

Hacettepe Üniversitesi

VBM655 Proje Raporu



Ad: Sinan

Soyad: Ballı

Öğrenci No: N20133460

GİRİŞ

Bu projede en küçük kareler algoritması kullanılarak, Türkiye nüfusunun gelecekte ulaşacağı değerleri tahmin eden bir model geliştirilmiştir.

TÜİK internet sitesinden alınan 2016 ile 2019 yıllarına ait ölüm, doğum, içe ve dışa göç ve 2015-2018 yıllarına ait nüfus veri kümeleri kullanılmıştır. Her bir veri kümesi girdi olarak yıl alan doğrusal denklem ile modellenmiştir. Model parametrelerini bulmak için Matlab yazılımının *lsqr* fonksiyonu kullanılmıştır.

Modelleme sonucu yapılan gelecek tahminlerinin TÜİK'in yaptığı tahminler ile tutarlı olduğu gözlemlenmiştir.

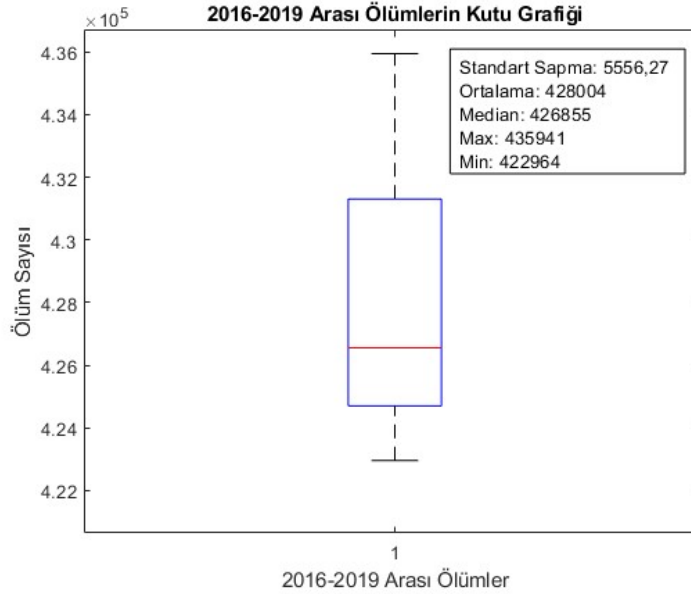
Ayrıca modele geçmiş yıllar girdi olarak verildiğinde gerçek verilere yakın sonuçlar çıktığı gözlemlenmiştir.

Veri Özellikleri

Ölüm Sayısı

Yıl	Ölüm Sayısı
2016	422964
2017	426662
2018	426449
2019	435941

Tablo 1: 2016-2019 yılları arası ölüm sayısı.

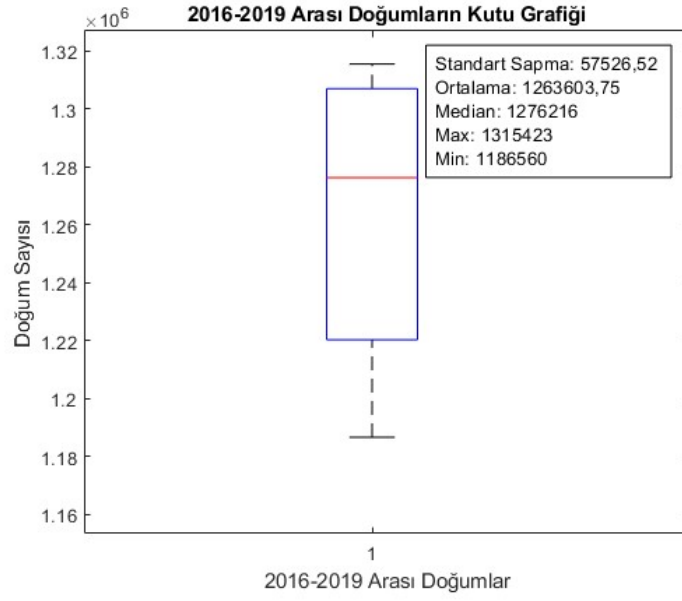


Grafik 1: 2016-2019 yılları arası ölümlerin kutu grafiği.

Doğum Sayısı

Yıl	Doğum Sayısı
2016	1315423
2017	1298451
2018	1253981
2019	1186560

Tablo 2: 2016-2019 yılları arası doğum sayısı.

