

UOT 004.94

DOI: 10.25045/jpit.v12.i1.08

**Sadiyeva F.İ.**

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan  
[sadiyeva.firengiz@gmail.com](mailto:sadiyeva.firengiz@gmail.com)

## **AZƏRBAYCANDA COVID-19 KORONAVİRUSUNA YOLUXMANIN PROQNOZLAŞDIRILMASI ÜÇÜN ARIMA MODELƏRİNİN TƏTBİQİ**

Daxil olmuşdur: 10.11.2020 Düzəliş olunmuşdur: 13.11.2020 Qəbul olunmuşdur: 20.11.2020

*Məqalədə COVID-19 pandemiyasını proqnozlaşdırmaq üçün avtoregressiv integrasiya edilmiş hərəkətli ortalama (ing. ARIMA. Autoregressive Integrated Moving Average) modeli təklif edilmişdir. COVID-19 dünyada sürətlə yayılan və hazırda davam edən yeni növ pandemiyadır. Son dövrlərdə pandemiya yoluxanların sayı Azərbaycanda rekord həddə çatmışdır. Məhz bu səbəbdən COVID-19 pandemiyasının proqnozu məsələsinə baxılmışdır və bir neçə ayı əhatə edən real verilənlərlə eksperimentlərdə təklif edilmiş ARIMA modelinin COVID-19 zaman sıralarının proqnozlaşdırılması üçün müxtəlif parametrlərlə tətbiq edilmişdir. Verilənlər dedikdə, 22.01.2020 – 22.10.2020 tarixləri arasında Azərbaycan Respublikasının Səhiyyə Nazirliyi ([www.sehiyye.gov.az](http://www.sehiyye.gov.az)) tərəfindən rəsmi olaraq qeydiyyatı alınan gündəlik yoluxma hallarının sayı nəzərdə tutulur. Bu verilənlərdən istifadə etməklə, növbəti zaman aralığında ölkəmizdə baş verəcək yoluxma halları proqnoz edilmişdir. Bunun üçün ARIMA modelinə müxtəlif parametrlər verilmiş və uyğun olaraq hər bir modelin səhv dərəcəsi qiymətləndirilmişdir. Səhvin qiymətləndirilməsi üçün MAPE (Mean Absolute Percentage Error), MAE (Mean Absolute Error) və RMSE (Root Mean Square Error) funksiyaları istifadə edilib. Müqayisələr nəticəsində ən uyğun model seçilmişdir. Alınmış nəticələr ölkəmizdə pandemiya dövründə həm səhiyyə sistemi, həm də adi vətəndaşlar üçün vacib amildir. Əldə edilmiş nəticələr statistik metodların koronavirusa aid qeyri-stasionar zaman sıralarının proqnozlaşdırılmasının digər məsələlərə tətbiqində də məhsuldar ola biləcəyini təsdiqləyir.*

**Açar sözlər:** COVID-19, koronavirus, ARIMA, proqnozlaşdırma, zaman sıraları.

### **Giriş**

2019-cu ilin dekabrında Çinin Hubei əyalətinin Wuhan şəhərində aşkarlanmış naməlum təbiətli sətəlcəm tezliklə bütün dünyada indiyədək görünməmiş miqyasda yayıldı. Xəstəliyin törədicisinin yeni növ koronavirus olduğu müəyyən edildi və ona Ümumdünya Səhiyyə Təşkilatı (ÜST) tərəfindən COVID-19 (ing. *Corona Virus Disease 2019*) adı verildi. İlk vaxtlar bu xəstəlik “2019 yeni koronavirus” yaxud “2019-nCoV” adlandırılmışdı. COVID-19 koronavirus xəstəliyi dünyada sürətlə yayıldı, qlobal böhrana çevrildi və ÜST 11 mart 2020-ci ildə bu xəstəliyi pandemiya elan etdi [1]. COVID-19 geniş yayılaraq potensial zərərləri səbəbindən bütün dünyada insanların sağlamlığı və təhlükəsizliyi üçün böyük bir təhlükə yaratmaqdadır. Bu səbəbdən, yeni növ koronavirus olan COVID-19 pandemiyası və onun gələcəkdəki inkişaf tendensiyası hazırda daha çox istifadə edilən tədqiqat mövzusunə çevrildi. Belə bir sual ortaya çıxır ki, bəs biz Azərbaycanda baş verən COVID-19 pandemiyasını müəyyən vaxt aralığında gələcək bir neçə gün üçün proqnoz edə bilərikmi? Bu suala cavab vermək üçün biz proqnoz dəqiqliyinin böyük ölçülü verilənlərin istifadə edilməsindən asılı olan klassik və müasir proqnoz üsullarını araşdırırıq.

Koronavirus yoluxmalarına məruz qalan ölkələr arasında Azərbaycan da yer alır. Azərbaycanda 27 fevral 2020-ci il tarixində koronavirus pandemiyasının səbəb ola biləcəyi təhlükənin qarşısının alınması, profilaktik və təxirəsalınmaz tədbirlərin görülməsi üçün Nazirlər Kabineti nəzdində operativ qərargah yaradıldı. Bu pandemiya ilə mübarizə aparmaq üçün riyazi və statistik vasitələrdən istifadə etməklə proqnoz etmək lazımdır. COVID-19-un zaman sırası verilənlər dəsti həm qeyri-xətti, həm də stasionar olmayan nümunələri özündə cəmləşdirdiyi üçün yalnız bir model əsasında qərarlar qəbul etmək çətin olardı.

