

## Comparison of Models in Predicting Cumulative Cases of Hospitalization<sup>1</sup> and Death of Covid-19 (Case Study: Bahabad city)



### Abstract

Karimizarchi MH<sup>2</sup>, Shishebori D<sup>3</sup>

**Introduction:** Coronavirus disease 2019 is a respiratory disease caused by acute respiratory syndrome coronavirus-2. Forecasting the number of new cases and deaths during today's can be a useful step in predicting the costs and facilities needed in the future. This study aims to modeling, comparing the performance of models, and predict new cases and deaths efficiently in the future.

**Methods:** In this article nine popular forecasting techniques are tested on the data of Covid-19 in Bahabad city as a case study. Using the evaluation criteria of mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and the mean absolute percentage of error (MAPE) of the models are compared.

**Results:** The results of the analysis showed that the best model according to the evaluation criteria for forecasting cumulative cases of hospitalization of Covid-19 is the cubic spline smoothing model, and cumulative cases of death, KNN regression model. Also, autoregressive neural network and theta models for hospitalization cases, and for death cases, autoregressive neural network model has the worst performance among other models.

**Conclusion:** This study can provide a proper understanding of the spread of covid-19 disease in this region so that by taking precautionary measures and formulating appropriate policies, this epidemic can be effectively overcome. Also, unlike other studies, this study uses 9 different techniques and their comparison, which in turn increases the confidence factor in decision making. Also, an important point is that the data should be updated in real time.

**Keywords:** Covid-19, forecasting, pandemic, Statistical modeling, time series.

---

1- **Cite this article:** Karimizarchi, Mohammad Hossein; Shihbari, Daoud (1402). Comparison of models in predicting cumulative cases of hospitalization and death of covid-19 (case study: Bahabad city). *Health Management*, 14(49): 17-7. DOI: 10.30495/JHM.2023.72246.11112

2- M.Sc. degree, Department of Industrial Engineering, Yazd University, Yazd, Iran

3- Associate Professor, Department of Industrial Engineering, Yazd University, Yazd, Iran, (Corresponding Author), shishebori@yazd.ac.ir

## مقایسه مدل‌ها در پیش‌بینی موارد تجمعی بستری و فوت کووید-۱۹ (مطالعه موردی: شهرستان بهاباد)<sup>۱</sup>

محمدحسین کریمی‌زارچی<sup>۲</sup> / داود شیشه‌بری<sup>۳</sup>

چکیده

**مقدمه:** بیماری کووید-۱۹، یک بیماری تنفسی است که در اثر سندرم تنفسی حاد کرونا ویروس-۲ ایجاد شده است. پیش‌بینی تعداد موارد جدید و مرگ‌ومیر می‌تواند گام مفیدی در پیش‌بینی هزینه‌ها و امکانات مورد نیاز در آینده باشد. هدف از این مطالعه مدل‌سازی، مقایسه عملکرد مدل‌ها و پیش‌بینی موارد جدید بستری و مرگ‌ومیر در آینده نزدیک است.

**روش پژوهش:** در این مقاله ۹ تکنیک پیش‌بینی بر روی داده‌های کووید-۱۹ شهرستان بهاباد استان یزد تحت آزمایش قرار گرفت و با استفاده از معیارهای ارزیابی میانگین مربعات خطا (MSE)، جذر میانگین مربعات خطا (RMSE)، میانگین قدر مطلق خطا (MAE) و میانگین درصد قدرمطلق خطا (MAPE) مدل‌ها باهم مقایسه شدند.

**یافته‌ها:** نتایج تحلیل نشان داد، بهترین مدل با توجه به معیارهای ارزیابی مذکور برای پیش‌بینی موارد تجمعی بستری کووید-۱۹ مدل هموارسازی اسپلاین مکعبی و برای موارد تجمعی فوت مدل رگرسیون KNN می‌باشد. همچنین مدل شبکه‌های عصبی اتورگرسیو و مدل تتا برای موارد بستری و برای موارد فوت مدل شبکه‌های عصبی اتورگرسیو دارای بدترین عملکرد را در میان دیگر مدل‌ها دارا می‌باشد.

**نتیجه‌گیری:** این مطالعه می‌تواند درک مناسبی از روند شیوع بیماری کووید-۱۹ در این منطقه ارائه کند تا با اتخاذ اقدامات احتیاطی و تدوین سیاست‌های مناسب بتوان به نحو احسن از این بیماری عبور کرد. همچنین برخلاف مطالعات دیگر این مطالعه، از ۹ تکنیک متفاوت و مقایسه آن‌ها، استفاده کرده است که به نوبه خود ضریب اطمینان را در تصمیم‌گیری بالا برده است. همچنین نکته‌ای که حائز اهمیت می‌باشد این است که باید داده‌ها در زمان واقعی بروز شوند.