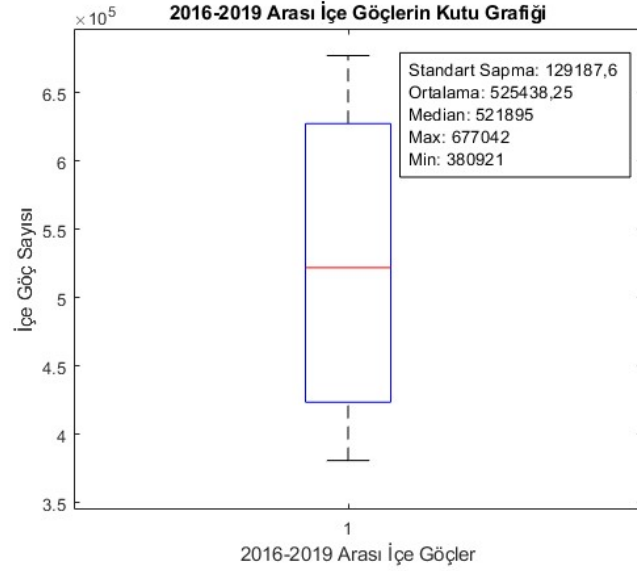


Grafik 2: 2016-2019 yılları arası doğumların kutu grafiği.

İç Göç Sayısı

Yıl	İç Göç Sayısı
2016	380921
2017	466333
2018	577457
2019	677042

Tablo 3: 2016-2019 yılları arası iç göç sayısı.

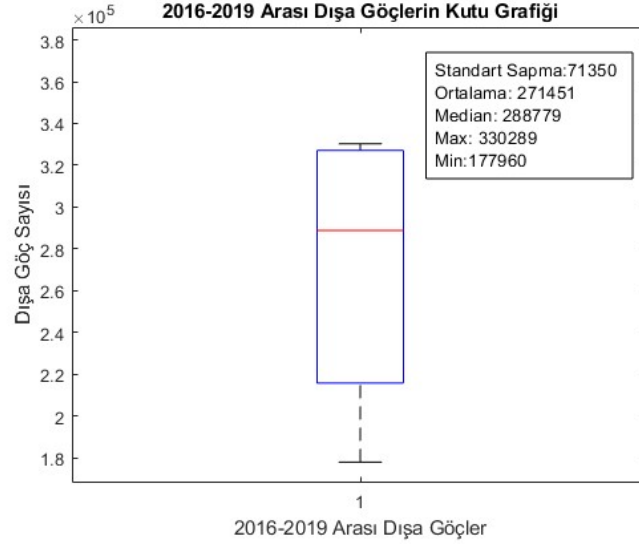


Grafik 3: 2016-2019 yılları arası iç göçlerin kutu grafiği.

Dışa Göç Sayısı

Yıl	Dışa Göç Sayısı
2016	177960
2017	253640
2018	323918
2019	330289

Tablo 4: 2016-2019 yılları arası dışa göç sayısı.

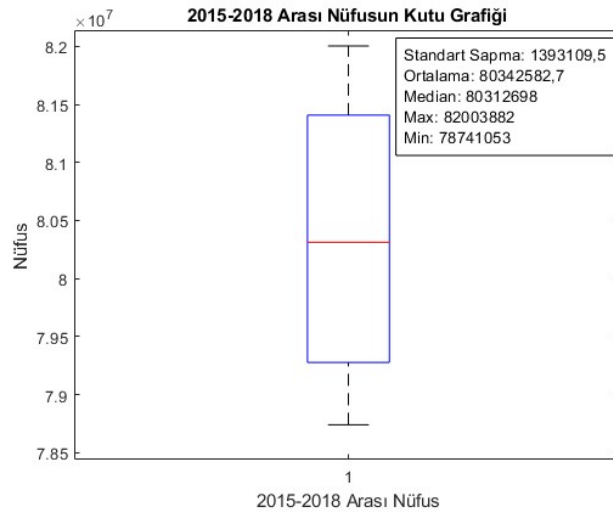


Grafik 4: 2016-2019 yılları arası dışa göçlerin kutu grafiği.

Nüfus Sayısı

Yıl	Nüfus Sayısı
2015	78741053
2016	79814871
2017	80810525
2018	82003882

Tablo 5: 2015-2018 yılları arası nüfus sayısı.



Grafik 5: 2015-2018 yılları arası nüfus sayısının kutu grafiği.