Bu tədqiqatın məqsədi koronavirus xəstəliyinin proqnozlaşdırılması üçün statistik metodlardan istifadə etməkdir. Tətbiq obyektini olaraq COVID-19 pandemiyası verilənləri götürülmüşdür [2].

### Əlaqədar işlərin analizi

Koronaviruslar ilk dəfə virusları adi soyuqdeymə xəstələrindən yetişdirən D.A.Tyrell və M. L.Bynoe tərəfindən 1966-cı ildə təyin edilmişdi [3]. Onları nüvəsi və səthdəki çıxıntıları ilə sferik virionlar şəklindəki formasına görə günəş tacına oxşadıqları üçün koronavirus adlandırılmışdır (latınca Corona – tac deməkdir). Koronaviruslar ailəsi xeyli genişdir, onların dörd altailəsi mövcuddur: alfa-, beta-, qamma- və delta-koronavirusları. Güman edilir ki, alfa- və beta-koronavirusları məməlilərdən, xüsusən də yarasalardan, qamma- və delta-koronavirusları donuz və quşlardan qaynaqlanır.

COVID-19 ilə Şiddətli Kəskin Tənəffüs Sindromu (*ing. Severe Acute Respiratory Syndrome, SARS*) və Yaxın Şərq Tənəffüs Sindromu (*ing. Middle East Respiratory Syndrome, MERS*) koronavirusları arasında genetik oxşarlıq vardır. COVID-19 aşağı tənəffüs yollarına təsir edən və insanlarda sətəlcəm kimi özünü göstərən SARS-CoV-2 adlı bir beta-koronavirusdan qaynaqlanır.

Əksər hallarda COVID-19 xəstəliyi yüngül keçir, lakin bəzi insanlarda bu xəstəlik olduqca ağır keçir və bəzi hallarda ölümlə nəticələnə bilər. Risk qrupuna yaşlı insanlar, həmçinin somatik xəstəlikləri (məsələn, ürək xəstəliyi və ya diabet) olan şəxslər aiddir. Hazırkı məlumatlara görə, COVID-19-un ölümlə nəticələnmə faizi (dünya ortalaması) SARS üçün 9.6% və MERS üçün 34.4% ilə müqayisədə təxminən 3.4%-dir (təəssüf ki, bu rəqəm yüksələ bilər) [4]. COVID-19-un inkubasiya dövrü – yoluxmadan xəstəliyin kliniki əlamətlərinin yaranmasına qədər olan müddət uzundur (təxminən 1–14 gün), SARS ilə müqayisədə, o, daha yoluxucudur və olduqca sürətlə yayılır.

[5]-də maşın təlimi metodlarından istifadə etməklə, bəşəriyyət üçün potensial təhdid olaraq qəbul edilən COVID-19 virusuna yoluxmuş xəstələrin sayı proqnozlaşdırılır. Xüsusilə, bu işdə COVID-19 virusunun təhdid amillərini proqnozlaşdırmaq üçün dörd standart proqnozlaşdırma modelləri olaraq, xətti reqressiya (LR), minimal mütləq daralma və seçim operatoru (LASSO), dəstək vektor maşını (SVM) və eksponensial hamarlaşdırma (ES) istifadə edilmişdir. Bu modellərin hər birinin köməyi ilə növbəti 10 gün ərzində yeni yoluxma hallarının, ölənlərin və sağalanların sayı üzrə üç növ proqnozlaşdırma aparılır. Nəticələr göstərir ki, ES modeli istifadə edilən bütün modellər arasında ən yaxşı nəticəni göstərir və bununla yanaşı, yeni yoluxma halları, ölüm və sağalma sayını proqnoz etməkdə yaxşı nəticə LR və LASSO modellərində müşahidə edildiyi halda, SVM modeli mövcud datasetlərdə verilən bütün proqnozlaşdırma ssenarilərində pis nəticə göstərmişdir.

[6]-da sadə və güclü metodla COVID-19 pandemiyasının davam etməsini proqnozlaşdırmaq üçün obyektiv yanaşma təklif edilir. Burada planlaşdırma və qərar qəbulu üçün böyük potensial təsirləri olan canlı proqnoz işinin trafiki təsvir edilir və təsdiq edilmiş COVID-19 üçün obyektiv proqnoz verilir.

Çin Milli Səhiyyə Komissiyasından 13 fevral 2020-ci ilə qədər verilən gündəlik verilənləri istifadə etməklə, 5 və 10 gün əvvəl hesablar təqdim edilir. Xüsusilə, əvvəllər fərqli yoluxucu xəstəliklərin yayılmasını proqnozlaşdırmaq üçün tətbiq edilən ümumiləşdirilmiş bir *logistic growth* modeli, *Richards growth* modeli və *sub-epidemic wave* modelindən istifadə edərək proqnozlar verilir. Üç modeldən istifadə etməklə sonrakı 10 gün üçün alınan nəticələr 23 fevral 2020-ci ilə qədər Quancaqdonda (*ing. Guangdong*) 65–81 əlavə (yuxarı hədd: 169–507) və 44–354 (yuxarı hədd: 141–7575) yoluxma olacağını proqnozlaşdırır. Ən yaxşı ssenaridə mövcud verilənlər hər iki vilayətdə də ötürülmənin yavaş olduğunu göstərir [7].

