



Semnan University

Journal of Modeling in Engineering

Journal homepage: <https://modelling.semnan.ac.ir/>

ISSN: 2783-2538



Research Article

Modeling Cumulative Cases of Covid-19 in Yazd City Using Various Time Series Techniques and Machine Learning and Comparing Their Efficiency

Mohammad Hossein Karimizarchi ^a, Davood Shishebori ^{b,*} 

^a M.Sc. degree, Department of Industrial Engineering, Faculty of Industrial Engineering, Yazd University, Yazd, Iran

^b Associate Professor, Department of Industrial Engineering, Faculty of Industrial Engineering, Yazd University, Yazd, Iran

PAPER INFO

Paper history:

Received: 09 February 2023

Revised: 12 July 2023

Accepted: 03 September 2023

Keywords:

Covid-19,
Time series,
Forecasting,
Statistical modeling.

ABSTRACT

Coronavirus disease 2019 or Covid-19, which is also called acute respiratory disease NCAV-2019 or commonly called corona, 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 model and predict new cases and deaths efficiently in the future. Nine popular forecasting techniques are tested on the data of Covid-19 in Yazd 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. According to the selected evaluation criteria, the results of the comprehensive analysis emphasize that the most efficient models are the ARIMA model for predicting the cumulative cases of hospitalization of Covid-19 and the Theta model for the cumulative cases of death. Also, in terms of evaluation criteria, the autoregressive neural network model has the worst performance in cumulative prediction of hospitalization and death cases. This study can provide a proper understanding of the spread of the 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, in this study, nine different techniques are used and their comparison, which in turn increases the comprehensiveness of the investigation and the assurance of the efficiency of the approach used in decision-making.

DOI: <https://doi.org/10.22075/jme.2023.29849.2404>

© 2023 Published by Semnan University Press.

This is an open access article under the CC-BY 4.0 license. (<https://creativecommons.org/licenses/by/4.0/>)

* Corresponding author.

E-mail address: shishebori@yazd.ac.ir

How to cite this article:

Karimizarchi, M., & Shishebori, D. (2023). Modeling cumulative cases of Covid-19 in Yazd city using various time series techniques and machine learning and comparing their efficiency. *Journal of Modeling in Engineering*, 21(75), 309-323. doi: 10.22075/jme.2023.29849.2404

مدلسازی موارد تجمعی کووید-۱۹ شهرستان یزد با استفاده از انواع تکنیک‌های سری زمانی و یادگیری ماشین و مقایسه کارایی آن‌ها

محمدحسین کریمی زارچی^۱، داود شیشه‌بری^{۲*}

| اطلاعات مقاله | چکیده |
|--|--|
| دریافت مقاله: ۱۴۰۱/۱۱/۰۲ بازنگری مقاله: ۱۴۰۲/۰۴/۲۱ پذیرش مقاله: ۱۴۰۲/۰۶/۱۲ | |
| واژگان کلیدی: کووید-۱۹، سری زمانی، پیش‌بینی، مدلسازی آماری. | بیماری کووید-۱۹، یک بیماری تنفسی است که در اثر سندرم تنفسی حاد کرونا ویروس-۲ ایجاد می‌شود. پیش‌بینی تعداد موارد جدید و مرگ‌ومیر می‌تواند گام مفیدی در پیش‌بینی هزینه‌ها و امکانات مورد نیاز در آینده باشد. هدف از مطالعه حاضر، مدلسازی و پیش‌بینی موارد جدید و مرگ‌ومیر در آینده است. ۹ تکنیک پیش‌بینی بر روی داده‌های کووید-۱۹ شهرستان یزد به عنوان یک مطالعه موردی 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 قدر مطلق خطا (MAPE) مدل‌ها باهم مقایسه شدند نتایج تحلیل نشان داد، بهترین مدل با توجه به معیارهای ارزیابی مذکور برای پیش‌بینی موارد تجمعی بستری کووید-۱۹ مدل رگرسیون KNN و برای موارد تجمعی فوت مدل BATS می‌باشد. همچنین از نظر معیارهای ارزیابی، بدترین عملکرد در پیش‌بینی تجمعی موارد بستری و فوت، مدل شبکه‌های عصبی اتورگرسیو دارد. 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 همه‌گیری عبور کرد. همچنین برخلاف مطالعات دیگر، در مطالعه حاضر، از ۹ تکنیک متفاوت و مقایسه آن‌ها، استفاده می‌شود که به نوبه خود، جامعیت بررسی و اطمینان از کارایی رویکرد به کار گرفته شده در تصمیم‌گیری را بالا می‌برد. |
| | DOI: https://doi.org/10.22075/jme.2023.29849.2404 |
| | © 2023 Published by Semnan University Press. This is an open access article under the CC-BY 4.0 license. (https://creativecommons.org/licenses/by/4.0/) |

۱-مقدمه^۱

بیماری کرونا ویروس ۲۰۱۹ یا کووید-۱۹ که به آن بیماری تنفسی حاد NCOV-2019 یا به شکل عمومی به آن کرونا نیز می‌گویند، یک بیماری تنفسی است که در اثر سندرم تنفسی حاد کرونا ویروس-۲ ایجاد می‌شود، که این ویروس یک سویه از ویروس‌های کرونا می‌باشد. این ویروس اولین بار در دسامبر ۲۰۱۹ شناسایی شد. گروهی از بیماران مبتلا به این ویروس دارای شکل جدیدی از ذات الریه

ویروسی بودند که همگی سابقه بازدید از بازار تره‌بار در ووهان چین را داشتند [۱]. در ۱۱ مارس سال ۲۰۲۰، سازمان بهداشت جهانی شیوع ویروس کرونا را به عنوان بیماری ویروسی همه‌گیر جهانی اعلام کرد. از زمانی که شیوع این بیماری به عنوان بیماری همه‌گیر اعلام شد، بسیاری از کشورهای جهان به دلیل ابتلا به بیماری ویروس کووید-۱۹ به شدت تحت تأثیر قرار گرفتند و اقدامات پیشگیرانه متعددی از جمله قرنطینه، تست‌های سریع

* پست الکترونیک نویسنده مسئول: shishebori@yazd.ac.ir

۱. کارشناسی ارشد، دانشکده مهندسی صنایع، دانشگاه یزد

۲. دانشیار، دانشکده مهندسی صنایع، دانشگاه یزد

استناد به این مقاله:

بیماری‌های عفونی مانند سل [۶]، تب دانگ [۷] و تب مالت [۸] استفاده کرده‌اند.