**کلید واژه‌ها:** کووید-۱۹، پاندمیک، سری زمانی، پیش‌بینی، مدل‌سازی آماری.

۱- استناد به این مقاله: کریمی‌زارچی، محمدحسین؛ شیشه‌بری، داود (۱۴۰۲). مقایسه مدل‌ها در پیش‌بینی موارد تجمعی بستری و فوت کووید-۱۹ (مطالعه

موردی: شهرستان بهاباد). مدیریت بهداشت و درمان، ۱۴(۴۹): ۷-۱۷. DOI: 10.30495/JHM.2023.72246.11112

۲- کارشناسی ارشد مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه یزد، یزد، ایران

۳- دانشیار، گروه مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه یزد، یزد، ایران، (نویسنده مسئول)، پست الکترونیک: shishebori@yazd.ac.ir

## مقدمه

بیماری کووید-۱۹، یک بیماری تنفسی است که در اثر سندرم تنفسی حاد کرونا ویروس-۲ ایجاد شده است. این ویروس اولین بار در دسامبر ۲۰۱۹ شناسایی شد [۱]. در ۱۱ مارس ۲۰۲۰، سازمان بهداشت جهانی کرونا را به عنوان بیماری همه‌گیر جهانی اعلام کرد. مدل‌سازی و پیش‌بینی ویروس‌ها و ویژگی‌های اپیدمیولوژیک از موارد مهم در تهیه تجهیزات مورد نیاز برای کنار آمدن با پیامدهای آن‌ها است. بنابراین مدل‌سازی و پیش‌بینی تعداد موارد کووید-۱۹ تأیید شده و مرگ‌ومیر، نقش مهمی در برنامه‌ریزی برای کنترل کووید-۱۹ در جهان دارد. مدل‌های مختلفی برای پیش‌بینی بیماری‌های عفونی و تعداد موارد کووید-۱۹ در آینده استفاده شدند [۲]. ژانگ و همکاران [۳] برای پیش‌بینی تب حصبه از مدل سری زمانی اتورگرسیو میانگین متحرک تلفیق شده فصلی (SARIMA) و سه مدل شبکه‌های عصبی شامل: شبکه‌های عصبی پس انتشار، شبکه‌های عصبی تابع شعاعی و شبکه‌های عصبی بازگشتی المان استفاده کردند. همچنین پژوهشگران دیگری بطور مشابه از مدل‌های اتورگرسیو میانگین متحرک تلفیق شده (ARIMA) برای پیش‌بینی بیماری‌های عفونی مانند سل [۴]، تب دانگ [۵] و تب مالت [۶] استفاده کرده‌اند. ابوالقاسم گلخندان و سمیه صحرائی [۷] برای پیش‌بینی سرانه مخارج بهداشتی در ایران تا دوره زمانی ۱۴۲۰ از الگوریتم‌های ژنتیک و بهینه‌سازی انبوه ذرات استفاده کردند. سیلان [۸] از مدل‌های ARIMA برای پیش‌بینی روند کووید-۱۹ در ایتالیا، اسپانیا و اروپا استفاده کرد. در این پژوهش چندین مدل ARIMA با پارامترهای مختلف با استفاده از معیار ارزیابی درصد میانگین خطای مطلق برای کشورها انتخاب کرده‌اند. لوتفی و بورکو [۹] برای پیش‌بینی تعداد موارد تأیید شده و مرگ‌ومیر کووید-۱۹ از مدل ARIMA در مرکز کنترل و پیشگیری اروپا استفاده کردند. خان و گوپتا [۱۰] برای پیش‌بینی تعداد موارد کووید-۱۹ در هند از مدل ARIMA و داده‌های جمع‌آوری شده از

۳۱ ژانویه ۲۰۲۰ تا ۲۵ مارس ۲۰۲۰ پیاده‌سازی و آن را با داده‌های جمع‌آوری شده از ۲۶ مارس ۲۰۲۰ تا ۴ آوریل ۲۰۲۰ مورد ارزیابی قرار دادند. الیاسی و همکاران [۱۱] از مجموعه داده‌های ارائه شده توسط مرکز پیشگیری و کنترل بیماری اروپا تا ۳ دسامبر ۲۰۲۰، برای پیش‌بینی کووید-۱۹ در مالزی، مراکش و عربستان سعودی استفاده کرده‌اند. و از مدل‌های یادگیری عمیق برای پیش‌بینی در این سه کشور مورد ارزیابی قرار گرفت. همچنین مخایرز و آلف [۱۲] از الگوریتم‌های پیش‌بینی کوتاه‌مدت برای پیش‌بینی تعداد مبتلایان و مرگ‌ومیرها در آینده کوتاه استفاده کردند. پیش‌بینی با استفاده از مدل ARIMA، ARIMA ترکیبی، هولت-وینترز و هموارسازی نمایی با تعداد ۵۷ داده انجام شد. پس از ارزیابی، متوجه شدند که هموارسازی نمایی بهترین مدل برای موارد بهبودیافته و تأیید شده بوده و هولت-وینترز برای موارد مرگ‌ومیر می‌باشد. در مقاله‌ای دیگر که در سال ۲۰۲۲ انجام شد لی و همکاران [۱۳] از مدل ARIMA برای پیش‌بینی کرونا در بریتانیا، آفریقای جنوبی، برزیل و نیجریه استفاده کرده‌اند.

در مطالعه حاضر، موارد تجمعی کووید-۱۹ (موارد بستری و مرگ‌ومیر) برای شهرستان بهاباد از تاریخ ۱۳۹۸/۱۲/۰۷ الی ۱۴۰۰/۰۹/۲۸ با استفاده از تکنیک‌های سری‌های زمانی و یادگیری ماشین مدل‌سازی شده و براساس معیارهای ارزیابی بهترین مدل انتخاب شده است.

