**Türkiye’de COVID-19 : Günlük ve Kümülatif Verilerle Zaman Serisi Analizleri**

Emine Kıskanç

Harran Üniversitesi,Bilgisayar Mühendisliği

**Özet**

2019 yılı sonunda Çin’in Wuhan kentinde ortaya çıkan COVID-19, 10 Nisan tarihinden itibaren Dünya Sağlık Örgütü(DSÖ) tarafından Küresel Salgın (Pandemi) olarak değerlendirilmiştir.Bu çalışmanın yapıldığı sırada bir tedavi veya aşı bulunamamıştır.Bu nedenle tüm süreç içerisinde en önemli olan adım, hastalığın bulaşma riskini azaltacak hamleler yapmaktır. Bu çalışmada ortaya konacak sonuçlar gelecekte atılacak adımların belirlenmesinde yardımcı olacaktır.Türkiye Cumhuriyeti Sağlık Bakanlığı , COVID-19 verilerini günlük olarak paylaşıyor.Bu çalışmada Türkiye’de görülen ilk 61 günlük veriler kullanılarak zaman serisi analizleri yapılmıştır.Araç olarak Facebook’un geliştirdiği Prophet tahmin prosedürü kullanılmıştır.

**Anahtar Kavramlar:**büyüme eğrileri,COVID-19, hastalık seyrinin tahmini, Prophet,zaman serisi analizi

**Giriş**

2019 Aralık ayında Çin’in Wuhan kentinde ortaya çıkan yeni tip Coronavirüs(COVID-19), tüm dünya genelinde 204 ülkeye yayılmış bir küresel salgın durumundadır.Kullanılan verisetinin son günü olan 5 Mayıs 2020 tarihine kadar tüm dünyada onaylanan vaka sayısı 3.525.081’dir [1] .Bulaşıcı bir hastalık olduğundan dolayı vaka sayısı hızla artmıştır.Bu nedenle kısa zamanda dünyanın bir numaralı sorunu haline gelmiştir.

Türkiye’de ilk vaka 10 Mart tarihinde görülmüştür.İlk vakanın görülmesinden itibaren eğitim,sağlık, sosyal vb. alanlarda çeşitli önlemler alınmıştır.Bunlardan bazıları şöyledir: tüm okul ve üniversitelerde eğitime ara verilmiştir,uçuşlar durdurulmuştur,şehiriçi ve şehirlerarası ulaşıma çeşitli kısıtlamalar getirilmiştir,tüm sosyal alanlar kapatılmıştır,”Sosyal Mesafe” kavramı geliştirilmiştir,haftasonları dışarıya çıkma kısıtlaması getirilmiştir,65 yaş üstü ve 20 yaş altında olanlara sokağa çıkma kısıtlaması getirilmiştir. Sağlık Bakanlığı tarafından COVID-19 ile ilgili veriler günlük olarak paylaşılmıştır[2].Yapılan çalışmada bu verilerden faydalanılmıştır.Çalışmanın amacı, alınan önlemleri değerlendirmek ve gelecekte atılacak olan adımlara dair bir fikir oluşturabilmektir.

**Literatür Taraması**

Koronavirüs, henüz tedavisi olmayan bir salgın türü olduğundan dolayı tedavi ve aşı araştırmalarının yanısıra alınan tedbirlerin analizi ve hastalık seyrinin modellenmesi gibi çalışmalarda yapılmaktadır.Özellikle eldeki imkanların en verimli şekilde kullanılabilmesi adına hastalık seyrinin global veya lokal olarak modellenmesi önem arz etmektedir.[3]’ deki çalışmada COVID-19’un epidemiyolojik analizi ve hastalık seyrinin modellemesi gerçekleştirilmektedir.Çin, İtalya,İran,İspanya gibi virüsten en fazla etkilenen ve virüsün yayılmasında etkili olan 9 ülkenin verileri modellenmiştir.Bu model 9 ülkenin günlük verileri kullanılarak oluşturulmuştur.Aynı zamanda Türkiye’ninde ilk 10 günlük verileri kullanılmıştır.Çalışmada model başarısının genel olarak %95 üzerinde olduğu belirtilmiştir. Ülkenin sağlık politikası, ön hazırlık tedbirleri, yaş ortalaması, ekonomik düzeyi, şehirleşme oranı gibi etkenler başta olmak üzere sosyo-kültürel yapısı ve teknolojik gelişmişliği sonuçlarda farklılık oluşmasında etkili olmuştur.

