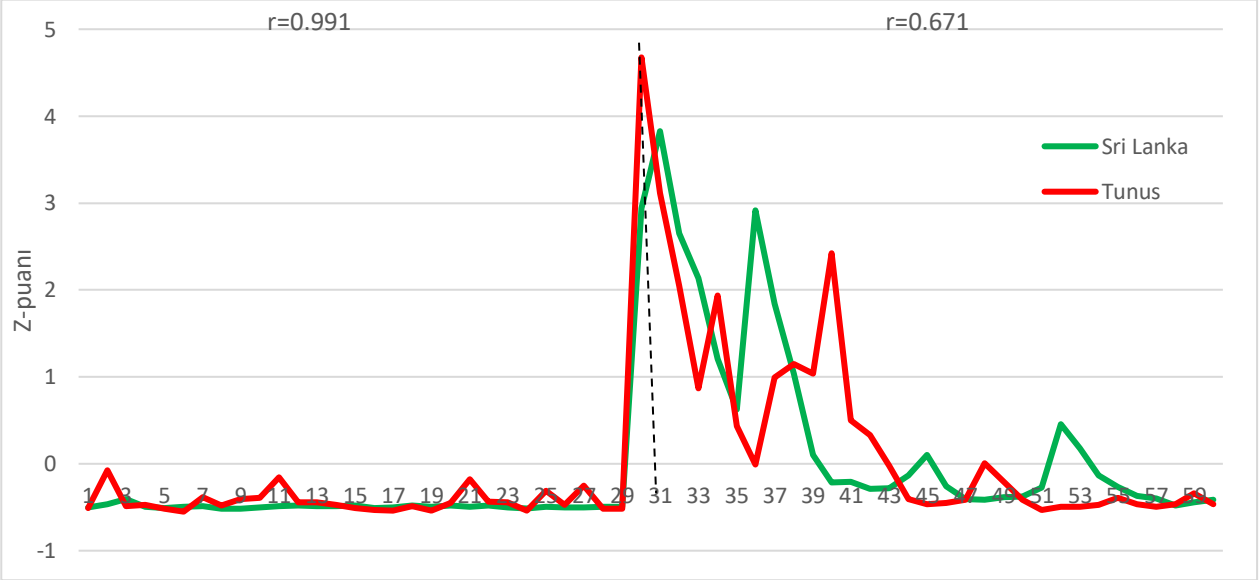
Şekil 2: Guyana 2017-02-05 – 2017-03-07 (siyah çizginin solu) ve 2017-03-08 – 2017-04-06 (siyah çizginin sağı), Tunus (2013-06-26 - 2013-07-25 ve 2013-07-26 – 2013-08-24) ile karşılaştırılması



Şekil 3: Angola 2015-05-10-2015-06-09 (siyah çizginin solu) ve 2015-06-10 – 2015-07-09 (siyah çizginin sağı), Tunus (2013-06-26 – 2013-07-25 ve 2013-07-26 – 2013-08024) ile karşılaştırılması