## روش پژوهش

در این مطالعه از مدل‌های ARIMA، هموارسازی نمایی، هولت-وینترز (Holt-Winters)، رگرسیون KNN، تنه، شبکه عصبی اتورگرسیو (ARNN)، هموارسازی اسپلاین مکعبی، روش STL و روش BATS برای ایجاد یک پیش‌بینی ۱۴ روزه از موارد تجمعی کووید-۱۹ برای شهرستان بهاباد استفاده شده است. همچنین برای محاسبات و کد نویسی از نرم‌افزار R بهره گرفته شده است.

## فرآیند اتورگرسیو میانگین متحرک تلفیق شده (ARIMA)

مدل ARIMA اولین بار توسط باکس-جنکینز در سال ۱۹۷۶ معرفی شد [۱۴]. که می‌توان از آن برای پیش‌بینی داده‌های ایستا غیرفصلی استفاده کرد، که بصورت  $ARIMA(p,d,q)$  است. الگوی کلی ARIMA بصورت زیر خواهد بود:

$$\varphi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)a_t$$

که در آن عملگر AR ایستا  $\varphi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$  و عملگر MA وارون‌پذیر  $\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$  می‌باشد.

## فرآیند اتورگرسیو میانگین متحرک تلفیق شده فصلی (SARIMA)

الگوی فصلی ARIMA معروف باکس-جنکینز بصورت روابط زیر تعریف می‌شود:

$$ARIMA: \varphi_p(B)(1-B)^d Z_t = \theta_q(B)a_t$$

$$\phi_p(B^s) = 1 - \phi_1 B^s - \phi_2 B^{2s} - \dots - \phi_p B^{ps}$$

$$\vartheta_q(B^s) = 1 - \vartheta_1 B^s - \vartheta_2 B^{2s} - \dots - \vartheta_q B^{qs}$$

$a_t$  یک فرآیند تصادفی محض با میانگین صفر و  $\varphi_p(B)$  و  $\theta_q(B)$  را به ترتیب عوامل اتورگرسیو و میانگین متحرک و  $\phi_p(B^s)$  و  $\vartheta_q(B^s)$  را عوامل اتورگرسیو و میانگین متحرک فصلی می‌نامند و بصورت  $ARIMA(p,d,q)(P,D,Q)_s$  نشان می‌دهند که اندیس  $s$  مربوط به دوره فصلی است.

## روش هموارسازی نمایی (ES)

روش هموارسازی نمایی به جای این که یک میانگین ساده را حساب کند میانگین وزنی مقادیر گذشته را می‌گیرد بطوری که وزن‌ها بصورت نمایی متمایل به داده‌های گذشته کاهش می‌یابد. روش هموارسازی نمایی برای سری‌های فاقد روند و فصلی می‌باشد. پیش‌بینی در زمان  $t+1$  بصورت رابطه زیر می‌باشد:

$$F_{(t+1)} = \alpha y_t + \alpha(1-\alpha)y_{(t-1)} + \alpha(1-\alpha)^2 y_{(t-2)} + \dots$$

در رابطه بالا  $\alpha$  ثابت هموارسازی و یک مقدار ثابت بین صفر و یک است.

## روش هولت-وینترز (HW)

روش هولت-وینترز توسعه‌ای از هموارسازی نمایی می‌باشد. مدل هولت-وینترز، شامل سه بخش میانگین، روند و تغییرات فصلی است. پیش‌بینی توسط ترکیب این سه مولفه  $(\alpha, \beta, \gamma)$ ، صورت می‌پذیرد. شکل ساده مدل به صورت زیر است:

$$S_t = \alpha \frac{y_t}{I_{t-1}} + (1-\alpha)(S_{t-1} + S b_{t-1})$$

که در آن  $y_t$  مشاهدات در زمان  $t$  و  $S_t$  مشاهده هموار شده در زمان  $t$  است. از طرفی  $I$  نیز شاخص فصلی و  $L$  نیز طول دوره فصلی است. اگر مدل دارای روند باشد به صورت رابطه زیر خواهد بود.

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)(b_{t-1})$$

پارامتر  $\gamma$  نیز در مدل اضافه شده است. در نهایت مدل فصلی به صورت زیر خواهد بود:

$$I_t = \beta \frac{y_t}{S_t} + (1-\beta)I_{t-L}$$

پیش‌بینی در مدل هولت-وینترز، بصورت زیر است.

$$F_{t+m} = (S_t + m b_t) I_{t-L+m}$$

## شبکه عصبی اتورگرسیو (ARNN)

در مدل شبکه عصبی اتورگرسیو پیش‌بینی در دو فاز صورت می‌گیرد. اول، مرتبه مدل اتورگرسیو تعیین و در مرحله بعد، شبکه عصبی با در نظر گرفتن مرتبه اتورگرسیو آموزش داده می‌شود. در این روش، مدل با الگوی غیرفصلی از دو جزء  $p$  و  $k$  تشکیل شده که  $p$  تعداد لگ‌های ورودی و  $k$  تعداد نوروهای پنهان است. و بصورت  $ARNN(p,k)$  نشان داده می‌شود.