[4]’deki çalışmada ise 22 ülkenin verilerinden yararlanılarak çeşitli analizler yapılmıştır.Çalışmanın bir diğer aşamasında Türkiye’deki seyir durumunu tahmin etmek için büyüme eğrileri ve zaman serisi analizinden yararlanılmıştır.Türkiye için ilk bir aylık veriler kullanılmıştır.Yapılan tahmninler sonucunda toplam vaka sayısı ve ölüm sayısının tahmin edilmesinde Üstel düzleştirme (Box-Cox) yönteminin kullanılması kararlaştırılmıştır.

Dünya ve Türkiye genelinde artan vaka sayıları ve ölüm oranları eldeki imkanların verimli kullanılması problemini beraberinde getirmiştir. Bu amaçla [5]’de yapılan çalışmada salgın sürecinde ihtiyaç duyulacak yoğun bakım yatak ve solunum cihazı sayılarının direkt tahmini amaçlanmıştır.İki farklı model kullanılmıştır. Veriler,Gompertz ve Zaman serileri modelleri ile incelenmiştir.Modellerin tahmin başarıları benzer bulunmuştur.Modeller,13 günlük veriler kullanıldığı için sadece 1 haftalık öngürülerde bulunabilmişlerdir.Yazar, veri sayısının artış göstermesi halinde modellerin ileriye dönük uzun süreli tahminlerde bulunabileceğini belirtmiştir.

**Kullanılan Araç**

Dünya üzerinde her geçen gün artan veriler sebebiyle veri analizinde hız çok önemli bir faktördür.Bu hızlı analiz gücünü veri bilimcilere vermek adına birçok firma çeşitli prosedürler ve araçlar geliştirmiştir.Bu geliştirilen araçlardan biride bu çalışmada tercih edilen Facebook tarafından açık kaynak olarak geliştirilen Prophet tahmin prosedürüdür.Prophet, hızlıdır ve tamamen otomatik tahminler sağlar. Kullanıcıların tahminleri düzeltmesi ve ayarlayabilmesi için birçok olasılık içerir.Ayarlanabilir parametreleri ile daha verimli sonuçlar alınabilir.Aykırı ve eksik değerlere karşı sağlamdır[6].Prophet aşağıdaki modeli verilere uydurur:

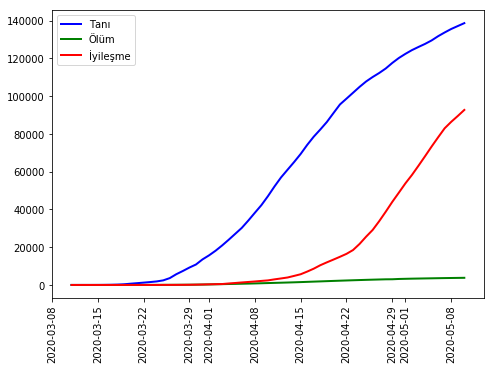
* y(t) tahmin ettiğimiz değişken
* g(t) "büyüme / trend" - yavaş değişen bir işlev
* s(t) "mevsimsellik" - aylık / haftalık modelleri dikkate alacak periyodik fonksiyonlar
* h(t) "tatil" dir - düzenli kalıpları takip etmeyen "özel" günlerin etkisini dikkate alır
* [bağımsız özdeşçe ayrılmış](https://tr.bab.la/sozluk/turkce-ingilizce/ba%C4%9F%C4%B1ms%C4%B1z-%C3%B6zde%C5%9F%C3%A7e-ayr%C4%B1lm%C4%B1%C5%9F) olduğu ve normal olarak dağıtıldığı kabul edilen artıklardır.