سیلان [۹] از مدل‌های ARIMA برای پیش‌بینی روند اپیدمیولوژیک شیوع کووید-۱۹ در ایتالیا، اسپانیا و اروپا استفاده کرده است. در این پژوهش چندین مدل ARIMA با پارامترهای مختلف با استفاده از معیار ارزیابی درصد میانگین خطای مطلق (MAPE) برای هر کدام از کشورها انتخاب شد. لوتفی و بورکو [۱۰] برای پیش‌بینی تعداد موارد تأیید شده و مرگ‌ومیر کووید-۱۹ از مدل ARIMA در مرکز کنترل و پیشگیری اروپا استفاده کردند. تاندن و همکاران [۱۱] مدل ARIMA را برای پیش‌بینی موارد کووید-۱۹ در هند بکار بردند. این مطالعه نشان‌دهنده روند صعودی موارد کووید-۱۹ در روزهای آینده بود. تجزیه و تحلیل سری‌های زمانی نیز افزایش نمایی تعداد موارد را نشان می‌دهد. در این مطالعه فرض بر این بود که مدل‌های پیش‌بینی فعلی به دولت و پرسنل پزشکی کمک می‌کند تا برای شرایط آینده آماده باشند. پرون [۱۲] در این مقاله یک مدل اتورگرسیو میانگین متحرک تلفیق شده (ARIMA) برای پیش‌بینی روند اپیدمی طی دوره زمانی بعد از ۴ آوریل ۲۰۲۰ (بیش از ۴۰ روز)، با استفاده از داده‌های اپیدمیولوژیک ایتالیا در سطح ملی و منطقه‌ای استفاده کرده است. نتایج آن‌ها نشان داد که تعداد موارد کووید-۱۹ در توسکانی (ایتالیا) در پنجاه و پنجمین روز مسطح خواهد شد.

بعلاوه چندین مطالعه پیش‌بینی کوتاه مدت بیماری همه‌گیر کووید-۱۹ را با استفاده از مدل‌های یادگیری ماشین به غیر از مدل ARIMA و SARIMA ارائه نموده‌اند. غسال و همکاران [۱۳] از روش‌های رگرسیون خطی و رگرسیون خطی چندگانه برای پیش‌بینی تعداد تلفات در هند برای مدت کوتاه به مدت شش هفته استفاده کرده‌اند. نویسندگان گزارش داده‌اند که اگر اقدامات پیشگیرانه کووید-۱۹ بدون تغییر یا اجرا نشود، تلفات در هند دو برابر خواهد شد. پارات و چاکرابورتی [۱۴] برای پیش‌بینی موارد کووید-۱۹ در هند به مدت ۶۰ روز براساس داده‌های سری‌زمانی گزارش شده برای دوره زمانی ۱ مارس ۲۰۲۰ تا ۳۰ آوریل ۲۰۲۰، از رگرسیون بردار پشتیبان (SVR) استفاده کرده‌اند. نتایج آن‌ها نشان می‌دهد که مدل

تشخیص کرونا، ماسک زدن، خود قرنطینه کردن، رعایت فاصله اجتماعی توسط کشورها برای جلوگیری از شیوع بیماری همه‌گیر کووید-۱۹ در حال اجرا می‌باشد. علیرغم این اقدامات، کووید-۱۹ به دلایل مختلفی از جمله تراکم جمعیت، سبک زندگی، سفرهای جهانی، ظهور سویه‌های جدید از این ویروس و... به سرعت در حال انتشار می‌باشد و بر سلامت انسان و اقتصاد جهانی تأثیر گذاشته است. مدارکی که کووید-۱۹ از انسان به انسان دیگر منتقل می‌شود توسط مراکز کنترل و پیشگیری از بیماری^۲ تأیید شده است و همچنین گزارش شده که کووید-۱۹ با لمس سطوح، تماس نزدیک با اشیاء و هوای حاوی ذرات ویروسی قابل انتقال است [۲].

آماده‌سازی و کنترل شیوع بیماری کووید-۱۹ نیاز به برنامه‌ریزی و سیاست‌های دقیق دارد. مدلسازی، برآورد و پیش‌بینی شیوع ویروس‌ها و ویژگی‌های اپیدمیولوژیک از موارد مهم، در تهیه تجهیزات مورد نیاز برای مقابله با پیامدهای آن‌ها است. برای داشتن یک برنامه مناسب برای کووید-۱۹، پیش‌بینی موارد تأیید شده در آینده بسیار مهم و حیاتی است. بنابراین مدل‌سازی و پیش‌بینی تعداد موارد کووید-۱۹ تأیید شده، بهبود یافته و مرگ‌ومیر، نقش مهمی در برنامه‌ریزی و کنترل شیوع کووید-۱۹ در جهان دارد.

مدل‌های آماری مختلفی برای پیش‌بینی شیوع بیماری‌های عفونی در آینده نزدیک و همچنین پیش‌بینی تعداد موارد کووید-۱۹ در آینده استفاده می‌شود [۳]. ژانگ و همکاران [۴] برای پیش‌بینی تب حصبه از مدل سری زمانی اتورگرسیو میانگین متحرک تلفیق شده فصلی (SARIMA) و سه مدل مختلف برگرفته از شبکه‌های عصبی شامل: شبکه‌های عصبی پس انتشار (BPNN)، شبکه‌های عصبی تابع شعاعی (RBFNN) و شبکه‌های عصبی بازگشتی المان (ERNN) استفاده کردند. در مطالعه‌ای دیگر چن و همکاران [۵] با استفاده از مدل SARIMA شیوع آنفولانزا را در مناطق شهری و روستایی شنیانگ چین پیش‌بینی کرده‌اند. و در نهایت نتایج این پژوهش به عنوان مرجعی برای استراتژی‌های پیشگیری و کنترل آنفولانزا در آینده مورد استفاده قرار گرفت. همچنین پژوهشگران دیگری بطور مشابه از مدل‌های اتورگرسیو میانگین متحرک تلفیق شده (ARIMA) برای پیش‌بینی

^۲CDC(Centers for Disease Control and Prevention)

مورد ارزیابی قرار گرفت. در این مطالعه، پیش‌بینی‌ها را با داده‌هایی که تا ۳ دسامبر ۲۰۲۰ در دسترس بود، انجام شد. چایون و همکاران [۲۰] در مطالعه‌ای، مدل ARIMA را برای تحلیل زمانی جهانی کووید-۱۹ در بازه زمانی از ۲۲ ژانویه ۲۰۲۰ تا ۷ آوریل ۲۰۲۰ استفاده کردند.

در مقاله‌ای دیگر، نایر و همکاران [۲۱] موارد مرگ‌ومیرهای تجمعی کووید-۱۹ در هند را براساس مدل‌های آماری مختلف مانند سری زمانی، یادگیری ماشین و یادگیری گروهی پیش‌بینی کردند. مدل ARIMA و هموارسازی نمایی و Holt-Winters در سری‌های زمانی، رگرسیون بردار پشتیبان (SVR) و رگرسیون خطی (LR) در یادگیری ماشین (ML) و رگرسیون جنگل تصادفی در یادگیری گروهی (EL) برای پیش‌بینی‌ها اجرا شده‌اند. دقت مدل‌های برازش شده با استفاده از معیارهایی مانند مقدار R-squared، جذر میانگین مربعات خطا (RMSE)، میانگین مربعات خطا (MSE)، میانگین خطاهای مطلق (MAE) و میانگین درصد خطا مطلق (MAPE) ارزیابی می‌شوند.