## رگرسیون KNN

KNN مخفف K-Nearest Neighbors است. به طور خلاصه، KNN مجموعه‌ای از نمونه‌های آموزشی را که با  $n$  ویژگی توصیف شده‌اند ذخیره می‌کند. هر نمونه آموزشی نشان‌دهنده یک نقطه در فضای  $n$  بعدی است. با توجه به یک نمونه جدید، KNN نزدیک‌ترین  $k$  نمونه‌های خود را در فضای  $n$  بعدی پیدا می‌کند که

فرض کنید یک سری زمانی تک متغیره، با روند غیرخطی داریم. ما علاقه‌مند به پیش‌بینی سری با برون‌یابی روند با یک تابع خطی برآورد شده از سری زمانی مشاهده شده هستیم. برای سری‌های زمانی با فواصل مساوی، یک هموارسازی اسپلاین مکعبی را می‌توان به عنوان تابع  $f(t)$  با پارامتر  $\lambda$  تعریف کرد که پارامتر هموارسازی  $\lambda$  نرخ تغییرات بین خطای باقی‌مانده و تغییرات محلی، را کنترل می‌کند. مقادیر بزرگ  $\lambda$  تابع  $f(t)$  را نزدیک به یک خط صاف و مقادیر کوچک  $\lambda$  تابع  $f(t)$  را بسیار پرنحنا نشان می‌دهد. برای اطلاعات بیشتر به رفرنس [۱۸] مراجعه شود.

### روش BATS

شناسه BATS مخفف ویژگی‌های کلیدی مدل است: تبدیل Box-Cox، خطاهای ARMA، روند و فصلی. برای نشان دادن پارامتر Box-Cox، پارامتر تعدیل، پارامترهای ARMA یعنی  $p$  و  $q$ ، و دوره‌های فصلی  $(m_1, \dots, m_T)$ ، با آرگومان‌هایی  $(\omega, \varphi, p, q, m_1, m_2, \dots, m_T)$  استفاده می‌شود. مدل BATS برای سری‌هایی که دارای چندین دوره فصلی است، بکار می‌رود. برای اطلاعات بیشتر به رفرنس (۱۹) مراجعه شود.

### معیارهای ارزیابی

میانگین مربعات خطا (MSE)

$$MSE(Y, \hat{Y}) = 1/n \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = 1/n \sum_{i=1}^n e_i^2$$

جذر میانگین مربعات خطا (RMSE)

$$RMSE(Y, \hat{Y}) = \sqrt{MSE(Y, \hat{Y})} = \sqrt{1/n \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$$

میانگین قدر مطلق خطا (MAE)

$$MAE(Y, \hat{Y}) = 1/n \sum_{i=1}^n |Y_i - \hat{Y}_i| = 1/n \sum_{i=1}^n |e_i|$$

میانگین درصد قدر مطلق خطا (MAPE)

$$MAPE(Y, \hat{Y}) = 100/n \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| = 100/n \sum_{i=1}^n \left| \frac{e_i}{Y_i} \right|$$

اهداف آن‌ها مشابه هدف ناشناخته‌اش باشد. حال، KNN می‌تواند برای پیش‌بینی سری‌های زمانی اعمال شود. در این مورد، هدف مرتبط با یک نمونه آموزشی مجموعه‌ای از مقادیر سری زمانی است و ویژگی‌هایی که نمونه را توصیف می‌کنند مقادیر عقب مانده هدف هستند. یعنی ما یک مدل اتورگرسیو داریم. بطور مثال اگر  $k = 2$  باشد،  $2 -$  نزدیک‌ترین همسایگان نمونه جدید پیدا و اهداف آن‌ها برای پیش‌بینی ماه آینده تجمیع می‌شود. برای اطلاعات بیشتر به رفرنس (۱۵) مراجعه شود.

### مدلسازی تتا (Theta)

این مدل بر اساس مفهوم اصلاح انحنای محلی است. این تغییر از ضریبی به نام تتا بدست می‌آید. اگر مقدار تتا به تدریج کاهش، سری زمانی نیز کاهش می‌یابد. اگر ضریب تتا افزایش یابد ( $\theta < 1$ ) آنگاه سری زمانی بازتر می‌شود. فرمول کلی تتا بشرح زیر می‌باشد.

$$Data = \frac{1}{2}(L(\theta = 0) + L(\theta = 2))$$

که در آن  $L(\theta = 0)$  مخفف خط تتا برابر با صفر که رگرسیون خطی داده‌ها است و دومی دقیقاً دو برابر سری زمانی اولیه می‌باشد. اولین خط تتا به روش معمول برای یک روند خطی و دومی از طریق هموارسازی نمایی ساده برون‌یابی می‌شود. با ترکیب این دو پیش‌بینی، پیش‌بینی نهایی مدل تتا به دست می‌آید (۱۶).

### روش STL

STL یک روش برای تجزیه سری زمانی به اجزای روند، فصلی و باقیمانده است. STL شامل دنباله‌ای از عملیات هموارسازی است که از رگرسیون وزن محلی (Loess) استفاده می‌کند. STL شش پارامتر دارد،  $n(p)$ : تعداد مشاهدات در هر چرخه فصلی /  $n(i)$ : تعداد عبور از حلقه داخلی /  $n(o)$ : تعداد تکرارهای استوارحلقه بیرونی /  $n(l)$ : پارامتر هموارسازی برای فیلتر پایین‌گذر /  $n(t)$ : پارامتر هموارسازی برای مؤلفه روند /  $n(s)$ : پارامتر هموارسازی برای مؤلفه فصلی. برای اطلاعات بیشتر به رفرنس (۱۷) مراجعه شود.