Bu işleve [uyum](http://mc-stan.org/) , güçlü bir istatistiksel modelleme paketi olan [STAN](http://mc-stan.org/) tarafından gerçekleştirilir[7].

Veri setindeki verilerin zaman aralığında tatil olmamasından dolayı bu çalışmada h(t)  gözardı edilmiştir.

**Veri Seti**

Veri Seti olarak Türkiye Cumhuriyeti Sağlık Bakanlığı tarafından günlük olarak açıklanan veriler kullanılmıştır[8].Veri setinde ilk günden itibaren 61 günlük kayıt bulunmaktadır.Bu kayıtlar bir önceki günlerin verileriyleriyle toplanarak elde edilmiştir yani kümülatif bir şekildedir.Veri setinde onaylanan vaka sayısı,ölüm sayısı,iyileşme sayısı, şehir, ülke ve tarih özellikleri vardır.Şekil-1’de veri setinin görselleştirilmesi bulunmaktadır.



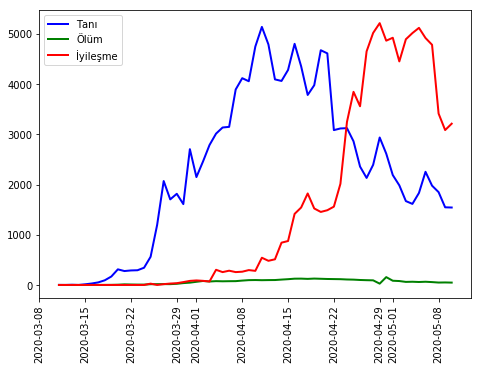
**Şekil-1** Veri setiningörselleştirilmesi

**Veri Önişleme**

Veri önişleme, veri setini standartlaştırmak adına yapılması gereken adımları içerir.Veri önişlemede eksik verileri tamamlama ,hatalı verileri düzeltme,aykırılıkları saptama ve temizleme, tutarsızlıkları giderme,veri dönüştürme, özellik ekleme ve azaltma gibi çeşitli işlemler gerçekleştirilir.

Bu veri setinde kullanılan özelliklerde eksik veri olmadığı için bunları tamamlamak adına herhangi bir işlem gerçekleştirilmemiştir.Veri setinde özellikler kümülatif olarak toplandığı için günlük tanı,ölüm ve iyileşme sayıları yeni özellikler olarak eklenmiştir.Şekil-2’ de özelliklerin günlük değerlerinin eğrisi bulunmaktadır.

Veri setinde şehir ve ülke özelliği kullanılmamıştır.

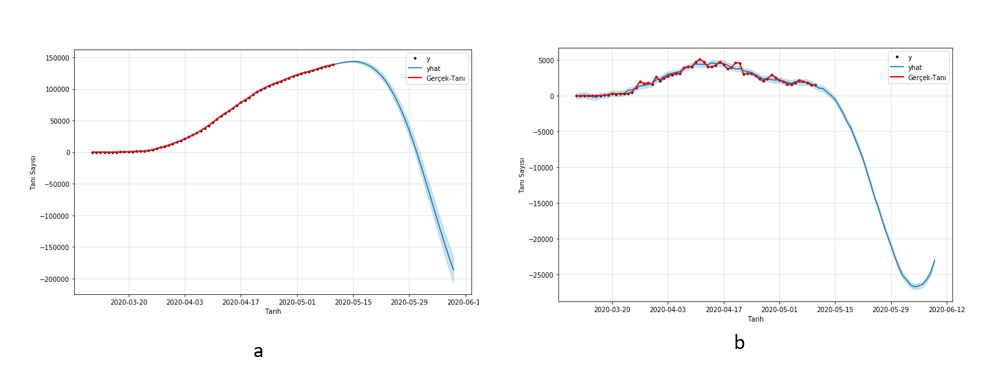


**Şekil-2** Özelliklerin günlük eğrisi

**Deneyler ve Yorumlar**

Veri seti ilk aşamada 30 günlük seriler olarak analiz edilmiştir.İkinci aşamada haftalık ve yıllık analizler olarak detaylandırılmıştır.Ayrıca veriler hem kümülatif değerler üzerinden hem de günlük değerler üerinden bir analize tabi tutulmuştur.Özelliklerin zaman serisi analizleri bağımsızdır.