Kullanılan Analitik Metot

Bu proje kapsamında 4 farklı model oluşturulmuştur. Her bir model doğrusal bir denklem olan ve girdi olarak yıl parametresini alan bir fonksiyondur.

Örnek bir model fonksiyonu: $F(yıl) = C1.yıl + C2$ şeklindedir.

Model fonksiyonları doğum oranı, ölüm oranı, dışa göç oranı ve içe göç sayısı olarak seçilmiştir.

Doğum Oranı

$F_{DoğumOranı}(yıl) = C1.yıl + C2$ şeklinde modellenmiştir.

Sonuç olarak parametre olarak verilen yılda gerçekleşen doğum sayısının, bir önceki yılın nüfusuna oranı elde edilir.

Örneğin $F_{DoğumOranı}(2017)$ sonucunun 2016 yılındaki nüfus ile çarpımı 2017 yılında doğan bebek sayısını verir.

Bir yıl içinde doğan bebek sayısı, bir önceki yılın nüfusuyla yakından ilgilidir. Bu yüzden doğan bebek sayısındaki trendi doğru modellemek için sonuç bir önceki yılın nüfusuna göre normalleştirilmiştir.

Yukarıda bahsedilen veri kümesi üzerinden A, x ve b matrisleri aşağıdaki gibi oluşturulmuştur.

$$A = \begin{bmatrix} 2016 & 1 \\ 2017 & 1 \\ 2018 & 1 \\ 2019 & 1 \end{bmatrix}, x = \begin{bmatrix} C1 \\ C2 \end{bmatrix}, b = \begin{bmatrix} 0.01670 \\ 0.01626 \\ 0.01551 \\ 0.01446 \end{bmatrix}$$

Bu modeldeki C1 ve C2 katsayılarını bulmak için, $A.x = b$; denklem kümesi üzerinde EKK algoritması çalıştırılmıştır.

Ölüm Oranı

$F_{ÖlümOranı}(yıl) = C1.yıl + C2$ şeklinde modellenmiştir.

Sonuç olarak parametre olarak verilen yılda gerçekleşen ölüm sayısının, bir önceki yılın nüfusuna oranı elde edilir.

Örneğin $F_{ÖlümOranı}(2017)$ sonucunun 2016 yılındaki nüfus ile çarpımı 2017 yılındaki ölüm sayısını verir.

Bir yıl içinde ölen kişi sayısı, bir önceki yılın nüfusuyla yakından ilgilidir. Bu yüzden ölen kişi sayısındaki trendi doğru modellemek için sonuç bir önceki yılın nüfusuna göre normalleştirilmiştir.

Yukarıda bahsedilen veri kümesi üzerinden A, x ve b matrisleri aşağıdaki gibi oluşturulmuştur.

$$A = \begin{bmatrix} 2016 & 1 \\ 2017 & 1 \\ 2018 & 1 \\ 2019 & 1 \end{bmatrix}, x = \begin{bmatrix} C1 \\ C2 \end{bmatrix}, b = \begin{bmatrix} 0.00537 \\ 0.00534 \\ 0.00527 \\ 0.00531 \end{bmatrix}$$

Bu modeldeki C1 ve C2 katsayılarını bulmak için, $A.x = b$; denklem kümesi üzerinde EKK algoritması çalıştırılmıştır.

Dışa Göç Oranı

$F_{DışaGöçOranı}(yıl) = C1.yıl + C2$ şeklinde modellenmiştir.

Sonuç olarak parametre olarak verilen yılda gerçekleşen dışa göç sayısının, bir önceki yılın nüfusuna oranı elde edilir.

Örneğin $F_{DışaGöçOranı}(2017)$ sonucunun 2016 yılındaki nüfus ile çarpımı 2017 yılındaki dışa göç sayısını verir.

Bir yıl içinde dışa göç eden kişi sayısı, bir önceki yılın nüfusuyla yakından ilgilidir. Bu yüzden dışa göç eden kişi sayısındaki trendi doğru modellemek için sonuç bir önceki yılın nüfusuna göre normalleştirilmiştir.

Yukarıda bahsedilen veri kümesi üzerinden A, x ve b matrisleri aşağıdaki gibi oluşturulmuştur.

$$A = \begin{bmatrix} 2016 & 1 \\ 2017 & 1 \\ 2018 & 1 \\ 2019 & 1 \end{bmatrix}, x = \begin{bmatrix} C1 \\ C2 \end{bmatrix}, b = \begin{bmatrix} 0.00226 \\ 0.00317 \\ 0.00400 \\ 0.00400 \end{bmatrix}$$

Bu modeldeki C1 ve C2 katsayılarını bulmak için, $A.x = b$; denklem kümesi üzerinde EKK algoritması çalıştırılmıştır.