[8]-də COVID-19-un yayılma dinamikası üçün epidemioloji modellərə alternativ olaraq, Çin ərazisi boyunca yayılma və bitmə müddətini qiymətləndirmək məqsədilə real vaxt proqnozlaşdırılması üçün süni intellektə əsaslanan metodlar təklif edilir.

[9]-da Çinin əyalətlərində COVID-19 fəaliyyətini real-vaxt müddətində etibarlı şəkildə proqnozlaşdırmaq üçün, şərh olunan maşın təlimi metodları vasitəsi ilə xəstəlik izlərini rəqəmsal izlərlə birləşdirən yeni bir metodologiya təqdim edilir.

[10]-də Kanada və dünya üzrə mövcud COVID-19-un yayılmasının meyllərini və mümkün kəsilməsini proqnozlaşdırmaq üçün əsas əlamətlər götürülür. Burada COVID-19 hallarını proqnozlaşdırmaq üçün dərin təlim metodları olan Uzun-Qısamüddətli Yaddaş (*ing. Long-Short Term Model, LSTM*) şəbəkələrini təqdim edirik. LSTM metodunun nəticələrinə əsasən, bu yayılmanın potensial son nöqtəsinin 2020-ci ilin iyun ayı ətrafında olacağını proqnozlaşdırdıq. Bu yazıdakı təxminlərimiz 31 mart 2020-ci ilə qədər mövcud məlumatlara əsaslanır. Bildiyimizə görə, bir neçə tədqiqatçılar yoluxucu xəstəliklərin proqnozlaşdırılması üçün LSTM metodlarından istifadə edirlər.

[11]-də həm İranda, həm də onun əyalətləri səviyyəsində təsdiqlənmiş hallar üçün zaman sıralarının proqnozlaşdırılmasında uzunmüddətli yaddaşa əsaslanan LSTM dərin təlim metodu istifadə edilir. Mənbələrdən toplanmış verilənlər 2020-ci ilin 19 fevral – 22 mart və 19 fevral – 13 may zaman aralığını əhatə edir. Bundan başqa, rekurrent neyron şəbəkələri, mövsümi avtoregressiv inteqrasiya olunan hərəkətli ortalama (SARIMA), Holt qışının eksponensial hamarlanması və hərəkətli ortalama metodlarından istifadə edilir. Modelləri qiymətləndirmək üçün isə, MAE və MSE səhvlər funksiyasından istifadə edilib. İranın verilənlər bazasından toplanmış COVID-19 pandemiyasının zaman sıralarını proqnozlaşdırmaq üçün LSTM modeli digər modellərə nisbətən daha yaxşı nəticə göstərmişdir.

[12] John Hopkins alətlər panelindən real-vaxt verilənlərindən istifadə edərək bütün ölkələr üçün COVID-2019-un gələcəyini proqnozlaşdırmaqla yanaşı, gündəlik eksponensial davranışını anlamaq üçün maşın təlimi və dərin təlim modellərindən istifadə etməyi təklif edir.

COVID-2019 pandemiyasının Təsdiqlənmiş – Neqativ – Ölüm (*ing. Confirmed–NegativeReleased – Death*) halları ilə bağlı proqnozlaşdırılan nəticənin orijinal verilənlərə nə qədər yaxın olduğunu qiymətləndirmək üçün maşın təlimi metodlarından istifadə etməyin mümkün olub-olmadığı müəyyən edilir. Burada LSTM və GRU (*ing. Gated Recurrent Unit*) modelləri tətbiq edilir və proqnoz nəticələri həkimlər tərəfindən aparılan proqnoz nəticələrinə uyğun gəlir. Təklif olunan metod yüksək proqnoz dərəcəsinə malik olmuş və virusu düzgün müəyyənləşdirmək üçün daha tez işləmişdir [13].

COVID-2019 pandemiyasının yoluxma dərəcələrinin dəqiq müəyyənləşdirilməsi bu virusun təkamülünü başa düşmək və baş verən yoluxucu xəstəliklərin riskini müəyyən etmək üçün vacibdir. [14]-də müxtəlif ölkələrin verilənlər bazasından toplanan məlumatlar əsasında bütün xəstələrin genetik ardıcılıqla yoluxma dərəcəsi araşdırılır. Toplanmış verilənlər bazası nukleotid və kodon yoluxmalarını ayrıca müəyyən etmək üçün işlənir. Bundan əlavə, verilənlər bazasının ölçüsünə əsasən, müəyyən edilmiş yoluxma dərəcəsi dörd fərqli bölgəyə klassifikasiya edilir: Çin, Avstraliya, ABŞ və dünyanın qalan hissəsi. COVID-2019-un gələcək yoluxma dərəcəsini proqnozlaşdırmaq üçün rekurrent neyron şəbəkəsinin növü olan LSTM modeli tətbiq edilmişdir. LSTM modeli RMSE (Root Mean Square Error) testdə 0.06 və təlimdə 0.04 nisbətini verir ki, bu da optimallaşdırılmış bir dəyərdir.

Bütün dünya üzrə COVID-2019 pandemiyası qlobal təhlükə hesab edildi və onun mümkün yayılması hadisəsini proqnozlaşdırmaq üçün müxtəlif riyazi modellərdən istifadə edərək bir neçə tədqiqatlar aparılır. Müxtəlif faktorlara və analizlərə əsaslanan bu riyazi modellər potensial səhvlərə məruz qalır. [15]-də COVID-2019-un yayılmasını proqnozlaşdırmaqda faydalı ola biləcək sadə bir ekonometrik model təklif edilir. COVID-2019-un yayılması və hadisənin epidemioloji meylini proqnoz etmək üçün Johns Hopkins epidemioloji verilənlər üzərində ARIMA modelinin tətbiqi həyata keçirilmişdir.