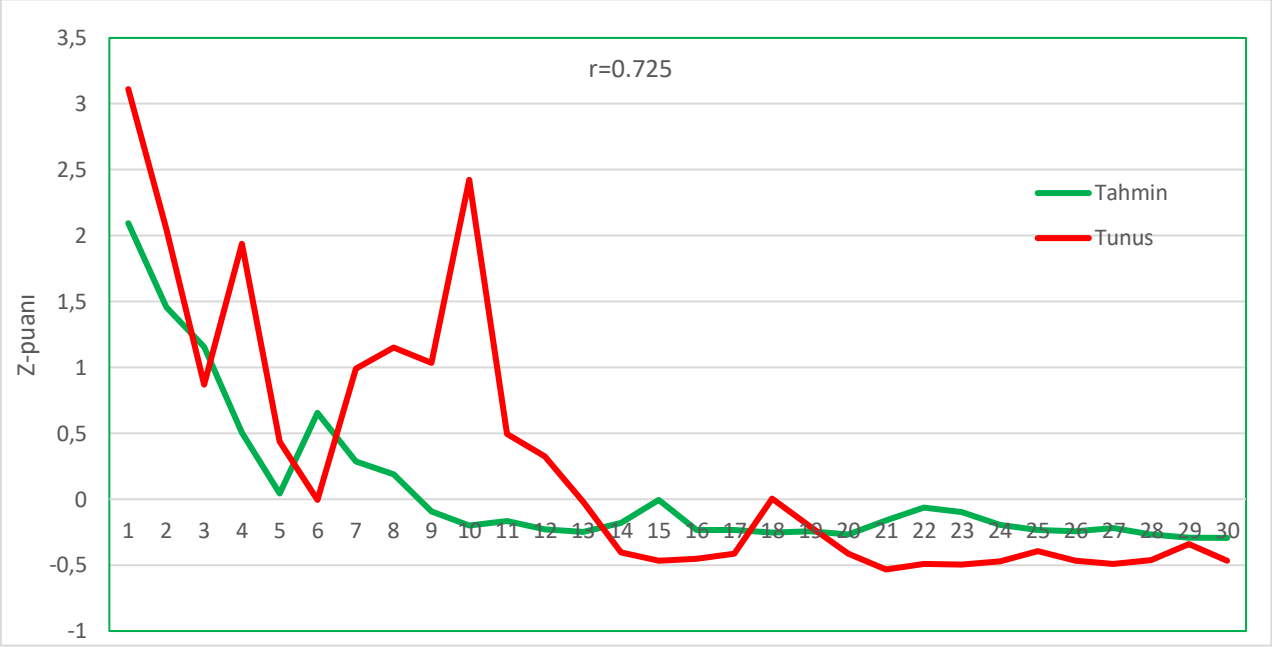
Şekil 4: Sri Lanka 2019-03-22 – 2019-04-21 (siyah çizginin solu) ve 2019-04-22 – 2019-05-21 (siyah çizginin sağı), Tunus (2013-06-26 - 2013-07-25 ve 2013-07-26 – 2013-08-24) ile karşılaştırılması



Yukarıdaki grafiklerde siyah kesikli çizgiler, Tunus'un 2013-07-25 tarihidir. Tunus'taki Halkçı Cephe Milletvekili Muhammed El-Brahimi'nin öldürülmesi ve Mısır'daki darbenin etkisiyle ortaya çıkan yeni bölgesel denklemde Nahda Hareketi başkanlıktan çekilmiş ve ülkeyi seçimlere taşıyacak yeni bir teknokrat hükümetine iktidarı devrettiği 25 Temmuz 2013'te çatışmalar zirveye ulaşmıştır. Yukarıdaki grafikler korelasyon bulgularını desteklemekte ve ülkeler arasında incelenen bu dönemde güçlü ilişkilerin olduğu görülmektedir. Burada Tunus'un ikinci 30 günlük periyotta (2013-07-26 – 2013-08-24) korelasyon katsayısında azalma yaşanmasına rağmen ülkeler arasında hala güçlü bir ilişkinin olduğu görülmektedir.

Son olarak, bu dört ülkenin ikinci 30 günlük ortalamaları alınırsa bu önümüzdeki bir ay boyunca Tunus'ta (2013-07-25 – 2013-08-24) yaşanabilecekler için olası bir tahmini verecektir. Elde edilen tahmin sonuçlarının Tunus ile karşılaştırılması aşağıdaki Şekil 5'te verilmiştir.

Şekil 5: İlişkili zaman dilimlerinden sonraki 30 gün içinde Tunus (kırmızı) ile dört ülkenin ortalama 30 günü (yeşil) arasındaki karşılaştırma



Şekil 5'te Tunus'ta gerçekte yaşananlarla (kırmızı), tahmin edilenler (yeşil) karşılaştırılmıştır. Şekildeki sonuçlar benzer olup, r=0.725 gibi yüksek bir korelasyon katsayısı bulunmuştur. Bu nedenle, tarihi 30 günlük pencerelerin her birinin ardında olanlar ile Tunus'ta olanlar arasında güçlü bir benzerlik ilişkisi bulunmuştur.

Elde edilen tüm bu bulgular; büyük veri çağında gelişen algoritmalar sayesinde ülkelerde yaşanan önemli çatışma olaylarından sonraki süreçte ülkelerde neler yaşanabileceği konusunda önemli tahminlerde bulunmanın mümkün olduğunu göstermektedir. Kısacası; bugün artık gelişen teknoloji ve algoritmalar sayesinde geçmiş tarihi kullanarak geleceği tahmin etmek mümkündür.