### روش هموارسازی اسپلاین مکعبی (CSS)

برای پیش‌بینی موارد تجمعی فوت کووید-۱۹ در شهرستان بهاباد توصیه می‌شود. (جدول ۱)

برای بررسی بهتر عملکرد این ۹ روش شکل‌های ۲ تا ۵ ترسیم شده‌اند. بر اساس بررسی شکل‌های ۲ و ۳، به نظر می‌رسد که مدل‌های BATS و هموارسازی اسپلاین مکعبی برای پیش‌بینی موارد جدید بستری کووید-۱۹ دارای خطای کمتری هستند به عبارتی این مدل‌ها را برای پیش‌بینی موارد تجمعی بستری شهرستان بهاباد در یک سطح می‌توان در نظر گرفت که از این میان مدل هموارسازی اسپلاین مکعبی برای پیش‌بینی موارد تجمعی بستری کووید-۱۹ دارای کمترین خطا می‌باشد. همچنین با بررسی و تحلیل شکل‌های ۴ و ۵ به این نتیجه رسیدیم که روش‌های رگرسیون KNN و روش تتا دارای خطاهای کمتری نسبت به بقیه روش‌ها هستند که روش رگرسیون KNN دارای کمترین خطا برای پیش‌بینی موارد تجمعی فوت می‌باشند. همچنین مدل شبکه‌های عصبی اتورگرسیو و مدل تتا برای موارد بستری و برای موارد فوت مدل شبکه‌های عصبی اتورگرسیو دارای بدترین عملکرد را در میان دیگر مدل‌ها از خود نشان دادند. (اشکال ۱ - ۴)

### بحث و نتیجه‌گیری

مهم‌ترین نکته‌ای که می‌توان از نتایج تحلیل برداشت کرد، این است که با استفاده از مدل‌هایی که در سال‌های خیلی قبل معرفی شده‌اند و یا به اصطلاح مدل‌های قدیمی می‌توان در عمل نتایج بهتری کسب نمود، درحالی‌که نتایج به دست آمده از مدل‌های جدیدتر مثل مدل شبکه‌های عصبی اتورگرسیو نتایج بدتری در پیش‌بینی کووید-۱۹ دارد. بنابراین می‌توان نتیجه گرفت که استفاده از مدل‌های جدید در عمل همیشه جوابگو نیست و با توجه به نوع داده‌ها ممکن است مدل‌های قدیمی نتیجه و کارایی بهتری در پیش‌بینی داشته باشند. با توجه به ناشناخته بودن رفتار این ویروس و پژوهش‌های صورت گرفته در پیش‌بینی کووید-۱۹ مهم‌ترین جنبه‌ی نوآوری این مطالعه

که در آن مقدار پیش‌بینی  $\hat{Y}_i$  و مقدار واقعی  $Y_i$  می‌باشد.

### یافته‌ها

در این بخش نتایج برازش مدل و پیش‌بینی مقادیر تجمعی کرونا (مقادیر تجمعی بستری و فوت) را برای شهرستان بهاباد، با توجه به مدل‌های مذکور آورده شده است. داده‌های کرونا مورد استفاده در این مقاله، از تاریخ ۱۳۹۸/۱۲/۰۷ الی ۱۴۰۰/۰۹/۲۸ به صورت روزانه می‌باشد، که از معاونت بهداشت و درمان استان یزد اخذ شده است. لازم به ذکر است که آمار موجود برای افرادی که در بیمارستان شهرستان بهاباد پذیرش و بستری شده‌اند، می‌باشد و افرادی که در مراکز درمانی سرپائی مورد پذیرش قرار گرفته باشند، را شامل نمی‌شود. در این پژوهش ۱۴ داده آخر را به عنوان داده‌های تست در نظر گرفته شد و با الباقی داده‌ها و مدل‌های مذکور، برای ۱۴ روز بعد، موارد تجمعی کووید-۱۹ (بستری و فوت) شهرستان بهاباد پیش‌بینی شد. و در نهایت بهترین مدل با توجه به معیارهای ارزیابی انتخاب شد. لازم به ذکر است تاریخی که اولین مورد اعم از بستری یا فوت گزارش شده را روز شروع سری زمانی در نظر گرفته شده است. برای توسعه مدل‌ها، فرض شده است که داده‌ها حاوی داده‌های پرت نیستند، داده‌ها تک متغیری هستند، داده‌ها ایستا هستند و پارامترهای مدل‌ها و خطاها با توجه به گذشت زمان ثابت‌اند.

با توجه به تجزیه و تحلیل داده‌ها و نتایج بدست آمده مطابق جدول ۱، بهترین مدل برای موارد تجمعی بستری کووید-۱۹ شهرستان بهاباد با توجه به معیارهای ارزیابی، مدل هموارسازی اسپلاین مکعبی می‌باشد که برای آینده از این روش برای پیش‌بینی موارد تجمعی بستری کووید-۱۹ در شهرستان بهاباد پیشنهاد می‌شود. بهترین روش و مدل برای موارد تجمعی فوت کووید-۱۹ شهرستان بهاباد با توجه به معیارهای ارزیابی یا همان خطاهای پیش‌بینی، مدل رگرسیون KNN می‌باشد، که برای آینده از این روش