[16]-da milli və regional səviyyədə İtalyanın epidemioloji verilənlərindən istifadə etməklə, 4 aprel 2020-ci ildən sonrakı dövrdə epidemiya meylini proqnozlaşdırmaq üçün ARIMA modeli tətbiq edilir. Verilənlər İtaliya Səhiyyə Nazirliyi tərəfindən rəsmi qeydiyyatla alınan 2020-ci ilin

20 fevral – 4 aprel tarixlərinə aid gündəlik təsdiqlənmiş halların sayından ibarətdir (www.salute.gov.it). Bu modelin əsas üstünlüyü ondan ibarətdir ki, idarə etmək və uyğunlaşdırmaq asandır. Üstəlik, hipotetik epidemiyanın infeksiya nöqtəsini və son ölçüsünü təklif edərək əsas meyllər barədə ilk anlayış verə bilər.

### COVID-19-un proqnozlaşdırılması üçün tədqiqat metodologiyası

Bu məqalədə COVID-19-un proqnozlaşdırılması üçün statistik model olan ARIMA modelləri tətbiq edilmişdir. İlk olaraq ARIMA modellərinin işləmə prinsipi müzakirə edilir, növbəti mərhələdə isə onların tətbiqi reallaşdırılır. Bu bölmədə MAPE, MAE və RMSE funksiyaları isə modellərin nəticələrini qiymətləndirmək üçün təsvir edilmişdir.

#### A. ARIMA modeli.

ARIMA modeli adətən zaman sıralarının analizi üçün istifadə edilir, çünki bir dəyişənin gələcək qiymətini, yalnız tarixi qiymətlərinin dəyişikliyinə əsaslanaraq proqnoz edə bilər. ARIMA modeli bir-biri ilə əlaqəli olan və hər biri ayrı parametrləri təyin edən üç əsas konseptə bölünür:

- AR: Proqnozlaşdırılan dəyişənin regressiyası əvvəlki qiymətlərə (gecikmələr, *ing. lags*) əsaslanır.  $p$  parametri bu proqnozların sayını müəyyənləşdirir.

$$\Delta^{(d)}y_t = c + \sum_{i=1}^p \varphi_i \Delta^{(d)}y_{t-i} + \varepsilon_t. \quad (1)$$

burada  $y_t$  və  $\varepsilon_t$  uyğun olaraq  $t$  zaman müddətində cari qiymət və təsadüfi səhv,  $\varphi_i$  ( $i = 1, 2, \dots, p$ ) modelin parametrləri və  $c$  sabitdir.

- I: regressiya üçün istifadə edilən verilənlərin qiymətləri stasionar olana kimi diferensiallanır.  $d$  parametri stasionarlıq üçün lazım olan diferensiallama tərtibini bildirir

$$\Delta y_t = y_t - y_{t-1}$$

$$\Delta^{(2)}y_t = \Delta y_t - \Delta y_{t-1}$$

...

$$\Delta^{(d)}y_t = \Delta^{(d-1)}y_t - \Delta^{(d-1)}y_{t-1}.$$

- MA: keçmişdən gələn səhvlərin birləşməsinə əsaslanır (səhvlərin gecikməsi, *ing. lags of errors*).  $q$  parametri gecikmiş proqnoz səhvlərinin sayını bildirir.

$$\Delta^{(d)}y_t = \mu + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t. \quad (2)$$

burada  $\mu$  sıranın orta qiymət göstəricisi,  $\theta_j$  ( $j = 1, 2, \dots, q$ ) modelin parametrləridir. Beləliklə, MA modelini zaman sırasına uyğunlaşdırmaq AR modelinə uyğunlaşdırmaqdan daha mürəkkəbdir, çünki təsadüfi səhv şərtləri əvvəlcədən məlum olmur.

Riyazi olaraq ARIMA ( $p, q$ ) modeli aşağıdakı kimi göstərilir:

$$\Delta^{(d)}y_t = c + \sum_{i=1}^p \varphi_i \Delta^{(d)}y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t. \quad (3)$$

Yuxarıda təsvir olunan ARMA modellərindən yalnız stasionar zaman sırası verilənləri üçün istifadə edilə bilər. Lakin praktikada bir çox zaman sırası, məsələn, sosial-iqtisadi və iş həyatı ilə bağlı olan qeyri-stasionar davranışlar mövcuddur. Trend və mövsümi komponentləri özündə

cəmləşdirən zaman sıraları da təbiətdə qeyri-stasionardır. Beləliklə, tətbiqi nöqtəyi- nəzərdən ARMA modelləri praktikada tez-tez rast gəlinən qeyri-stasionar zaman sıralarını düzgün təsvir etmək üçün yetərli deyildir. Bu səbəbdən ARIMA modeli təklif olunur. Bu model ARMA modelinin ümumiləşdirilməsidir və qeyri-stasionar hal da mövcuddur.