به منظور پیش‌بینی تعداد افراد مبتلا و فوت شده در آینده، مخایرز و آلف [۲۲] از الگوریتم‌های پیش‌بینی کوتاه‌مدت برای پیش‌بینی تعداد مبتلایان و مرگ‌ومیرها در آینده کوتاه استفاده کردند. پیش‌بینی با استفاده از مدل ARIMA، Holt-Winters ترکیبی، و هموارسازی نمایی با دوره زمانی ۵۷ روز انجام شد. پس از ارزیابی الگوریتم‌های آزمایش‌شده با استفاده از MAPE، متوجه شدند که هموارسازی نمایی بهترین الگوریتم برای پیش‌بینی موارد بهبود یافته با میانگین درصد خطا مطلق ۲/۶۶ و برای پیش‌بینی موارد تأیید شده با میانگین درصد خطا مطلق برابر با ۱/۷۷ درصد بوده و Holt-Winters بهترین الگوریتم برای پیش‌بینی موارد مرگ‌ومیر با MAPE برابر با ۵/۳۳ درصد می‌باشد. در مطالعه پراجاپاتی و همکاران [۲۳] از مدل‌های پیش‌بینی برای موارد تجمعی هر کشور با مقایسه داده‌های گزارش شده و مقدار پیش‌بینی شده و سپس رتبه‌بندی الگوریتم‌ها (Prophet, Holt-Winters, LSTM, ARIMA, ARIMA-NARNN) براساس مقادیر RMSE، MAE و MAPE آن‌ها انجام می‌شود. ترکیب مدل‌های ARIMA و شبکه عصبی اتورگرسیون غیرخطی (NARNN) بهترین نتیجه را در بین مدل‌های انتخابی نسبت به رایج‌ترین روش‌های پیش‌بینی سری‌های

SVR دارای دقت حدودی ۹۷ درصدی در پیش‌بینی موارد تلفات، موارد بهبود یافته و موارد تأیید شده تجمعی است. آنتونی و نایمولی [۱۵] برای پیش‌بینی موارد کووید-۱۹ در ایتالیا براساس داده‌های سری‌زمانی گزارش شده با استفاده از مدل خودرگرسیون ناهمگن (HAR) استفاده کرده‌اند. نتایج آن‌ها نشان می‌دهد که مدل HAR دارای دقت بیشتری نسبت به مدل‌های ARIMA در پیش‌بینی موارد کووید-۱۹ است.

کیبریا و همکاران [۱۶] از مدل ARIMA برای تخمین تعداد روزانه موارد کووید-۱۹ در بنگلادش بر اساس داده‌های ۲۰ آوریل ۲۰۲۱ تا ۴ ژوئیه ۲۰۲۱ استفاده کرده‌اند. مدل ARIMA بهترین نتایج را در بین چند مدل اجرا شده نسبت به مدل اتورگرسیون (AR)، میانگین متحرک (MA)، اتورگرسیون میانگین متحرک (ARMA) نشان داد. خان و گوپتا [۱۷] از مدل سری زمانی تک متغیره برای پیش‌بینی تعداد موارد آلوده به کووید-۱۹ که می‌تواند در روزهای آینده در هند انتظار داشت، استفاده کرده‌اند. آن‌ها یک مدل ARIMA را بر روی داده‌های جمع‌آوری شده از ۳۱ ژانویه ۲۰۲۰ تا ۲۵ مارس ۲۰۲۰ پیاده سازی نمودند و آن را با استفاده از داده‌های جمع‌آوری شده از ۲۶ مارس ۲۰۲۰ تا ۴ آوریل ۲۰۲۰ مورد ارزیابی قرار دادند. آن‌ها این مدل را برای پیش‌بینی روزانه موارد کووید-۱۹ برای ۵۰ روز آینده بدون هیچگونه مداخله اضافی استفاده کردند. نتایج نشان‌دهنده، روند افزایشی در تعداد واقعی و پیش‌بینی شده موارد کووید-۱۹ با حدود ۱۵۰۰ مورد در روز، براساس داده‌های موجود تا ۴ آوریل ۲۰۲۰ بود. چاودوری و همکاران [۱۸] از مجموعه داده موج سوم کووید-۱۹ که شامل پرونده‌های تقریباً سه ماهه موارد تأیید شده می‌باشد، با استفاده از سیستم استنتاج عصبی فازی تطبیقی (ANFIS) و شبکه‌های حافظه کوتاه مدت (LSTM) برای پیش‌بینی موارد جدید کووید-۱۹ در بنگلادش استفاده کردند. نتایج هر دو روش را مورد مقایسه قرار دادند که LSTM نتایج رضایت بخش‌تری را نشان داد. الاسافی و همکاران [۱۹] از مجموعه داده‌های عمومی ارائه شده توسط مرکز پیشگیری و کنترل بیماری اروپا برای توسعه یک مدل پیش‌بینی برای گسترش شیوع کووید-۱۹ در مالزی، مراکش و عربستان سعودی استفاده کرده‌اند. برای این منظور از برخی از مدل‌های یادگیری عمیق (DL) برای پیش‌بینی روند شیوع کووید-۱۹ موجود در این سه کشور

یزد استفاده کرده‌ایم. برای داشتن یک پیش‌بینی معنی‌دار آماری از داده‌های سری زمانی، حداقل حجم نمونه ۳۰ مشاهده مورد نیاز است [۲۵]. تعداد مشاهدات (یعنی حجم نمونه) مورد استفاده در کار حاضر بسیار بیشتر از حداقل اندازه مورد نیاز برای انجام پیش‌بینی سری‌های زمانی معنی‌دار است، زیرا داده‌های جمع‌آوری شده برای مدت حدوداً بیست و دو ماه می‌باشد.

داده‌های سری زمانی، دنباله‌ای از مقادیر عددی است که دارای یک لیبل زمانی مرتبط با هر مقدار است [۲۶]. بنابراین سری زمانی مجموعه‌ای از مشاهدات است که بر حسب زمان (یا هر کمیت دیگر) مرتب شده باشند. و معمولاً آن را به صورت $X_{t1}, X_{t2}, \dots, X_{tn}$ نمایش می‌دهند. داده‌های سری زمانی را می‌توان به دو دسته داده‌های ایستا و داده‌های غیرایستا طبقه‌بندی کرد. داده‌های سری زمانی ایستا هیچ الگویی نسبت به زمان ندارند درحالی‌که داده‌های سری زمانی غیرایستا دارای الگوهایی هستند که به عنوان فصلی نیز شناخته می‌شوند. بنابراین میانگین و واریانس داده‌های غیرایستا در طول زمان ثابت نیست.