کووید-۱۹ در شهرستان بهاباد استان یزد، استفاده شد و با معیارهای ارزیابی  $MAE$ ،  $RMSE$ ،  $MSE$  و  $MAPE$  با یکدیگر مقایسه شدند. نتایج نشان داد که بهترین مدل برای پیش‌بینی موارد تجمعی بستری کووید-۱۹ مدل هموارسازی اسپلاین مکعبی و موارد تجمعی مرگ‌ومیر مدل رگرسیون  $KNN$  با توجه به معیارهای ارزیابی یا همان خطاهای پیش‌بینی است. با توجه به نتایج این مطالعه به محققین آتی پیشنهاد می‌شود: پیش‌بینی کووید-۱۹ با ترکیب روش‌ها مثل روش تتا و  $STL$ ، پیش‌بینی با توجه به اثر واکسیناسیون علیه کووید-۱۹، از روش‌های استفاده شده، برای دیگر بیماری‌های واگیردار یا پاندمیک برای پیش‌بینی استفاده شود. و در نهایت می‌توان این روش‌ها را در مکانی دیگر مدل‌سازی و نتیجه‌گیری کرد.

**ملاحظات اخلاقی:** این مطالعه برگرفته از پایان نامه کارشناسی‌ارشد با کد رهگیری ۱۶۴۳۳۴۱۰ دانشگاه یزد، با کد اخلاق IR.YAZD.REC.1401.094 است، و از سوی هیچ سازمانی مورد حمایت مالی قرار نگرفته است.

استفاده از ۹ روش متفاوت در زمینه‌های مختلف برای موارد جدید بستری و مرگ‌ومیر کرونا شهرستان بهاباد، برخلاف پژوهش‌های انجام شده که از ۲ الی ۳ روش در زمینه خاصی بهره گرفته‌اند، می‌باشد. در این میان نکته‌ای که حائز اهمیت می‌باشد این است که باید داده‌ها در زمان واقعی بروز شوند، بدین معنی که با توجه به شرایط جدید جامعه (مثلاً واکسیناسیون، ظهور سویه‌های جدید) باید مدل‌ها بروز شوند.

مطالعه حاضر محدودیت‌هایی را به همراه داشت اول اینکه داده‌ها به دلیل محدودیت فقط شامل افرادی می‌باشد که در بیمارستان بستری شده‌اند و افرادی را که در مراکز درمانی سرپائی پذیرش شده باشند را شامل نمی‌شود. و به دلیل اینکه عوامل بسیاری در شیوع بیماری‌های همه‌گیر مثل کووید-۱۹ دخیل است دارای پیچیدگی‌های زیادی می‌باشد مانند ظهور سویه‌های جدید، که دومین محدودیت مطالعه حاضر است.

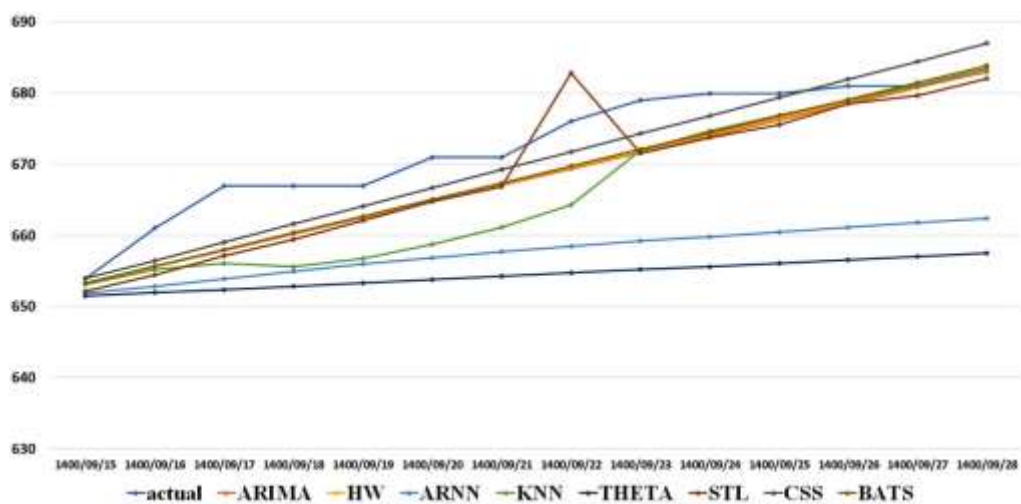
در مطالعه حاضر، از ۹ روش  $ARIMA$ ، تتا، هموارسازی نمائی، هولت-وینترز، شبکه عصبی اتورگرسیون، رگرسیون  $KNN$ ، روش  $BATS$ ، روش هموارسازی اسپلاین مکعبی و روش  $STL$  برای پیش‌بینی تعداد موارد تجمعی جدید و مرگ‌ومیر

جدول ۱ - معیارهای ارزیابی موارد تجمعی بستری و مرگ‌ومیر کووید-۱۹ در بهاباد

Table 1: Evaluation criteria of cumulative cases of hospitalization and death of Covid-19 in Bahabad

MSE	RMSE	MAE	MAPE	Model
<b>cumulative cases of hospitalization</b>				
27.1441	5.2100	4.4588	0.6636	ARIMA
27.7539	5.2682	4.5175	0.6723	ETS
26.3692	5.1351	4.4028	0.6553	HW
254.4886	15.9527	15.0418	2.2264	ARNN
61.6413	7.8512	6.4987	.9683	KNN
376.1039	19.3934	18.2220	2.6967	THETA