İçe Göç Sayısı

$F_{İçeGöçSayısı}(yıl) = C1.yıl + C2$ şeklinde modellenmiştir.

Sonuç olarak parametre olarak verilen yılda gerçekleşen içe göç sayısı elde edilir.

Örneğin $F_{İçeGöçSayısı}(2017)$ sonucu 2017 yılında gerçekleşen içe göç sayısını verir.

Bir yıl içinde içe göç eden kişi sayısı, bir önceki yılın nüfusuyla fazla ilgili değildir. Bu yüzden normalleştirilme yapılmamıştır.

Yukarıda bahsedilen veri kümesi üzerinden A, x ve b matrisleri aşağıdaki gibi oluşturulmuştur.

$$A = \begin{bmatrix} 2016 & 1 \\ 2017 & 1 \\ 2018 & 1 \\ 2019 & 1 \end{bmatrix}, x = \begin{bmatrix} C1 \\ C2 \end{bmatrix}, b = \begin{bmatrix} 380921 \\ 466333 \\ 577457 \\ 677042 \end{bmatrix}$$

Bu modeldeki C1 ve C2 katsayılarını bulmak için, $A.x = b$; denklem kümesi üzerinde EKK algoritması çalıştırılmıştır.

Kullanılan Yazılım ve Donanım

Proje kapsamında Windows 10 işletim sistemi çalıştıran Hp Probook 450 G7 bilgisayarı ve Matlab 2018b yazılımı kullanılmıştır.

Matlab yazılımının LSQR fonksiyonu, EKK problemini çözmek için kullanılmıştır. Bu fonksiyon $A.x = b$ şeklinde tanımlanan denklemleri çözmeye çalışır eğer çözüm bulamaz ise $\text{norm}(b - A.x)$ yani $\text{norm}(\text{error})$ değerini minimize eden bir çözüm bulur.

Endüstriye Dönük Uygulama Sahaları

Eğitim, sağlık, ulaşım ve istihdam gibi toplumun çoğunu ilgilendiren alanlarda, gelecekte ulaşılabilecek nüfus sayısına göre önceden planlama yapmak önemlidir. Gerçekçi nüfus tahmini yapabilmek, devletlerin bu planlamaları sağlıklı biçimde yapabilmelerine imkân sağlar ve böylece hizmetlerde oluşabilecek aksamaların önüne geçilir.

Yapay Zekâ ile İlişkisi

Model oluşturulurken yapay zekâ kullanılmamıştır. Fakat yapay zekâ kullanılarak daha gerçekçi sonuçlar elde edebilen modeller çıkarılabilir.

Örneğin kişilerin beslenme, tahlil sonuçları ve kalıtsal hastalıklarından yola çıkarak ömür tahmini yapabilen ve kişilerin internetteki ayak izlerini kullanarak kaç çocuk yapacaklarını, hangi ülkeye göç edeceklerini tahmin edebilen yapay zekâ yazılımları sayesinde daha gerçekçi modeller elde edilebilir.

Ayrıca kullanılan veri kümelerinin her yıl güncel verilerle değiştirilmesi modele bir öğrenme yetisi kazandırır ve gerçekliğini artırabilir.

Modelin Dünya Geneline Uygulanabilirliği

Ülkemizin içinde bulunduğu coğrafi ve siyasi koşullar sebebi ile dış ülkelere gelen göç sayısını modellemek zordur. Bu nedenden ötürü dıştan içe olan göç sayısı bir önceki yılın nüfusuna göre normalleştirilmemiştir.

Fakat bu durum diğer ülkeler için doğru olmayabilir ve modelin doğruluğu azalabilir.

Bunu düzeltmek için, içe göç modeli elde edilirken ülkenin göç politikası dikkate alınarak hazırlanan modeller kullanılabilir. Örneğin içe göç limiti olan bir ülkede model o değerin üstünde sonuçlar veremez.