۲-۱- فرآیند اتورگرسیو میانگین متحرک تلفیق شده (ARIMA)

مدل $ARIMA(p,d,q)$ اولین بار توسط باکس و جنکینز در سال ۱۹۷۶ معرفی شد [۲۷] که می‌توان از آن برای پیش‌بینی داده‌های سری زمانی ایستا غیرفصلی استفاده کرد. سری‌های $ARIMA$ یا اتورگرسیو میانگین متحرک تلفیق شده در واقع سری‌های $ARMA$ همراه با روند هستند. بنابراین به جای آنکه به طریقی روند را حذف کنیم و از همان روش‌های موسوم درباره سری‌های $ARMA$ استفاده کنیم، یک‌باره از مدل $ARIMA$ استفاده می‌کنیم. بنابراین یک سری زمانی که پس از d مرتبه تفاضل‌گیری به سری زمانی $ARMA(p,q)$ تبدیل شود، به سری زمانی $ARIMA(p,d,q)$ معروف است.

الگوی کلی $ARIMA$ بصورت رابطه (۱) خواهد بود:

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

که در آن عملگر AR ایستا $\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ و عملگر MA وارون‌پذیر $\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ برای $d=0$ و $d>0$ نقش‌های بسیار متفاوتی را بازی می‌کند.

زمانی (ARIMA) داشت.

در مقاله‌ای دیگر که در سال ۲۰۲۲ انجام شد لی و همکاران [۲۴] کشورهای بریتانیا، آفریقای جنوبی، برزیل و نیجریه به عنوان نمونه پژوهشی انتخاب کردند. مجموعه داده‌های تجزیه و تحلیل شده در این مطالعه از ۱ مارس ۲۰۲۰ تا ۳۱ ژانویه ۲۰۲۱ را شامل می‌شود و از وب سایت سازمان بهداشت جهانی به دست آمده است. این مطالعه از مدل $ARIMA$ برای پیش‌بینی شیوع ویروس کرونا استفاده کرده است.

در مطالعه حاضر، قرار است تا موارد تجمعی کووید-۱۹ (موارد تأیید شده و مرگومیر) برای شهرستان یزد به عنوان یک مطالعه موردی از تاریخ ۱۳۹۸/۱۲/۰۷ الی ۱۴۰۰/۰۹/۲۸ (داده‌ها از معاونت بهداشت و درمان استان یزد اخذ شده است) با استفاده از تکنیک‌های سری‌های زمانی و یادگیری ماشین از جمله مدل‌های اتورگرسیو میانگین متحرک تلفیق شده ($ARIMA$)، هموارسازی نمایی (ETS)، Holt-Winters، رگرسیون KNN، تتا (Theta)، شبکه عصبی اتورگرسیو (ARNN)، هموارسازی اسپلاین مکعبی (CSS)، روش STL و روش BATS مدل‌سازی و براساس معیارهای میانگین خطای مطلق (MAE)، درصد میانگین خطای مطلق (MAPE)، خطای میانگین مربعات (MSE) و خطای جذر میانگین مربعات (RMSE) به عنوان معیار انتخاب بهترین مدل استفاده شود.

بقیه مقاله به شرح زیر سازماندهی شده است: بخش ۲ مدل‌های آماری، همراه با ابزارهای تحلیلی و معیارهای ارزیابی توصیف می‌کند. در بخش ۳، انتخاب پارامتر مدل و روش بهینه‌سازی پارامتر با در نظر گرفتن تجزیه و تحلیل سری زمانی روی موارد تأیید شده تجمعی و مرگومیر تجمعی در شهرستان یزد مورد بحث قرار گرفته است. و در نهایت، بخش ۴ نتایج حاصل از کار حاضر را ارائه می‌دهد. همچنین بخش ۵ مربوط به پیشنهادات آتی می‌باشد.

۲-۲- مدل‌های آماری

این مطالعه از مدل‌های آماری اتورگرسیو میانگین متحرک تلفیق شده ($ARIMA$)، هموارسازی نمایی (ETS)، Holt-Winters، رگرسیون KNN، تتا (Theta)، شبکه عصبی اتورگرسیو (ARNN)، هموارسازی اسپلاین مکعبی (CSS)، روش STL و روش BATS برای ایجاد یک پیش‌بینی ۱۴ روزه از موارد تجمعی کووید-۱۹ برای استان

می‌باشد:

$$F_{t+1} = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \dots \quad (۸)$$

در رابطه بالا α ثابت هموارسازی نامیده می‌شود و یک مقدار ثابت بین صفر و یک است. رابطه بالا هموارسازی نمائی را با استفاده از میانگین وزنی از تمام مشاهدات گذشته با وزن‌های نزولی نمائی نشان می‌دهد. ثابت هموارسازی α توسط کاربر تعیین می‌شود. مقادیر پیش‌فرضی که برای α در نظر گرفته شده و نشان‌دهنده‌ی عملکرد خوب آن می‌باشد، در محدوده ۰/۱ تا ۰/۲ است. همچنین برای انتخاب α روش آزمون و خطا نیز می‌تواند راهکار مناسبی باشد. به هر حال باید در انتخاب α بسیار دقت خرج داد زیرا در غیر اینصورت باعث برآزش بیش از حد مدل و پایین آمدن دقت پیش‌بینی مدل در طول دوره اعتبارسنجی و آینده می‌شود.

۲-۴-روش Holt-Winters

روش پیش‌بینی Holt-Winters توسعه‌ای از روش هموارسازی نمائی می‌باشد. مدل Holt-Winters شامل سه بخش است. بخش اول، میانگین (مقدار ثابت) نامیده می‌شود که رفتار کلی مدل را نشان می‌دهد و مقادیر پیرامون آن نوسان می‌کنند. بخش دوم روند (شیب خط) است که برحسب زمان ثابت است ولی به عنوان مضرب برای متغیر در نظر گرفته می‌شود. بخش سوم نیز که بصورت تناوبی تغییر می‌کند، برای نمایش تغییرات فصلی به کار گرفته می‌شود. در چنین مدلی، پیش‌بینی مقدار آینده توسط ترکیب این سه مولفه، صورت می‌پذیرد. این مدل دارای چندین پارامتر است، که با α, β, γ نشان داده می‌شوند. شکل ساده (بدون روند و تغییرات فصلی) مدل سری زمانی Holt-Winters به صورت رابطه (۹) است:

$$S_t = \alpha \frac{y_t}{I_{t-1}} + (1 - \alpha)(S_{t-1} + b_{t-1}) \quad (۹)$$

که در آن y_t مشاهدات مربوط به زمان t و S_t مشاهده هموار شده در زمان t است. همانطور که در این مدل مشاهده می‌کنید، فقط پارامتر α حضور داشته و هر مشاهده به صورت ترکیب خطی از نقطه قبلی (S_{t-1}) دیده می‌شود. از طرفی I نیز شاخص تغییرات فصلی^۳ نامیده شده و L نیز طول دوره تغییرات فصلی است. در صورتی که مدل دارای