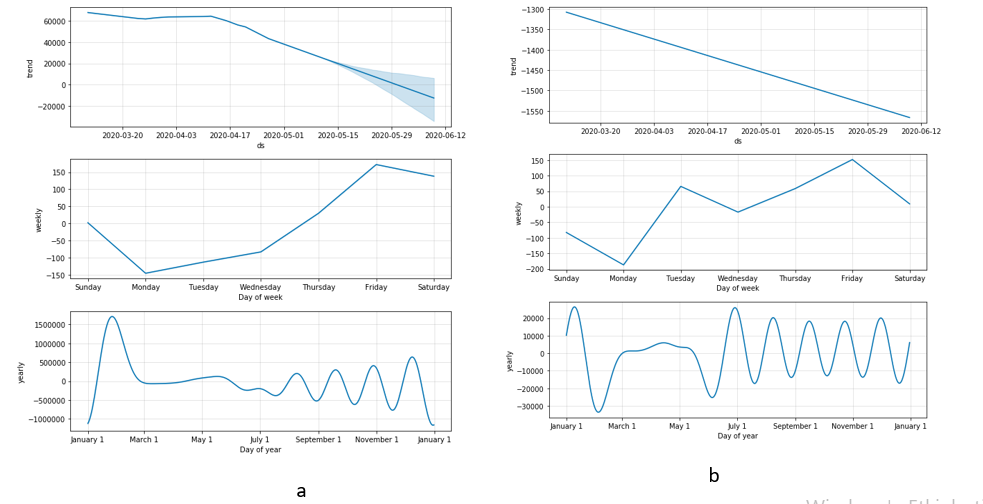
Tanı(vaka) verilerinin 30 günlük kümülatif ve günlük değerler zaman serisi analizi Şekil-3’de verilmiştir.Bu eğrilerde Prophet prosedürünün ön tanımlı özellik isimleri ‘y’ ve ‘yhat’ bulunmaktadır. y,tarihin eğride denk geldiği noktaları belirtmektedir. yhat belirlenen periyottaki tahmini göstermektedir.Kırmızı çizgi ise veri setinin orijinal değerlerine karşılık gelmektedir.



**Şekil-3** a: kümülatif vaka sayıları zaman serisi analizi, b: günlük vaka sayıları zaman serisi analizi

Şekil-3’ de görüldüğü gibi a eğrisinde kümülatif değerler eksponansiyel bir şekilde artmıştır.Devamında ki 30 günlük analizde eksponansiyel bir şekilde azaldığı görülmektedir.Ülkede alınan tedbirler çerçevesinde bu şekilde bir değişikliğin gözlenmesi bekleniyordu. b eğrisinde ise gerçek tanı değerleri belirli bir süre artarken daha sonra azalmaya başlamıştır.30 günlük zaman serisi tahmininde ise vaka sayısında mayıs ayının sonuna doğru ciddi bir azalma gözlemlenmektedir.Haziran başı itibariyle tekrar bir artışın olabileceği gözlemlenmiştir. Haziran başı itibariyle ülkede normal hayata dönüş adımlarının atılacağı bilinmektedir.Bu eğri analizine göre normalleşme sürecine geçişte vaka sayılarının artabileceği gözardı edilmemelidir.

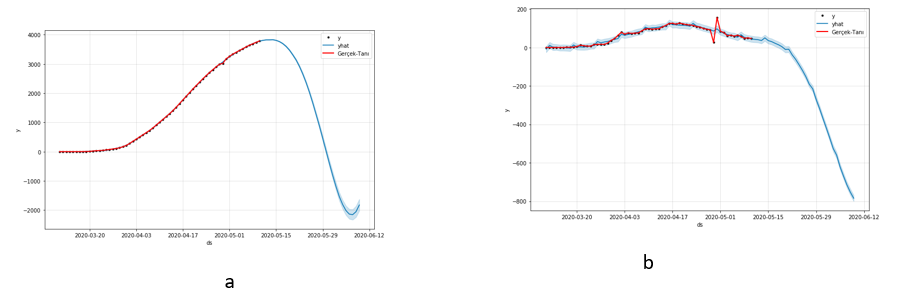
30 günlük analizin 2.aşamasında Şekil-5 de gösterildiği gibi haftalık, yıllık ve trend durumları incelenmiştir.