ARIMA modellərində verilənlərin sonlu fərqi uyğunlaşdırmaqla qeyri-stasionar zaman sırasını stasionar vəziyyətə gətirmək mümkündür. ARIMA  $(p,d,q)$  modelinin düsturu aşağıdakı kimidir [17, 18, 19].

$$\Delta^{(d)}y_t = c + \sum_{i=1}^p \varphi_i \Delta^{(d)}y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t. \quad (4)$$

#### B. COVID-19 datasetləri.

Açıq mənbələrin köməyi ilə koronavirus verilənlərini əldə etmək mümkündür. Bu mənbələrə kaggle.com, github.com və s. aiddir. Bizim istifadə etdiyimiz datasetlər github.com saytıdan link şəklində götürülmüşdür. Burada `css_COVID19_timeseries` qovluğunda bütün ölkələrin təsdiqlənmiş (*ing. confirmed*), ölüm (*ing. death*) və sağalma (*ing. recovered*) halları təsvir edilmiş üç növ csv faylları mövcuddur. 22.01.2020 – 22.10.2020 tarixləri arasında bu üç əlamətlərdən ibarət csv faylları yaradılmışdır. Tədqiqat işi göstərilən dövrdə baş verən halların yalnız Azərbaycana aid olan hissəsində aparılmışdır. Datasetdə mövcud olan əlamətlər: gündəlik tarix, təsdiqlənmiş və ya yoluxma halları, sağalma halları və ölüm halları mövcuddur. Datasetdə cəmi 326 gün ərzində baş verən bu halları nəzərə alaraq, növbəti iki həftə ərzində baş verəcək yoluxma halları proqnozlaşdırılmışdır.

Real verilənlər adətən natamam və müəyyən boşluqlardan ibarət olur və buna görə də ARIMA modellərinə təlim üçün verilmədən öncə verilənlərin əvvəlcədən işlənməsi (*ing. preprocessing data*) prosesi yerinə yetirilir [20, 21].

#### C. Nəticələrin qiymətləndirilməsi metrikaları.

Tədqiqatımızda təklif edilən modelin nəticəsini qiymətləndirmək üçün bəzi funksiyalar tətbiq edilmişdir, yuxarıda qeyd edildiyi kimi, bu funksiyalara MAPE, MAE və RMSE aiddir.

$$\text{MAPE} = \frac{1}{n} \sum \frac{|e_t|}{d_t}. \quad (5)$$

$$\text{MAE} = \frac{1}{n} \sum |e_t|. \quad (6)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum e_t^2}. \quad (7)$$

Burada  $n$  proqnozlaşdırılan dövrlərin sayı,  $e_t$  isə  $t$  zamanda səhvi (real qiymət ilə proqnoz qiymətin fərqi) göstərir [17].

#### D. ACF və PACF funksiyaları.

ARIMA modelinin tətbiqi zamanı ilkin olaraq ACF və PACF funksiyalarından istifadə etməklə modelin parametrləri seçilir. Zaman sırasında keçmiş və gələcək verilənlərin bir-biri ilə necə əlaqəli olduğunu müəyyən etmək üçün ACF istifadə edilir, qiyməti  $(-1, 1)$  aralığında dəyişir. ACF-in düsturu aşağıdakı kimi təyin edilir.

$$\rho_x(n) = \text{Corr}(x_t - x_{t-h}). \quad (8)$$

PACF funksiyası isə zaman sıralarının özünün keçmişi ilə qismən korrelyasiyasını verir. PACF-in qiyməti  $(-1, 1)$  aralığında dəyişir.

Qalıq testindən keçən bütün modellər AIC (*ing. Akaike information criterion*) istifadə edilməklə müqayisə edilir. Ən az AIC-ə sahib olan model ən yaxşı model olaraq seçilir.

$$\text{AIC} = 2k - 2 \ln(L), \quad (9)$$

burada,  $k$  modeldə qiymətləndirilən parametrlərin sayı,  $L$  isə model üçün ehtimal funksiyasının maksimum qiymətidir [19, 20, 22].

AIC müəyyən bir model tərəfindən itirilən nisbi verilənlərin miqdarını qiymətləndirir: bir model nə qədər az verilənlər itirirsə, o modelin keyfiyyəti bir o qədər də yüksək olur [23].

### Eksperimental nəticələr

Bu məqalədə Azərbaycan üçün COVID-19 verilənlər bazasında ARIMA proqnozlaşdırma modelləri tətbiq edilmişdir. Tədqiqatımızda 22.01.2020–22.10.2020 zaman intervalında Azərbaycanda baş verən COVID-19 pandemiyasının yoluxma hallarının sayı götürülmüşdür. Tətbiq edilən ARIMA modelinin müxtəlif parametrlərə görə nəticələri yoxlamaq üçün MAPE, MAE və RMSE funksiyaları ilə müqayisə edilmişdir. Tədqiqat Python proqramlaşdırma dilində yerinə yetirilmişdir.

Dataset təsdiqlənmiş (yoluxma), ölüm və sağalma halları olan üç əsas əlamətdən ibarətdir. İstifadə olunan datasetin 85%-i təlim, 15%-i isə test verilənlərinə ayrılmışdır. ARIMA modellərinin parametrləri və bu parametrlərə uyğun qiymətlər cədvəl 1-də, həmin parametrlərə uyğun nəticələrin qiymətləri isə cədvəl 2-də göstərilmişdir. Qeyd edilən parametrlər təsadüfi seçilmişdir və onların nəticələrinin qiymətləri bir-birinə yaxındır.