۲-۲-فرآیند اتورگرسیو میانگین متحرک تلفیق شده

فصلی (SARIMA)

الگوی ARIMA فصلی جمعی معروف باکس و جنکینز بصورت روابط (۲) تا (۷) تعریف می‌شود:

$$SARIMA: \Phi_p(B^s)\varphi_p(B)(1-B)^d(1-B^s)^q Z_t = \theta_q(B)\theta_q(B^s)a_t \quad (۲)$$

$$B^s)^D Z_t = \theta_q(B)\theta_q(B^s)a_t$$

$$ARIMA: \varphi_p(B)(1-B)^d Z_t = \theta_q(B)a_t \quad (۳)$$

$$\varphi_p(B) = (1 - \varphi_1 B - \dots - \varphi_p B^p) \quad (۴)$$

$$\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q) \quad (۵)$$

$$\Phi_p(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_p B^{Ps} \quad (۶)$$

$$\Theta_q(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_q B^{Qs} \quad (۷)$$

این دو عبارت چند جمله‌ای‌هایی از B^s هستند که ریشه مشترک ندارند. ریشه‌های این چند جمله‌ای‌ها در خارج از دایره واحد واقع‌اند و a_t یک فرآیند تصادفی محض با میانگین صفر است. برای سهولت، $\varphi_p(B)$ و $\theta_q(B)$ را به ترتیب عوامل اتورگرسیو و میانگین متحرک و $\Phi_p(B^s)$ و $\Theta_q(B^s)$ را عوامل اتورگرسیو و میانگین متحرک فصلی می‌نامند و اغلب بصورت $ARIMA(p,d,q)(P,D,Q)_s$ نشان می‌دهند که در آن اندیس s دوره فصلی را نشان می‌دهد.

۲-۳-روش هموارسازی نمایی

یکی از روش‌های پیش‌بینی رایج و محبوب روش هموارسازی نمائی است. محبوبیت آن ناشی از انعطاف‌پذیری، سهولت اتوماسیون، محاسبات ساده، و عملکرد خوب است. روش هموارسازی نمائی به جای اینکه یک میانگین ساده را حساب کند میانگین وزنی مقادیر گذشته را می‌گیرد بطوریکه وزن‌ها بصورت نمائی متمایل به داده‌های گذشته کاهش می‌یابد بدین معنا که به داده‌های اخیر اهمیت بیشتری نسبت به داده‌های قدیمی‌تر می‌دهد. روش هموارسازی نمائی برای پیش‌بینی سری‌هایی که فاقد روند و الگوی فصلی می‌باشند، استفاده می‌شود. بنابراین ابتدا باید روند و فصلی بودن را از سری حذف کنیم و سپس این روش را برای سری باقیمانده‌ها بکار بگیریم. پیش‌بینی در زمان $t+1$ به روش هموارسازی نمائی بصورت رابطه (۸)

^۳Seasonal Index

فاصله استفاده می‌شود. به عنوان مثال، با فرض اینکه فاصله شباهت فاصله اقلیدسی است، فاصله بین نمونه جدید و نمونه آموزشی i به صورت رابطه (۱۳) محاسبه می‌شود:

$$\sqrt{\sum_{x=1}^n (f_x^i - q_x)^2} \quad (13)$$

k نمونه‌های آموزشی که نزدیک‌ترین نمونه به نمونه جدید هستند، k نزدیک‌ترین همسایه‌های آن‌ها در نظر گرفته می‌شوند. KNN مبتنی بر یادگیری از طریق قیاس است. با توجه به یک نمونه جدید، ما فکر می‌کنیم که اهداف نزدیک‌ترین همسایگان احتمالاً مشابه هدف ناشناخته این نمونه هستند. به این ترتیب، اهداف نزدیک‌ترین همسایه‌ها برای پیش‌بینی هدف نمونه جدید جمع می‌شوند. به عنوان مثال، با فرض اینکه اهداف یا k نزدیک‌ترین همسایه بردار t^1, \dots, t^k است می‌توان آن‌ها را برای پیش‌بینی هدف نمونه جدید به صورت میانگین محاسبه کرد:

$$\sum_{i=1}^k \frac{t^i}{k} \quad (14)$$

به طور خلاصه، KNN مجموعه‌ای از نمونه‌های آموزشی را که با n ویژگی توصیف شده‌اند پیدا 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 به این امید که اهداف آن‌ها مشابه هدف ناشناخته‌اش باشد. حال، بیایید ببینیم که چگونه KNN می‌تواند برای پیش‌بینی سری‌های زمانی استفاده 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. به عنوان مثال، ما با یک سری زمانی ماهانه شروع می‌کنیم که شامل ۱۳۲ مشاهده است، یعنی ۱۱ سال: $t = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, \dots, x_{132}\}$ و همچنین فرض می‌کنیم می‌خواهیم ماه آینده را پیش‌بینی کنیم. از آنجا که ما قصد داریم تنها یک مقدار را پیش‌بینی کنیم، هدف یک نمونه آموزشی، مقدار سری زمانی است. همچنین فرض می‌کنیم که ویژگی‌هایی که یک هدف را توصیف می‌کنند، دوازده مقدار اول تاخیر در سری‌های زمانی هستند که ما آن را با تاخیر ۱:۱۲ نشان می‌دهیم. بنابراین، نمونه‌های آموزشی یا مثال‌های مرتبط با سری زمانی t در جدول ۱ نشان داده شده است. حال، می‌خواهیم ببینیم که کدام

روند باشد شکل مدل به صورت رابطه (۱۰) خواهد بود.

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)(b_{t-1}) \quad (10)$$

همانطور که مشخص است پارامتر γ نیز در مدل اضافه شده است. در نهایت مدلی که به همراه تغییرات فصلی باشد، به صورت زیر خواهد بود:

$$I_t = \beta \frac{y_t}{S_t} + (1 - \beta)I_{t-L} \quad (11)$$

برای پیش‌بینی در مدل Holt-Winters، از رابطه (۱۲) استفاده می‌کنیم.

$$F_{t+m} = (S_t + mb_t)I_{t-L+m} \quad (12)$$

لازم به ذکر است که پارامترها در این مدل α, β, γ ثابت هموارسازی و بین عدد صفر و یک می‌باشند.

۲-۵- شبکه عصبی اتورگرسیو

در مدل شبکه عصبی اتورگرسیو پیش‌بینی در دو فاز صورت می‌گیرد. برای سری‌های زمانی مورد نظر، مرتبه مدل اتورگرسیو در فاز اول تعیین می‌شود. در مرحله دوم، شبکه عصبی توسط مجموعه داده آموزشی، با در نظر گرفتن مرتبه اتورگرسیو آموزش داده می‌شود. تعداد گره‌های ورودی یا تأخیرهای سری زمانی شبکه عصبی از مرتبه اتورگرسیو تعیین می‌شود. در این روش، مدل برازش شده با الگوی غیرفصلی از دو جزء p و k تشکیل شده است که p نشان‌دهنده تعداد لگ‌های ورودی و k نشان‌دهنده تعداد نورون‌های پنهان است. بنابراین این مدل بصورت ARNN(p,k) نشان داده می‌شود. همچنین مدل برازش شده با الگوی فصلی به صورت ARNN(p,P,k)_(m) می‌باشد که مشابه با ARIMA(p,0,0)(P,0,0)_(m) با توابع غیرخطی است.