SONUÇ

Bu çalışmanın şaşırtıcı yanı herhangi bir uzmanlık gerektirmeden tek bir SQL satırı ile büyük karmaşık veri kümelerinin basitçe analiz edilmesidir. BigQuery sayesinde tüm karmaşık paralellik ve veri yönetimi basit bir şekilde ele alınmaktadır. Küresel toplumun altında yatan modelleri ortaya çıkarmak için dakikalar içerisinde milyonlarca korelasyon tek bir SQL satırı ile analiz edilmiştir. BigQuery'nin benzeri görülmemiş hız ve sadeliği, evrenin en karmaşık ve temel sorularını bile gerçek zamanlı olarak keşfetmeyi mümkün kılmaktadır. Bununla birlikte, haber medyası, çevremizdeki küresel olayları ve dikkat çekici güçlü korelasyonları nasıl anladığımızı temel olarak şekillendirmektedir. Tunus'un karmaşık iç savaş döneminde bile, dünya haberlerinin objektifi ile görüldüğü gibi, küresel toplumun tarihsel döngülerine büyüleyici bir bakış açısı sunmaktadır.

Bu çalışmada, geçmişe bakmanın bugünün geleceği hakkında öngörülerde bulunup bulunamayacağına ilişkin hipotezi test edilmeye çalışılmıştır. Bu amaçla çalışmada öncelikle, Tunus'un 25 Temmuz 2013 tarihi referans alınarak bu tarihten sonraki 30 günlük süreçte yaşanılabilecek çatışmalar tahmin edilmeye çalışılmıştır. Bu çerçevede, BigQuery'nin korelasyon özelliği kullanılarak Tunus'un 25 Temmuz 2013'ten önceki 30 günü ile pozitif korelasyonlu toplam 59.526 dönem bulunmuştur. Elde edilen bu dönemlerden en yüksek korelasyonlu ilk dört dönem ve ülke Tablo 1'de verilmiştir. Tablo 1'deki dönemlerle Tunus'un 25 Temmuz 2013'ten önceki 30 günü arasındaki korelasyon katsayısı 1'e çok yakın (r=0.99) çıkmıştır. Tablo 1'deki dört ülkenin sonraki 30 günü ile Tunus'un 25 Temmuz 2013'ten sonraki 30 günü karşılaştırıldığında yüksek korelasyon katsayıları bulunmuş ve sonuçlar Tablo 3'te verilmiştir. Son olarak, Tablo 3'teki dört ülkenin ikinci 30 günlük ortalamaları alınarak Tunus'un 25 Temmuz 2013'ten sonraki 30 gününde yaşanılabilecekler için olası bir tahmin yapılmıştır. Elde edilen tahmin sonuçları, Tunus'ta yaşananlar ile güçlü bir (r=0.725) benzerlik ilişkisi göstermektedir. Elde edilen bu sonuç, geçmişe bakmanın bugünün geleceği hakkında öngörüler de bulunulabileceğini göstermektedir.

Bu çalışma, gerek hükümetlerin, gerekse devletlerin gelecekleri konusunda politik kararlar almada ve güvenlik stratejileri geliştirmede yol gösterici niteliği taşımaktadır. Özellikle ülkelerin gelecekleri hakkında öngörülere sahip olmaları büyük önem arz etmektedir. Tabi ki, bu kadar basit bir yaklaşım üzerine gerçek bir politik risk tahmin sistemi inşa etmek mümkün olmasa da, en azından haber medyası aracılığıyla gördüğümüz görüşün dünya tarihinin döngüsel ve öngörülebilir olduğunu ve keşfedilecek daha çok şey olduğu söylenebilir. Bu tarihi kalıplar, tüm insan yaşamını düzenleyen gerçek psiko-tarihsel denklemleri yakalamakta ve gazeteciliğin küresel olaylar hakkındaki anlayışımızı nasıl şekillendirdiğinin matematiksel bir tanımı olabilmektedir. Bu olaylar, gizil desenleri ortaya çıkarmak için Google BigQuery gibi yeni nesil "büyük veri" araçlarının emsalsiz gücünü göstermektedir. Elde edilen model her ülke ve süre boyunca geçerli olacak mı ve daha iyi veya daha kötü çalışan belirli bir yuvarlanma penceresi boyutu var mı? Farklı bir zaman aralığı veya farklı bir olay türü grubuna geçmek doğruluğu arttırır veya azaltır mı? Huzursuzluklardan hemen önce veya huzursuzluğun sadece ilk birkaç gününde daha iyi çalışıyor mu? Gibi soruların yanıtını bulmak için BigQuery de daha değişik çalışmalar yapmak gerekmektedir.