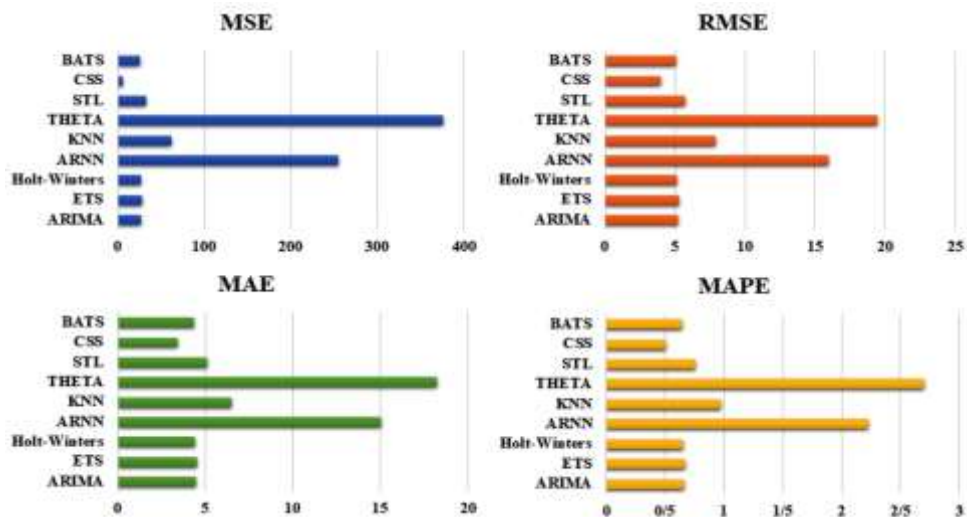
32.3840	5.6907	5.0726	0.7551	STL
15.8276	3.9784	3.4220	0.5088	CSS
25.2325	5.0232	4.3303	0.6445	BATS
<b>cumulative cases of death</b>				
0.9189	.9586	0.8157	1.2176	ARIMA
1.1524	1.0735	0.9463	1.4125	ETS
1.1763	1.0846	0.9561	1.4271	HW
12.7920	3.5766	3.4038	5.0803	ARNN
0.1304	0.3612	0.2708	0.4041	KNN
0.1868	0.4323	0.3808	0.5684	THETA
1.0351	1.0174	0.8801	1.3137	STL
2.3663	1.5383	1.3153	1.9631	CSS
0.7736	0.8796	0.7806	1.1651	BATS



شکل ۱ - نمودار پیش‌بینی ۱۴ روزه میزان موارد تجمعی بستری کووید-۱۹ در شهرستان بهاباد

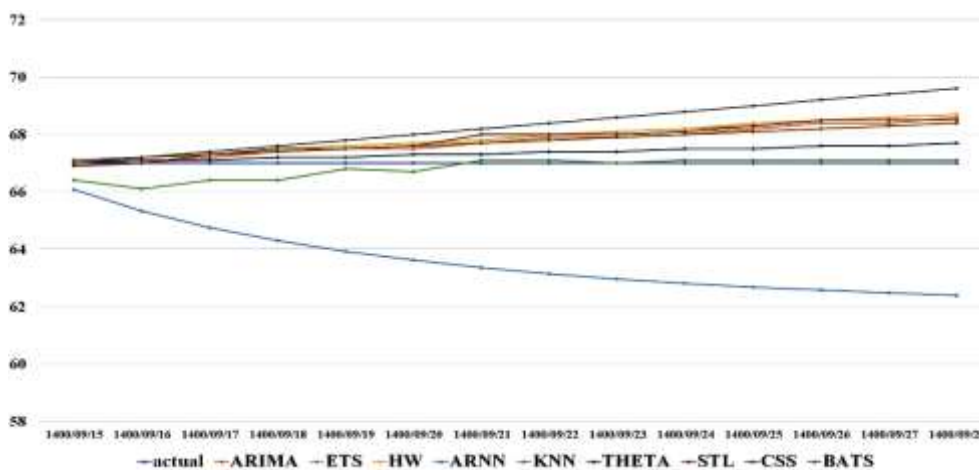
Figure 1 - 14-day prediction chart of the cumulative number of hospitalization cases of Covid-19 in Bahabad city





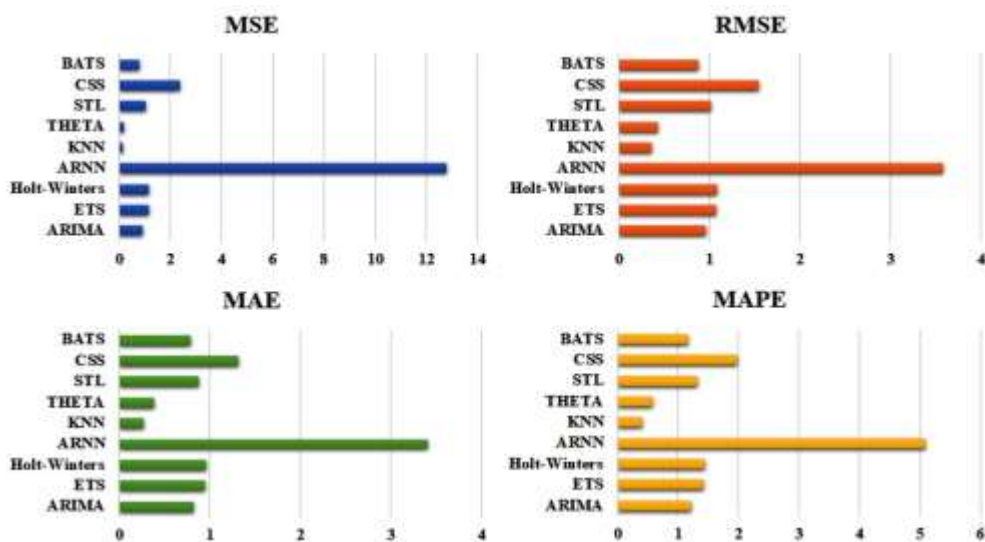
شکل ۲ - نمودار معیارهای ارزیابی موارد تجمعی بستری کووید-۱۹ در شهرستان بهاباد