Sonuçlar

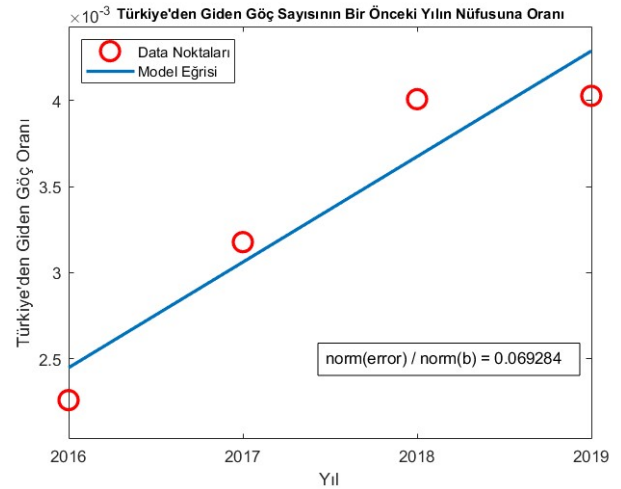
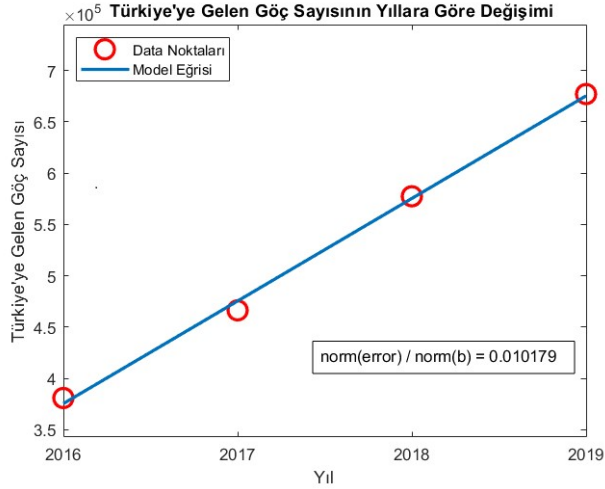
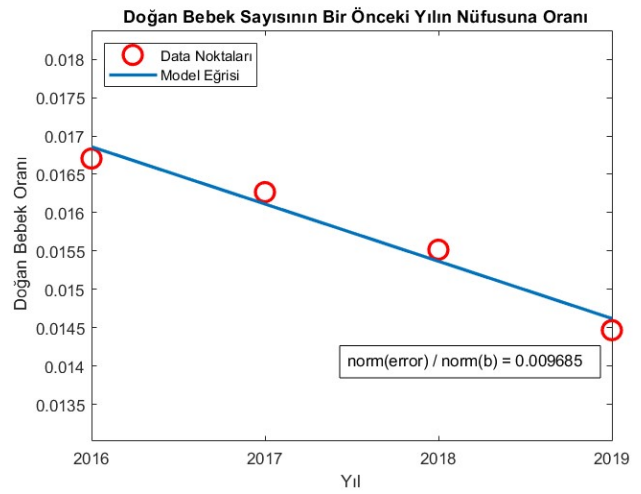
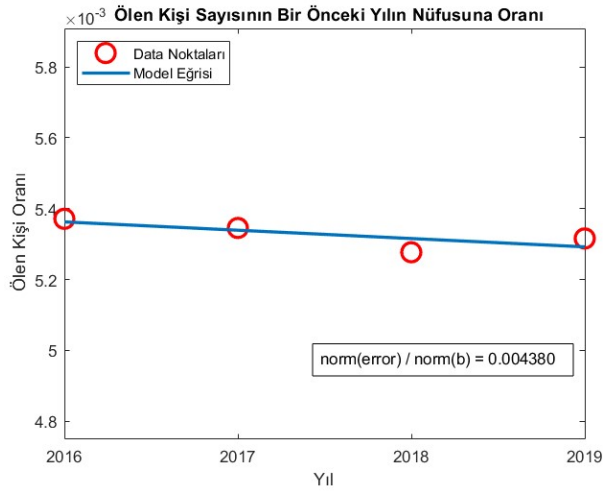
Yukarıda anlatılan analitik metoda göre EKK algoritmasının bulduğu model katsayıları ve model eğrileri aşağıda paylaşılmıştır.

$$F_{\text{ÖlümOranı}}(\text{yıl}) = -2.349378.\text{yıl} + 0.05272$$

$$F_{\text{DoğumOranı}}(\text{yıl}) = -0.000745.\text{yıl} + 1.52061$$

$$F_{\text{DışaGöçOranı}}(\text{yıl}) = 0.0006133.\text{yıl} - 1.23406$$

$$F_{\text{İçeGöçSayısı}}(\text{yıl}) = 99948.699.\text{yıl} - 201121063.99$$



Grafik 6: Model eğrileri ve gerçek veri noktaları.