Cədvəl 1  
Təsadüfi seçilmiş ARIMA modelinin parametrləri

Model	Parametrlər	Qiymətlər
ARIMA	(p, d, q)	(1,1,1)
		(2,2,0)
		(3,2,4)
		(4,1,2)
		(5,2,2)
		(7,2,4)

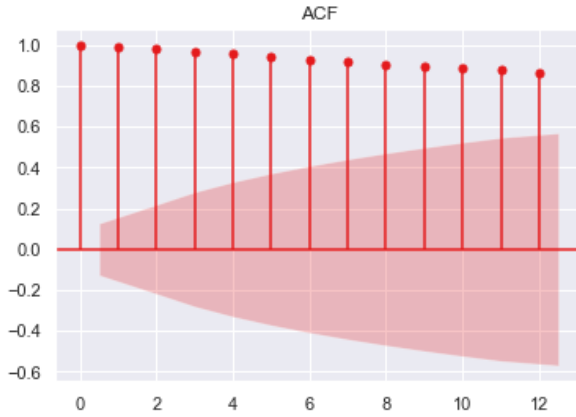
Cədvəl 2  
ARIMA modelinin nəticələrinin qiymətləndirilməsi üçün istifadə edilən funksiyaların nəticələri

Model	MAPE	MAE	RMSE
ARIMA(1,1,1)	17.4	11728.5	15443.4
ARIMA(2,2,0)	16.7	11237.0	14784.0
ARIMA(3,2,4)	15.7	10587.1	13957.8
ARIMA(4,1,2)	17.8	11958.2	15702.3
ARIMA(5,2,2)	16.9	11358.4	14920.2
ARIMA(7,2,4)	15.5	10465.9	13852.5
ARIMA(2,2,6)	15.7	10546.4	13907.9

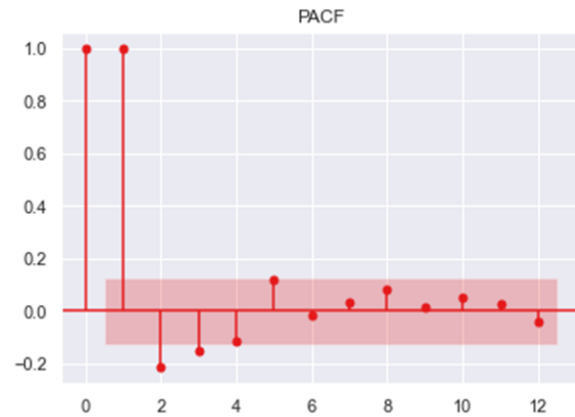
Şəkil 1 və şəkil 2-də ARIMA modelinin parametrlərini seçmək üçün istifadə olunan ACF və PACF funksiyalarının qrafikləri göstərilir.

ACF və PACF qrafiklərində görüldüyü kimi, gecikmə əlaqələri adlanan xətlər mövcuddur. Burada  $p$  parametrinin qiymətini PACF qrafikindəki,  $q$  parametrinin qiymətini isə ACF qrafikindəki gecikmə əlaqələri müəyyən edir. ACF qrafikində ilk əlaqə, PACF qrafikində isə ilk iki əlaqə vacib hesab edilir. Modelin dərəcəsi PACF qrafikində ilk iki gecikməyə aid əlaqə vacib olduğu üçün  $p=2$ -dir, ACF qrafikində isə əlaqələr arasında kiçik fərq mövcuddur və  $q=6$  qiymətini müəyyən edir, daha sonra isə, uyğun olaraq ARMA(2,6) modeli seçilir və model stasionar hala gətirilir.

Nəticə etibarlı ilə tədqiqatımızda ARIMA (2,2,6) modeli seçilmişdir, lakin effektivliyin yüksək olması və müqayisələr aparmaq üçün modelin parametrlərinə bir neçə müxtəlif qiymətlər verilmiş, sonra bu parametrlərə uyğun proqnoz modelləri, yuxarıda qeyd edildiyi kimi, səhvlər funksiyalarının köməyi ilə müqayisə edilmişdir (cədvəl 2).

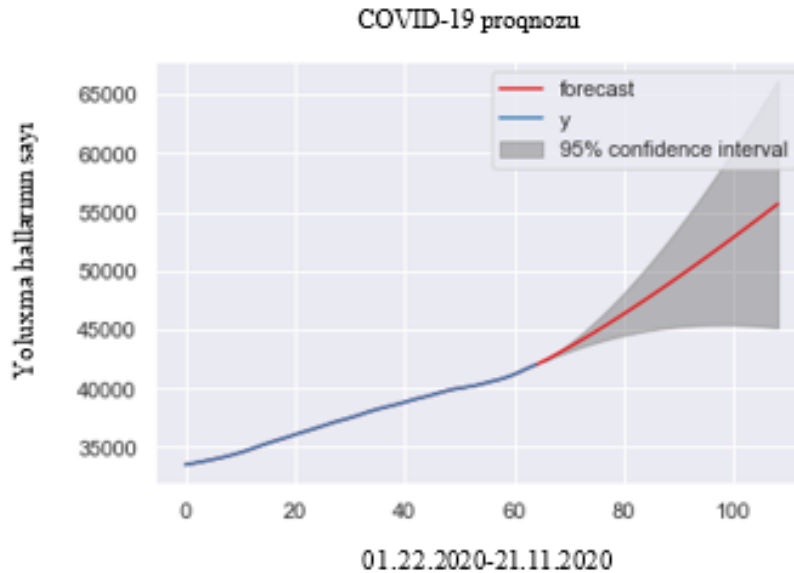


Şəkil 1. ACF funksiyası



Şəkil 2. PACF funksiyası

ARIMA modelindən istifadə etməklə Azərbaycanda COVID-19 pandemiyasının növbəti iki həftə ərzində yoluxma sayının proqnozlaşdırma nəticələri şəkil 3-də təsvir edilmişdir.