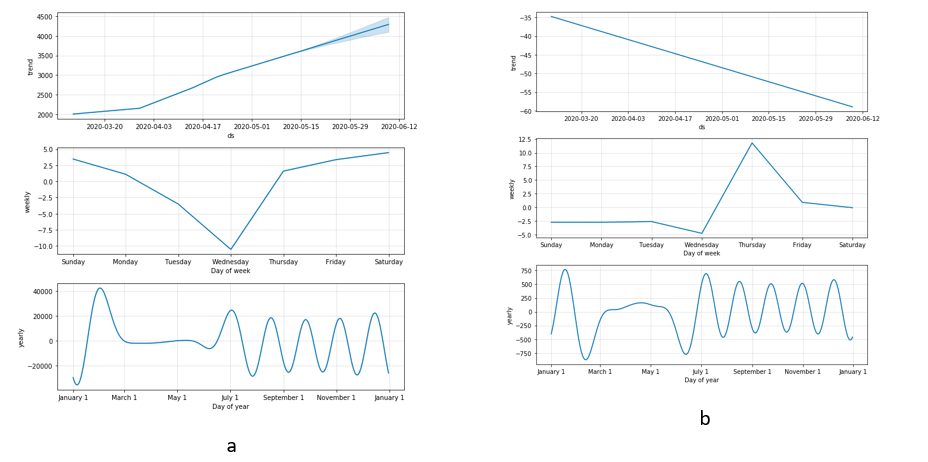
**Şekil-4** a:kümülatif vaka sayılarının yıllık,haftalık,trend analizleri, b: günlük vaka sayılarının yıllık,haftalık,trend analizleri

Şekil-4’de a ve b eğrilerinin trend analizi Şekil-3’de ki analizle uyum göstermektedir.İki analize göre vaka sayılarında azalma beklenmektedir.Şekil-4’de a ve b eğrilerinin haftalık analizlerinde belirli günlerde tutarlılık gözlemlenmektedir.Özellikle Cuma,Cumartesi ve Pazar günlerinde. Cuma gününe kadar vaka sayılarının arttığı gözlemlenmektedir.Cumartesi ve Pazar günlerinde vaka sayıları en düşük değerlerde gözlemlenmektedir. Bu sonucun gözlemlenmesinde ülkede haftasonları uygulanan sokağa çıkma kısıtlamasının etkili olduğu düşünülmektedir. Aynı şekilde a ve b eğrilerinin yıllık analizlerinde de bazı tutarlılıklar gözlemlenmektedir.Mart ve Mayıs ayları arasında vaka sayılarında bir stabilite durumu gözlemlenmektedir.Temmuz ayından Kasım ve Aralık aylarına kadar vaka sayılarında dalgalanmalar gözlemlenebileceği öngörülmektedir.

Ölüm sayılarının 30 günlük kümülatif ve günlük değerler zaman serisi analizi Şekil-5’de verilmiştir.