۲-۶- رگرسیون KNN

KNN مخفف عبارت K-Nearest Neighbors به معنی K نزدیک‌ترین همسایگی است که نام این الگوریتم اشاره به شیوه کار آن دارد. رگرسیون KNN مجموعه‌ای از نمونه‌های آموزشی در خود دارد. نمونه آموزشی i ام از یک بردار با n ویژگی بصورت $f_1^i, f_2^i, \dots, f_n^i$ توصیف نمونه و یک بردار هدف با m ویژگی بصورت $t_1^i, t_2^i, \dots, t_m^i$ تشکیل شده است، با توجه به یک نمونه جدید، که ویژگی‌های آن بصورت q_1, \dots, q_n است ولی هدف این نمونه ناشناخته است، ویژگی‌های نمونه جدید برای یافتن k مشابه‌ترین نمونه‌های آموزشی با توجه به بردار ویژگی‌ها از شباهت یا

نمونه جدید برای پیش‌بینی مقدار آینده بعدی سری زمانی استفاده می‌شود. از آنجایی که ما از تاخیرهای ۱ تا ۱۲ به عنوان بردار ویژگی استفاده می‌کنیم، بردار ویژگی مرتبط با نقطه آینده بعدی، بردار $x_{121}, x_{122}, \dots, x_{132}$ است. که از

دوازده مقدار آخر سری زمانی تشکیل شده است. به عنوان مثال، اگر k برابر با ۲ باشد، ۲-نزدیک‌ترین همسایگان نمونه جدید پیدا می‌شوند و اهداف آنها برای پیش‌بینی ماه آینده تجمیع می‌شود [۲۸].

جدول ۱. نمونه‌های آموزشی برای سری زمانی t برای پیش‌بینی یک گام جلوتر و تاخیرهای ۱:۱۲ به عنوان بردار ویژگی

| ویژگی‌ها |
|--|
| $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}$ |
| $x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}$ |
| $x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}$ |
| ... |
| $x_{120}, x_{121}, x_{122}, x_{123}, x_{124}, x_{125}, x_{126}, x_{127}, x_{128}, x_{129}, x_{130}, x_{131}$ |

۷-۲-مدلسازی تتا

این مدل بر اساس مفهوم اصلاح انحناهای محلی سری‌های زمانی است. این تغییر از ضریبی به نام θ (به عنوان نماد از حرف یونانی θ استفاده می‌شود) به دست می‌آید که مستقیماً به تفاضل‌های دوم سری زمانی اعمال می‌شود:

$$X_{new}''(\theta) = \quad (15)$$

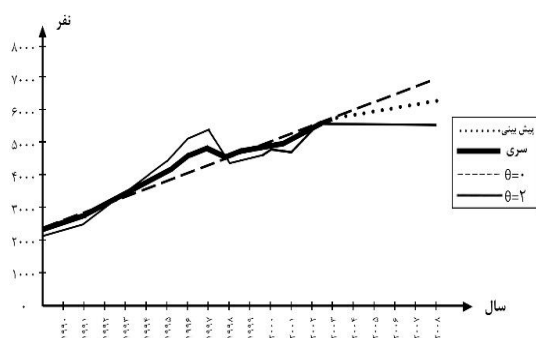
$$\theta \cdot X_{data}, \text{ where } X_{data}'' = X_t - 2X_{t-1} - X_{t-2} \text{ at time } t$$

اگر مقدار تتا به تدریج کاهش یابد، سری زمانی کاهش می‌یابد. هرچه مقدار ضریب تتا کوچکتر باشد، درجه قله‌ها کاهش پیدا می‌کند. در حالتی که $\theta=0$ است، سری زمانی به یک خط، رگرسیون خطی تبدیل می‌شود. ضریب θ نیز می‌تواند مقادیر منفی داشته باشد. برعکس، اگر ضریب تتا افزایش یابد ($\theta < 1$) آنگاه سری زمانی بازتر می‌شود. فرمول کلی روش تتا بشرح رابطه (۱۶) می‌باشد. سری زمانی اولیه به دو یا چند خط تتا یا سری زمانی جدید تجزیه می‌شود. هر یک از خطوط تتا به طور جداگانه برون‌یابی شده و پیش‌بینی‌ها به سادگی ترکیب می‌شوند. بطور مثال $\theta=0$ و $\theta=2$

$$Data = \frac{1}{2}(L(\theta = 0) + L(\theta = 2)) \quad (16)$$

که در آن $L(\theta = 0)$ مخفف خط تتا برای پارامتر θ برابر با صفر است. اولین خط تتا ($\theta = 0$) خط رگرسیون خطی داده‌ها است و دومی دقیقاً دو برابر سری زمانی اولیه می‌باشد. اولین جزء یعنی $L(\theta = 0)$ سری زمانی را از طریق یک روند خطی توصیف می‌کند. مورد دوم، $L(\theta = 2)$ ، انحناهای محلی را دو برابر کرده و رفتار کوتاه مدت را

بزرگ‌نمایی می‌کند. اولین خط تتا به روش معمول برای یک روند خطی برون‌یابی می‌شود. دومی از طریق هموارسازی نمایی ساده برون‌یابی می‌شود. و در نهایت ترکیب ساده این دو پیش‌بینی، پیش‌بینی نهایی مدل تتا برای سری‌های زمانی خاص، همانطور که در شکل (۱) نشان داده شده است، به دست می‌آید.



شکل ۱. نحوه انجام پیش‌بینی با روش تتا

در کل مراحل این روش بصورت زیر خواهد بود:
مرحله اول (بررسی فصلی بودن): ابتدا هر سری زمانی برای رفتار فصلی معنی‌دار آماری بررسی می‌شود. مرحله دوم (فصلی‌زدایی): سری‌های زمانی از طریق روش تجزیه کلاسیک (ضربی) غیر فصلی می‌شوند. مرحله سوم (تجزیه): هر سری زمانی به دو خط تتا تجزیه می‌شود. اولی به خط رگرسیون خطی با $\theta=0$ و دیگری به خط تتا با $\theta=2$. مرحله چهارم (برون‌یابی): خط رگرسیون خطی به روش معمول برون‌یابی می‌شود درحالی‌که خط دوم از طریق هموارسازی نمایی ساده برون‌یابی می‌شود. مرحله پنجم (ترکیب): پیش‌بینی‌های حاصل از برون‌یابی دو خط با وزن‌های مساوی ترکیب می‌شوند. نکته‌ای که حائز اهمیت می‌باشد، استفاده از ترکیب‌های مختلف خطوط تتا برای هر افق

$$\sum_{t=1}^n (y_t - f(t))^2 + \lambda \int_S (f''(u))^2 du \quad (17)$$

که بر روی تابع مشتق دوم f در S که در آن $[1, n] \subseteq S \subseteq \mathbb{R}$ به حداقل می‌رسد.