Şəkil 3. ARIMA modelindən istifadə etməklə yoluxma sayının proqnozu

2020-ci ilin yanvar–noyabr ayları arasında 310 gün ərzində COVID-19-un gündəlik yoluxma, ölüm və sağalma halları üçün verilənlər toplandı və Azərbaycan üzrə yoluxma halları proqnoz edildi. Eksperimentin nəticəsi göstərir ki, ölkəmizdə növbəti iki həftə ərzində yoluxma halları kəskin artacaq və bu sayın 55000-dən çox olacağı gözlənilir.

Cədvəl 2-də tətbiq edilmiş ARIMA modelinin parametrləri qiymətləndirilmişdir. Burada MAPE funksiyasına görə, parametrlər arasında ən az səhv verən ARIMA(3,2,4) və ARIMA(2,2,6) modelləri, MAE və RMSE funksiyalarına görə isə, ARIMA(7,2,4) və ARIMA(2,2,6) modelləri seçilmişdir. Nəticə olaraq, proqnozlaşdırma üçün ARIMA(2,2,6) və təsadüfi seçilən ARIMA(7,2,4) modelləri ən uyğun modellər hesab edilir.

### Nəticə

COVID-19 dünyada sürətlə yayılan və hazırda davam edən yeni növ pandemiyadır. Bildiyimiz kimi, son dövrlərdə pandemiya yoluxanların sayı Azərbaycanda da rekord həddə çatmışdır. Yoluxmalar kəskin artıqda, ölkədə bununla bağlı vacib qərarlar verilməlidir. Qərarlar sürətlə dəyişən bir mühitdə baş verir və səhv məlumatlandırılma bilər. Nəticə etibarilə proqnozlar, yoluxmaların qarşısının alınması strategiyalarının səmərəliliyi və vaxtı üçün vacib məsələdir. Bu tədqiqatda Azərbaycan üçün COVID-19 pandemiyasının proqnozu məsələsinə baxdıq və bir neçə ayı əhatə edən real verilənlərlə eksperimentlərdə təklif edilmiş ARIMA modelinin COVID-19 zaman sıralarının proqnozlaşdırılmasında müxtəlif parametrlərlə tətbiq etdik. Bundan əlavə, ARIMA modelini zaman sıralarının proqnozlaşdırılması üçün izah etdik və gələcək on dörd gün ərzində yoluxanların sayını proqnozlaşdırdıq. Əlbəttə, bu kimi proqnozlaşdırma ölkəmizdə COVID-19 pandemiyası dövründə həm səhiyyə sistemi, həm də adi vətəndaşlar üçün əhəmiyyətlidir. Bu tədqiqatımız, ilkin olaraq ARIMA modeli əsasında reallaşdırılmışdır, lakin gələcək tədqiqatlarda baxılan məsələnin həlli üçün rekurrent neyron şəbəkələrinin tətbiqi nəzərdə tutulur.

### Ədəbiyyat

1. The World Health Organization (WHO). Coronavirus disease (COVID-2019) situation reports. URL: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>
2. Chakraborty T., and Ghosh I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis // *Chaos, Solitons & Fractals*, 2020, vol.135, 109850. DOI: 10.1016/j.chaos.2020.109850.
3. Tyrrell D. A., and Bynoe M. L. Cultivation of viruses from a high proportion of patients with colds // *Lancet*, 1966, vol. 1 (7428), pp.76–77.
4. Velavan T.P., and Meyer C.G. The COVID-19 epidemic // *Tropical medicine & international health*, 2020, vol.25(3), pp.278–280.
5. Rustam F., Reshi A. A., Mehmood A., Ullah S., On B., Aslam W., and Choi G. S. COVID-19 future forecasting using supervised machine learning models // *IEEE Access*, 2020, vol.8, pp.101489–101499. DOI: 10.1109/ACCESS.2020.2997311.
6. Petropoulos F., and Makridakis S. Forecasting the novel coronavirus COVID-19 // *PloS one*, 2020, vol.15(3), e0231236. DOI: 10.1371/journal.pone.0231236.
7. Roosa K., Lee Y., Luo R., Kirpich A., Rothenberg R., Hyman J.M., and Chowell G. Short-term forecasts of the COVID-19 epidemic in Guangdong and Zhejiang, China: February 13–23, 2020 // *Journal of Clinical Medicine*, 2020, 9(2), 596. DOI: 10.3390/jcm9020596.
8. Hu Z., Ge Q., Jin L., and Xiong M. Artificial intelligence forecasting of COVID-19 in China. *arXiv preprint arXiv:2002.07112*, 2020, 20 p.
9. Liu D., Clemente L., Poirier C., Ding X., Chinazzi M., Davis J. T., Vespignani A., and Santillana M. A machine learning methodology for real-time forecasting of the 2019-2020 COVID-19 outbreak using Internet searches, news alerts, and estimates from mechanistic models. *arXiv preprint arXiv:2004.04019*, 2020, 23 p.
10. Chimmula V.K.R., and Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks // *Chaos, Solitons & Fractals*, 2020, vol.135, 109864. DOI: 10.1016/j.chaos.2020.109864.
11. Azarafza M., Azarafza M., and Tanha J. COVID-19 Infection Forecasting based on deep learning in Iran. *medRxiv*. 2020, 7 p.