**Şekil-5** a: kümülatif ölüm sayıları zaman serisi analizi, b: günlük ölüm sayıları zaman serisi analizi

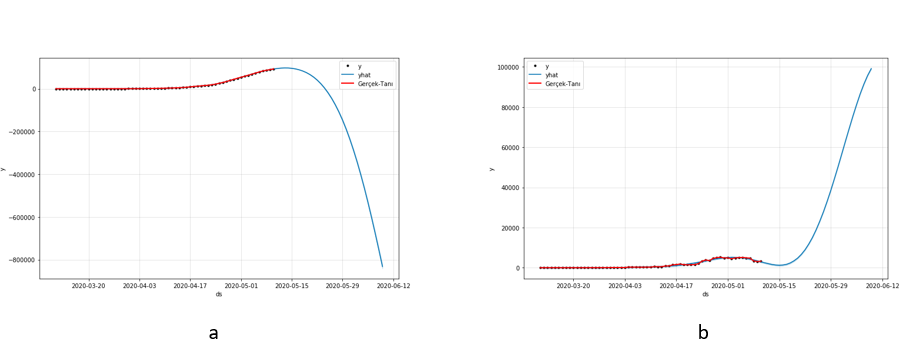
Şekil-5’ de Kümülatif ölüm sayıları veri setindeki değerlere göre eksponansiyel olarak artmıştır.Yapılan zaman serisi analizinde ise haziran ayına kadar eksponansiyel olarak azalması öngörülmektedir.Yapılan 30 günlük zaman serisi analizinde günlük ölüm sayılarıyla bir tutarlılık gözlemlenmektedir.



**Şekil-6** a:kümülatif ölüm sayılarının yıllık,haftalık,trend analizleri, b: günlük vaka sayılarının yıllık,haftalık,trend analizleri

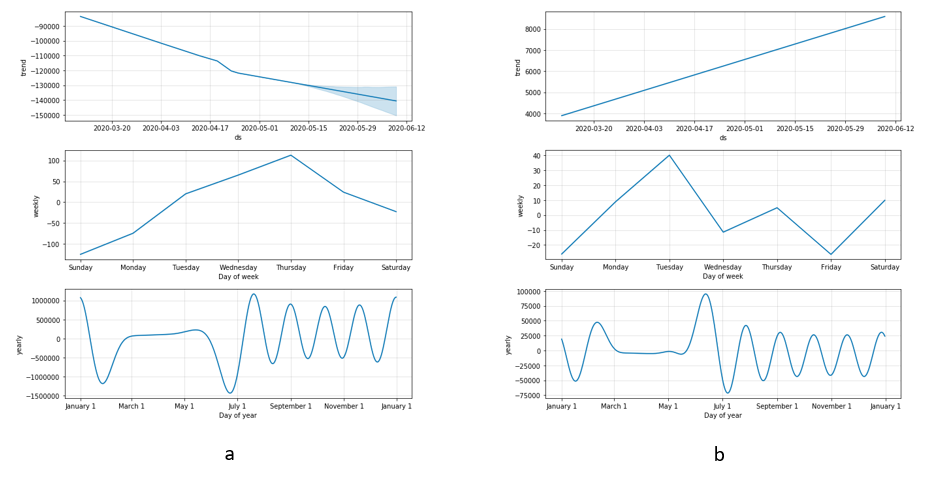
Şekil-6’ da a ve b trend eğrilerinde bir tutarsızlık gözlemlenmektedir.a trend eğrisinde artış gözlenirken b trend eğrisinde azalma görülmektedir.Haftalık ölüm sayılarında ise en tutarlı günün Çarşamba olduğu gözlemlenmektedir.Özellikle haftasonlarında a haftalık eğrisinde ölüm sayıları fazlayken b eğrisinde azalmış olması dikkat çekicidir.Yıllık eğriler ise Şekil-4’deki vaka sayılarının yıllık eğrileriyle tutarlılık göstermektedir.Vaka sayılarının artmasından dolayı ölüm sayılarında da bir artışın olması olası bir durumdur.

İyileşme verilerinin 30 günlük kümülatif ve günlük değerler zaman serisi analizi Şekil-7’de verilmiştir.



**Şekil-7** a: kümülatif iyileşme sayıları zaman serisi analizi, b: günlük iyileşme sayıları zaman serisi analizi

Şekil-7 a eğrisinin elde edilen önceki verilerle bir tutarlılık içinde olduğu gözlemlenebilir. Şekil-3’de vaka sayılarının azalacağı öngörülmüştü bu nedenle vaka sayısının azlığından dolayı iyileşme sayılarında da bir düşüş gözlemlenmesi olasıdır.b eğrisi ise a eğrisiyle bir tutarsızlık göstersede Şekil-5’deki b eğrisiyle bir tutarlılık gösterdiği gözlemlenmektedir.Şekil-5 b eğrisinde ölüm sayılarında düşüş olduğu gözlemlenmektedir.Ölüm sayıları ve vaka sayıları azalınca iyileşme sayılarında bir artışın gerçekleşmesi olası bir durumdur.