Model ve TÜİK tahminleri aşağıdaki tablo 6'da paylaşılmıştır. Sonuçlar incelendiğinde model tahminlerinin TÜİK ile tutarlı olduğu gözlemlenmiştir.

Ayrıca tabloda model ve TÜİK tahmini arasındaki farkın TÜİK tahminine göre yüzdesi de paylaşılmıştır. Bu normalleştirilmiş hata yüzdeleri 2048 yılına kadar yüzde 1'den azdır. Daha sonra model ve TÜİK tahminleri arasındaki fark giderek artmıştır ama yüzde 12'yi geçmemiştir.

Yıl	TÜİK Tahmini	Model Tahmini	100.(TÜİK Tahmini - Model Tahmini) / TÜİK Tahmini	Yıl	TÜİK Tahmini	Model Tahmini	100.(TÜİK Tahmini - Model Tahmini) / TÜİK Tahmini
2017	80 810 525	80 905 870	-0.12	2049	104 417 089	102 888 364	1.46
2018	81 867 223	81 997 160	-0.16	2050	104 749 423	102 920 268	1.75
2019	82 886 421	83 085 828	-0.24	2051	105 064 635	102 913 483	2.05
2020	83 900 373	84 168 949	-0.32	2052	105 362 568	102 869 434	2.37
2021	84 908 658	85 243 601	-0.39	2053	105 642 913	102 789 644	2.70
2022	85 911 035	86 306 882	-0.46	2054	105 905 418	102 675 721	3.05
2023	86 907 367	87 355 925	-0.52	2055	106 149 786	102 529 348	3.41
2024	87 885 571	88 387 912	-0.57	2056	106 375 976	102 352 270	3.78
2025	88 844 934	89 400 092	-0.62	2057	106 583 970	102 146 281	4.16
2026	89 784 584	90 389 792	-0.67	2058	106 773 484	101 913 215	4.55
2027	90 703 600	91 354 436	-0.72	2059	106 944 215	101 654 932	4.95
2028	91 601 117	92 291 555	-0.75	2060	107 095 998	101 373 309	5.34
2029	92 476 323	93 198 800	-0.78	2061	107 228 944	101 070 228	5.74
2030	93 328 574	94 073 956	-0.80	2062	107 343 229	100 747 565	6.14
2031	94 153 776	94 914 953	-0.81	2063	107 439 101	100 407 182	6.55
2032	94 951 512	95 719 873	-0.81	2064	107 516 923	100 050 918	6.94
2033	95 721 347	96 486 962	-0.80	2065	107 577 244	99 680 577	7.34
2034	96 463 090	97 214 637	-0.78	2066	107 621 025	99 297 924	7.73
2035	97 176 768	97 901 493	-0.75	2067	107 649 413	98 904 675	8.12
2036	97 862 549	98 546 306	-0.70	2068	107 663 430	98 502 493	8.51
2037	98 520 720	99 148 041	-0.64	2069	107 664 079	98 092 980	8.89
2038	99 151 467	99 705 851	-0.56	2070	107 652 538	97 677 673	9.27
2039	99 754 923	100 219 081	-0.47	2071	107 630 041	97 258 039	9.64
2040	100 331 233	100 687 266	-0.35	2072	107 597 814	96 835 473	10.00
2041	100 882 655	101 110 132	-0.23	2073	107 556 926	96 411 292	10.36
2042	101 409 507	101 487 592	-0.08	2074	107 508 425	95 986 735	10.72
2043	101 911 980	101 819 744	0.09	2075	107 453 417	95 562 962	11.07
2044	102 390 159	102 106 863	0.28	2076	107 392 765	95 141 052	11.41
2045	102 843 989	102 349 399	0.48	2077	107 326 897	94 722 001	11.74
2046	103 273 571	102 547 968	0.70	2078	107 256 148	94 306 726	12.07
2047	103 679 038	102 703 345	0.94	2079	107 180 781	93 896 063	12.39
2048	104 060 257	102 816 455	1.20	2080	107 100 904	93 490 768	12.71

Tablo 6: Model ve TÜİK tahminleri.

Kullanılan Matlab kodları aşağıdaki bağlantıda paylaşılmıştır.

https://drive.google.com/drive/folders/1Jq6ouZC5oYHXHF9ElPy3k_b6bj-Qiasa?usp=sharing