KAYNAKÇA

Alikhani, E., (2014), "Computational Social Analysis: Social Unrest Prediction Using Textual Analysis of News", State University of New York at Binghamton.

Cadena, J., Korkmaz, G., Kuhlman, C. J., Marathe, A., Ramakrishnan, N., & Vullikanti, A., (2015), "Forecasting social unrest using activity cascades". PloS one, 10(6), e0128879.

Cellan-Jones, R., (2014), "Can computers replace historians?", https://www.bbc.com/news/technology-28895098, Erişim Tarihi: 10.07.2019.

Fallahi, F., (2017), "Machine Learning on Big Data for Stock Market Prediction", Master of Science Thesis, *Southern Illinois University Carbondale*.

GDELT, (2019), "The GDELT Project", https://www.gdeltproject.org/data.html#googlebigquery, Erişim Tarihi: 18.01.2019.

Google Cloud Platform Blog, (2014), "World's largest event dataset now publicly available in BigQuery", https://cloudplatform.googleblog.com/2014/05/worlds-largest-event-dataset-now-publicly-available-in-google-bigquery.html, Erişim Tarihi 15.03.2018.

Gürsakal, N., (2013), "Çıkarımsal istatistik: MINITAB-SPSS uygulamalı", Dora Yayıncılık.

Hoffa, F., (2017), "What is BigQuery", https://www.quora.com/What-is-BigQuery, Erişim Tarihi 12.12.2018.

Hoffa, F., (2018), "GDELT correlations.jpynb", https://nbviewer.jupyter.org/github/fhoffa/notebooks/blob/master/GDELT%20correlations.ipynb, Erişim Tarihi: 10.10.2018.

Kallus, N., (2014), "Predicting crowd behavior with big public data", In Proceedings of the 23rd International Conference on World Wide Web (pp. 625-630). ACM.

Keneshloo, Y., Cadena, J., Korkmaz, G., & Ramakrishnan, N., (2014), "Detecting and forecasting domestic political crises: A graph-based approach", In Proceedings of the 2014 ACM conference on Web science (pp. 192-196). ACM.

Korkmaz, G., Cadena, J., Kuhlman, C. J., Marathe, A., Vullikanti, A., & Ramakrishnan, N., (2015), "Combining heterogeneous data sources for civil unrest forecasting", In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 (pp. 258-265). ACM.

Köseoğlu, M., & Yamak, R., (2001), "Uygulamalı İstatistik ve Ekonometri", Trabzon: Celepler. ss.275.

Leetaru, K., (2014), "Towards Psychohistory: Uncovering the Patterns of World History with Google BigQuery" , https://blog.gdeltproject.org/towards-psychohistory-uncovering-the-patterns-of-world-history-with-google-bigquery/, Erişim Tarihi: 10.02.2019.

Leetaru, K., Hoffa, F., (2015), "Analyzing the world's news: Exploring the GDELT Project through Google BigQuery", https://www.oreilly.com/ideas/analyzing-tworlds_news_exploring_the_gdelt_project_through_google_bigquery, Erişim Tarihi: 16.11.2018.

Muthiah, S., Huang, B., Arredondo, J., Mares, D., Getoor, L., Katz, G., & Ramakrishnan, N., (2015), "Planned protest modeling in news and social media", In Twenty-Seventh IAAI Conference.

Qiao, F., & Wang, H., (2015), "Computational approach to detecting and predicting occupy protest events", In 2015 International Conference on Identification, Information, and Knowledge in the Internet of Things (IIKI) (pp. 94-97). IEEE.

Qiao, F., & Chen, K., (2016), "Predicting protest events with Hidden Markov models", In 2016 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC) (pp. 109-114). IEEE.

Van der Aalst, W. M., Schonenberg, M. H., & Song, M., (2011), "Time prediction based on process mining", Information systems, 36(2), 450-475.

Yonamine, J. E., (2013), "Predicting future levels of violence in afghanistan districts using gdelt", Unpublished manuscript.