پارامتر هموارسازی λ نرخ تغییرات بین خطای باقی‌مانده و تغییرات محلی که توسط انتگرال مشتق دوم f به توان دو، نشان داده شده است، را کنترل می‌کند. مقادیر بزرگ λ تابع $f(t)$ را نزدیک به یک خط صاف نشان می‌دهد درحالی‌که مقادیر کوچک λ یک تابع $f(t)$ بسیار پر انحنای (پر پیچ و خم) را نشان می‌دهد.

پیش‌بینی‌ها معمولاً با استفاده از مدل‌هایی انجام می‌شوند که بیشترین وزن را به مشاهدات اخیر و وزن ناچیز به گذشته‌های دور می‌دهند. این بدان معنی است که پارامتر هموارسازی λ نباید برای اهداف پیش‌بینی خیلی بزرگ باشد. برخی از روش‌های پیش‌بینی خطی فرض می‌کنند که یک روند خطی زیربنایی وجود دارد. ما این فرض را نمی‌کنیم زیرا تابع پیش‌بینی ما خطی است، اما روند اصلی $f(t)$ مجاز است غیرخطی باشد. برای اطلاعات بیشتر در مورد این روش به رفرنس [۳۱] مراجعه شود.

۲-۱۰-۱ روش BATS

شناسه BATS مخفف ویژگی‌های کلیدی مدل است: تبدیل Box-Cox، خطاهای ARMA، روند و اجزای 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 $(\omega, \phi, p, q, m_1, m_2, \dots, m_T)$ تکمیل می‌شود.

مدل BATS واضح‌ترین تعمیم مدل‌های نوآوری فصلی سنتی است که برای سری‌های زمانی با دوره‌های فصلی متعدد به کار می‌رود. با این حال، این مدل نمی‌تواند دوره‌های فصلی غیر صحیح را در خود جای دهد، ولی می‌تواند تعداد بسیار زیادی حالت داشته باشد. مولفه فصلی اولیه به تنهایی حاوی m_T غیر صفر است. برای اطلاعات بیشتر در مورد این روش به رفرنس [۳۲] مراجعه شود.

۲-۱۱-۱ معیارهای ارزیابی

معیارهای ارزیابی بصورت زیر محاسبه می‌شوند:

میانگین مربعات خطا

$$MSE(Y, \hat{Y}) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (18)$$

جزر میانگین مربعات خطا

پیش‌بینی است. شواهد تجربی وجود دارد مبنی بر اینکه برای افق‌های طولانی‌تر، پیش‌بینی‌ها باید بیشتر به سمت رفتار بلندمدت گرایش داشته باشند، درحالی‌که برای پیش‌بینی‌های کوتاه‌مدت بیشتر باید روندهای اخیر را در نظر بگیریم. این را می‌توان به راحتی با استفاده از جفت‌های مختلف خطوط تتا برای هر افق پیش‌بینی انجام داد. به عنوان مثال، اگر از زوج $\theta=0$ و $\theta=1/5$ استفاده شود، تأکید بیشتری بر روند بلند مدت سری زمانی می‌شود، درحالی‌که در مورد خطوط تتا $\theta=0$ و $\theta=2/5$ ، رفتار کوتاه مدت اهمیت بیشتری پیدا می‌کند [۲۹].

۲-۸-۲ روش STL

STL یک روش تجزیه یک سری زمانی به اجزای روند، فصلی و باقیمانده است. یعنی فرض کنید داده‌ها، جزء روند، مؤلفه فصلی و مؤلفه باقیمانده به ترتیب با Y_v, T_v, S_v و R_v برای $v = 1, \dots, N$ نشان داده می‌شوند. داریم که:

$$Y_v = T_v + S_v + R_v$$

STL متشکل از دنباله‌ای از عملیات هموارسازی است که از رگرسیون وزن محلی یا Loess استفاده می‌کند. این روش بیشتر برای سری‌های زمانی پیچیده کاربرد دارد. STL چندین پارامتر دارد که باید توسط تحلیلگر داده انتخاب شود. STL ۶ پارامتر به شرح زیر دارد:

$n(p)$: تعداد مشاهدات در هر چرخه فصلی / $n(i)$: تعداد عبور از حلقه داخلی / $n(o)$: تعداد تکرارهای استوارحلقه بیرونی / $n(t)$: پارامتر هموارسازی برای فیلتر پایین-گذر / $n(i)$: پارامتر هموارسازی برای مؤلفه روند / $n(s)$: پارامتر هموارسازی برای مؤلفه فصلی.

انتخاب پنج مورد اول ساده است. با این حال، آخرین پارامتر، $n(s)$ ، باید به دقت برای هر برنامه تنظیم شود. برای اطلاعات بیشتر در مورد این روش به رفرنس [۳۰] مراجعه شود.

۲-۹-۲ روش هموارسازی اسپلاین مکعبی

فرض کنید یک سری زمانی تک متغیره $y_t, t = 1, \dots, n$ با روند غیرخطی مشاهده می‌کنیم. ما علاقه‌مند به پیش‌بینی سری با برون‌یابی روند با استفاده از یک تابع خطی برآورد شده از سری زمانی هستیم. برون‌یابی روند خطی بسیار مورد استفاده قرار می‌گیرد و در عمل عملکرد نسبتاً خوبی دارد. برای سری‌های زمانی با فواصل مساوی، یک هموارسازی اسپلاین مکعبی را می‌توان به عنوان تابع $f(t)$ تعریف کرد:

به میانگین و واریانس در طول زمان ندارند.

- پارامترهای مدل‌ها و خطاها با توجه به گذشت زمان ثابت فرض می‌شوند.

برای محاسبات و کد نویسی از نرم افزار برنامه نویسی R استفاده کرده‌ایم.