12. Pun N. S., Sonbhadra S. K., and Agarwal S. COVID-19 epidemic analysis using Machine Learning and DeepLearning algorithms. medRxiv. 2020, 10 p.
13. Bandyopadhyay S. K., and Dutta S. Machine learning approach for confirmation of COVID-19 cases: Positive, negative, death and release. medRxiv, 2020, 10 p.
14. Pathan R.K., Biswas M., and Khandaker M.U. Time series prediction of COVID-19 by mutation rate analysis using recurrent neural network-based LSTM model // Chaos, Solitons & Fractals, 2020, vol. 138, 110018. DOI: 10.1016/j.chaos.2020.110018.
15. Benvenuto D., Giovanetti M., Vassallo L., Angeletti S., and Ciccozzi M. Application of the ARIMA model on the COVID-2019 epidemic dataset // Data in brief, 2020, vol.29, 105340. DOI: 10.1016/j.dib.2020.105340.
16. Perone G. An ARIMA model to forecast the spread and the final size of COVID-2019 epidemic in Italy. medRxiv. 2020, 14 p.
17. Vandepu N. Forecast KPI: RMSE, MAE, MAPE and Bias. Data Science for Supply Chain Forecast, 2019, 237 p.
18. Medium, <https://medium.com/@kangeugine/time-series-arima-model-11140bc08c6>
19. Liu Q., Liu X., Jiang B., and Yang W. Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model // BMC Infectious Diseases, 2011, vol.11 (1), 7 p. DOI: 10.1186/1471-2334-11-218.
20. GitHub, <https://github.com/owid/covid-19-data/tree/master/public/data>
21. Kaggle, <https://www.kaggle.com/datasets>
22. Shahid F., Zameer A., and Muneeb M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM // Chaos, Solitons & Fractals, 2019, vol. 140, Article 110212. DOI: 10.1016/j.chaos.2020.110212.
23. Wikipedia, [https://en.wikipedia.org/wiki/Akaike\\_information\\_criterion](https://en.wikipedia.org/wiki/Akaike_information_criterion)

#### **УДК 004.94**

**Садиева Фирангиз И.**

Институт Информационных Технологий НАНА, Баку, Азербайджан  
[sadiyeva.firengiz@gmail.com](mailto:sadiyeva.firengiz@gmail.com)

#### **Применение моделей ARIMA для прогнозирования COVID-19 в Азербайджане**

В статье предлагается модель авторегрессионного интегрированного скользящего среднего (Autoregressive Integrated Moving Average, ARIMA) для прогнозирования пандемии COVID-19. COVID-19 – это пандемия нового типа, которая быстро распространяется по миру и в настоящее время продолжается. В последнее время количество случаев пандемии в Азербайджане достигло рекордного уровня. По этой причине был рассмотрен прогноз пандемии COVID-19, а временной ряд COVID-19 модели ARIMA, предложенной в экспериментах с реальными данными за несколько месяцев, использовался с разными параметрами для прогнозирования. Согласно данным, учитывается количество ежедневных инфекций, официально зарегистрированных Министерством здравоохранения Азербайджанской Республики ([www.sehiyye.gov.az](http://www.sehiyye.gov.az)) в период с 22.01.2020 по 22.10.2020 года. На основе этих данных прогнозируется заболеваемость в нашей стране в ближайший период. С этой целью модели ARIMA были присвоены различные параметры и соответственно оценена частота ошибок каждой модели. Для оценки ошибки использовались функции MAPE (Mean Absolute Percentage Error), MAE (Mean Absolute Error) и RMSE (Root Mean Square Error). В результате сравнений была выбрана наиболее подходящая модель. Полученные результаты являются важным фактором как для системы здравоохранения, так и для простых граждан во время пандемии в нашей стране. Результаты подтверждают, что статистические методы могут быть эффективными при применении прогнозов нестационарных временных рядов коронавируса к другим вопросам.

**Ключевые слова:** COVID-19, коронавирус, ARIMA, прогноз, временные ряды.

**Firangiz I. Sadiyeva**

Institute of Information Technology of ANAS, Baku, Azerbaijan

[sadiyeva.firengiz@gmail.com](mailto:sadiyeva.firengiz@gmail.com)

**Application of ARIMA Models for forecasting COVID-19 in Azerbaijan**

The article proposes an Autoregressive Integrated Moving Average (ARIMA) model to predict the COVID-19 pandemic. COVID-19 is a new type of pandemic that is spreading rapidly around the world and is currently continuing. Recently, the number of pandemic cases in Azerbaijan has reached the highest rate. For this reason, the forecast of the COVID-19 pandemic is reviewed, and the COVID-19 time series of the ARIMA model proposed in experiments with real data covering several months is used with different parameters for forecasting. According to the data, the number of daily infections officially registered by the Ministry of Healthcare of the Republic of Azerbaijan ([www.sehiyye.gov.az](http://www.sehiyye.gov.az)) between 22.01.2020 - 22.10.2020 is considered. Using these data, the incidence of infections in our country in the next period is forecasted. For this purpose, various parameters are given to the ARIMA model and the error rate of each model was evaluated accordingly. MAPE (Mean Absolute Percentage Error), MAE (Mean Absolute Error), and RMSE (Root Mean Square Error) functions are used to estimate the error. As a result of comparisons, the most suitable model is selected. The results obtained are an important factor for both the health care system and ordinary citizens during the pandemic in our country. The results confirm that statistical methods can be effective in applying non-stationary coronavirus time-series predictions to other issues.

**Keywords:** *COVID-19, coronavirus, ARIMA, forecast, time-series.*