Figure 2 - Chart of evaluation criteria for cumulative cases of hospitalization of Covid-19 in Bahabad city



شکل ۳ - نمودار پیش‌بینی ۱۴ روزه میزان موارد تجمعی مرگ‌ومیر کووید-۱۹ در شهرستان بهاباد

Figure 3 - 14-day prediction chart of the cumulative number of deathscases of Covid-19 in Bahabad city



شکل ۴ - نمودار معیارهای ارزیابی موارد تجمعی بستری کووید-۱۹ در شهرستان بهاباد

Figure 4 - Chart of evaluation criteria for cumulative cases of deaths of Covid-19 in Bahabad city

**Reference:**

- 1- Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. *Acta bio medica: Ateneiparmensis*, 2020; 91(1): 157.
- 2- Cheng ZJ, Shan J. 2019 Novel coronavirus: where we are and what we know. *Infection*, 2020 Apr; 48: 155-63.
- 3- Zhang X, Liu Y, Yang M, Zhang T, Young AA, Li X. Comparative study of four time series methods in forecasting typhoid fever incidence in China. *PloS one*, 2013 May 1; 8(5): e63116.
- 4- Olanrewaju SO, Ojo EO, Oguntade ES. Time Series Analysis on Reported Cases of Tuberculosis in Minna Niger State Nigeria. *Open Journal of Statistics*, 2020 May 8; 10(3): 412-30.
- 5- Nayak MS, Narayan KA. Forecasting dengue fever incidence using ARIMA analysis. *International Journal of Collaborative Research on Internal Medicine & Public Health*, 2019; 11(6): 924-32.
- 6- Wu W, An SY, Guan P, Huang DS, Zhou BS. Time series analysis of human brucellosis in mainland China by using Elman and Jordan recurrent neural networks. *BMC infectious diseases*, 2019 Dec; 19(1): 1-1.
- 7- Golkhandan A, Sahraei S. The Prediction of Iran's Per Capita Health Expenditures up to 2041 Horizon Using the Genetic and Particle Swarm Optimization Algorithms. *Journal of Healthcare Management*, 2019; 9(4): 66-53. [In Persian]
- 8- Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Science of The Total Environment*, 2020 Aug 10; 729: 138817.
- 9- Bayyurt L, Bayyurt B. Forecasting of COVID-19 cases and deaths using ARIMA models. *Medrxiv*, 2020 Apr 22: 2020-04.
- 10- Khan FM, Gupta R. ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. *Journal of Safety Science and Resilience*, 2020 Sep 1; 1(1): 12-8.
- 11- Alassafi MO, Jarrah M, Alotaibi R. Time series predicting of COVID-19 based on deep learning. *Neurocomputing*, 2022 Jan 11; 468: 335-44.
- 12- Mukhairez HH, Alaff AJ. Short-term Forecasting of COVID-19. *Computational Intelligence for COVID-19 and Future Pandemics: Emerging Applications and Strategies*, 2022: 257-66.22.
- 13- Li C, Sampene AK, Agyeman FO, Robert B, Ayisi AL. Forecasting the severity of COVID-19 pandemic amidst the emerging SARS-CoV-2 variants: adoption of ARIMA model. *Computational and Mathematical Methods in Medicine*, 2022 Jan 13; 2022.
- 14- Kufel T. ARIMA-based forecasting of the dynamics of confirmed Covid-19 cases for selected European countries. *Equilibrium. Quarterly Journal of Economics and Economic Policy*. 2020; 15(2): 181-204.
- 15- Martínez F, Frías MP, Charte F, Rivera AJ. Time Series Forecasting with KNN in R: the tsfknn Package. *R J*. 2019 Dec 1; 11(2): 229.
- 16- Nikolopoulos K, Assimakopoulos V, Bougioukos N, Litsa A, Petropoulos F. The theta model: An essential forecasting tool for supply chain planning. In *Advances in Automation and Robotics*, Vol. 2: Selected Papers from the 2011 International Conference on Automation and Robotics (ICAR 2011), Dubai, December 1–2; 2011: 431-437. Springer Berlin Heidelberg.
- 17- Wen Q, Gao J, Song X, Sun L, Xu H, Zhu S. RobustSTL: A robust seasonal-trend decomposition algorithm for long time series. In *Proceedings of the AAAI Conference on Artificial*

Intelligence, 2019 Jul 17; 33(01): 5409-5416.

18- Wang Y. Smoothing splines: methods and applications. CRC press; 2011 Jun 22.

19- De Livera AM, Hyndman RJ, Snyder RD. Forecasting time series with complex seasonal patterns using exponential smoothing. Journal of the American statistical association, 2011 Dec 1; 106(496): 1513-27.