**Şekil-8** a:kümülatif iyileşme sayılarının yıllık,haftalık,trend analizleri, b: günlük iyileşme sayılarının yıllık,haftalık,trend analizleri

Şekil-8 a ve b trend analizleri Şekil-7 ‘deki analizlerle tutarlılıklar göstermektedir.a ve b haftalık analizlerinde ise Pazar gününden Salı gününe kadar iyileşme sayılarında artış görülmektedir.a haftalık eğrisinde Salı gününden Perşembe gününe kadar iyileşmelerde devamlı bir artış olduğu halde b eğrisinde aynı günler içerisinde düzenli olmayan davranışlar gözlemlenmiştir.Perşembe-Cuma günlerinde her iki eğri için benzer durumlar gözlemlenmiştir.Cuma-Cumartesi günlerinde ise a eğrisinde iyileşme sayılarında azalma gözlenirken b eğrisinde iyileşme sayılarında artış gözlemlenmiştir.Yıllık eğrilerde ise bir istisna dışında eğriler genel olarak benzer davranışlar sergilemektedir.Bu istisna Haziran ayında a eğrisinde herhangi bir değişiklik olmazken b eğrisinde ciddi bir artışın gözlemlenmesidir.

**Yapılabilecek Çalışmalar**

Bu çalışmada Türkiye’deki 61 günlük vaka,ölüm ve iyileşme sayılarının bulunduğu bir veri seti kullanıldı.Bu veri sayıları arttıkça daha doğru ve tutarlı analizler yapılabilecektir.Özellik sayıları artırılarak il,ilçe bazında analizler yapılabilir.Yapılacak bu analizler eldeki imkanların daha verimli kullanılmasını sağlayabilecektir.Yapılan çalışmada zaman serisi analizi yapılacak periyot 30 gün olarak belirlenmiştir.Bu periyot artırılıp, azaltılarak uzun ve kısa süreli tahminlerde ve analizlerde bulunulabilir.

**Referanslar**

1.WHO, www.who.int

2.Türkiye Cumhuriyeti Sağlık Bakanlığı,<https://covid19.saglik.gov.tr/>

3. ANKARALI, H , ANKARALI, S , ERARSLAN, N . (2020). COVID-19, SARS-CoV2, Enfeksiyonu: Güncel Epidemiyolojik Analiz ve Hastalık Seyrinin Modellemesi. Anatolian Clinic the Journal of Medical Sciences , 25 (Supplement 1) , 1-22 . DOI: 10.21673/anadoluklin.707038

4. ERGÜL, B , ALTİN YAVUZ, A , GÜNDOĞAN AŞIK, E , KALAY, B . (2020). Türkiye’deki COVID-19 Enfeksiyonu: Erken Dönem İstatistikleri ve Hastalık Seyrinin İstatistiksel Olarak Modellenmesi. Anatolian Clinic the Journal of Medical Sciences , 25 (Supplement 1) , 130-141 . DOI: 10.21673/anadoluklin.719629

5. ANKARALI, H . (2020). Türkiye’de COVID-19 Salgın Sürecinde İhtiyaç Duyulacak Yoğun Bakım Yatak ve Solunum Cihazı Sayılarının Direkt Tahmini. Anatolian Clinic the Journal of Medical Sciences , 25 (Supplement 1) , 59-62 . DOI: 10.21673/anadoluklin.715628

6. <https://facebook.github.io/prophet/>

7. <http://www.degeneratestate.org/posts/2017/Jul/24/making-a-prophet/>

8. <https://www.kaggle.com/gkhan496/covid19-in-turkey?select=covid_19_data_tr.csv>