پس از تجزیه و تحلیل داده‌ها، مدلسازی و نتایج بدست آمده، معیارهای ارزیابی (خطاهای پیش بینی) هر کدام از روش‌ها برای موارد بستری و فوت کووید-۱۹ در شهرستان یزد در جداول ۲ و ۳ آورده شده است. مطابق جدول ۲، بهترین روش و مدل برای موارد تجمعی بستری کووید-۱۹ شهرستان یزد با توجه به معیارهای ارزیابی یا همان خطاهای پیش‌بینی، مدل ARIMA فصلی یا همان SARIMA با معیارهای ارزیابی $MSE=415/7072$ ، $MAE=17/8935$ و $RMSE=20/3889$ و $MAPE=0/09825$ می‌باشد چرا که کمترین مقادیر ارزیابی را در میان دیگر مدل‌ها دارد و برای آینده، از این روش برای پیش‌بینی موارد تجمعی بستری کووید-۱۹ در شهرستان یزد استفاده می‌کنیم. مدل انتخابی، $ar1=0/9086$ با ضرایب $ma2=0/7553$ و $sma1=0/1909$ ، $ma1=-1/7324$ و $sar1=0/2801$ ، $sar2=0/2801$ و $sma2=0/3008$ است. همچنین با توجه به جدول ۲ بیشترین مقادیر معیارهای ارزیابی مربوط به مدل شبکه عصبی اتورگرسیو است که به معنای عملکرد نامطلوب این مدل در پیش‌بینی موارد تجمعی بستری کووید-۱۹ در شهرستان یزد می‌باشد. مطابق جدول ۳ بهترین روش و مدل برای موارد تجمعی فوت کووید-۱۹ شهرستان یزد با توجه به معیارهای ارزیابی یا همان خطاهای پیش‌بینی (کمترین مقدار)، مدل تتا با معیارهای ارزیابی $MSE=997/8649$ ، $MAE=31/4745$ و $RMSE=31/589$ می‌باشد چرا که کمترین مقادیر ارزیابی را در میان دیگر مدل‌ها دارد و برای آینده، از این روش برای پیش‌بینی موارد تجمعی فوت کووید-۱۹ در شهرستان یزد استفاده می‌کنیم. همچنین بیشترین مقادیر معیارهای ارزیابی مربوط به مدل شبکه عصبی اتورگرسیو است که به معنای عملکرد نامطلوب این مدل در پیش‌بینی موارد تجمعی فوت کووید-۱۹ در شهرستان یزد می‌باشد.

$$RMSE(Y, \hat{Y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (19)$$

میانگین قدر مطلق خطا

$$MAE(Y, \hat{Y}) = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (20)$$

میانگین درصد قدر مطلق خطا

$$MAPE(Y, \hat{Y}) = \frac{100}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \quad (21)$$

که در آن \hat{Y}_i مقدار پیش بینی و Y_i مقدار واقعی می‌باشد.

۳-مدلسازی و تجزیه و تحلیل داده‌ها

در این بخش نتایج برازش مدل و پیش‌بینی مقادیر تجمعی کرونا (مقادیر تجمعی بستری و مقادیر تجمعی فوت) را برای شهرستان یزد، با توجه به مدل‌های ذکر شده در بخش دوم 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 شامل افرادی است که در بیمارستان‌های شهرستان یزد پذیرش و بستری شده‌اند، می‌باشد و افرادی که در مراکز درمانی دیگر یا به اصطلاح مراکز درمان سرپائی مورد پذیرش قرار گرفته باشند، را شامل نمی‌شود. در این پژوهش ۱۴ داده مشاهده شده آخر را به عنوان داده‌های تست^۴ در نظر می‌گیریم و با الباقی داده‌ها و مدل‌های تشریح شده در بخش دوم، برای ۱۴ روز آینده موارد تجمعی کووید-۱۹ (بستری و فوت) را برای شهرستان یزد پیش‌بینی می‌کنیم. و بعد بهترین مدل پیشنهادی به وسیله معیارهای ارزیابی ذکر شده در بخش قبلی، انتخاب می‌شود. لازم به ذکر است ما تاریخی که اولین مورد اعم از بستری یا فوت گزارش شده است را به عنوان روز شروع سری زمانی در نظر گرفته‌ایم.

برای توسعه مدل‌های آماری براساس داده‌های سری‌های زمانی، فرضیات زیر در نظر گرفته شده‌اند:

- داده‌های سری‌های زمانی حاوی ناهنجاری‌ها و داده‌های پرت نیستند.
- داده‌ها تک متغیری هستند. بدین معنی که داده‌های سری زمانی از یک متغیر تشکیل شده‌اند.
- این مدل‌ها فرض می‌کنند داده‌ها ایستا هستند و نیاز

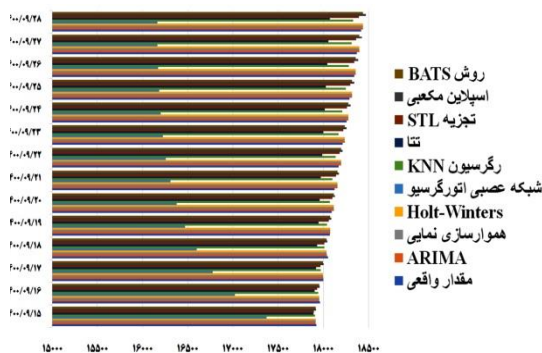
⁴testing data

جدول ۲. معیارهای ارزیابی مربوط به موارد تجمعی بستری کووید ۱۹ در شهرستان یزد

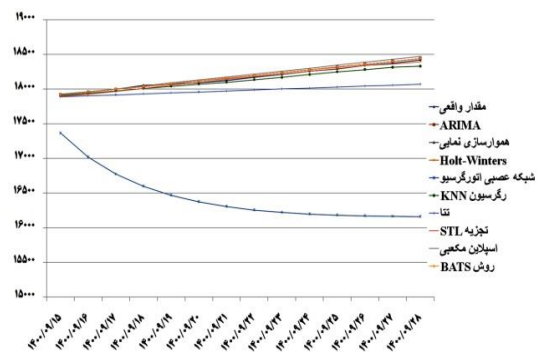
| مدل | MSE | RMSE | MAE | MAPE |
|---------------------|----------|----------|----------|----------|
| ARIMA | ۴۱۵/۷۰۷۲ | ۲۰/۳۸۸۹ | ۱۷/۸۹۳۵ | ۰/۰۹۸۲۵ |
| هموارسازی نمایی | ۴۳۸/۳۵۸۹ | ۲۰/۹۳۷۰ | ۱۸/۳۸۷۱ | ۰/۱۰۰۹۶۸ |
| Holt-Winters | ۴۳۲/۳۱۱۲ | ۲۰/۷۹۲ | ۱۸/۲۶۵ | ۰/۱۰۰۳۰۱ |
| شبکه عصبی اتورگرسیو | ۳۱۸۵۹۷۸ | ۱۷۸۴/۹۳۱ | ۱۷۱۵/۱۱۹ | ۹/۴۲۱۴ |
| رگرسیون KNN | ۱۹۰۳/۰۱۹ | ۴۳/۶۲۳۶ | ۳۹/۳۵۷۸ | ۰/۲۱۵۹ |
| تتا | ۴۲۸۲۰/۸۲ | ۲۰۶/۹۳۱۹ | ۱۸۳/۲۶۹۳ | ۱/۰۰۴۵ |
| روش تجزیه STL | ۴۲۳/۴۷۰۵ | ۲۰/۵۷۸۴ | ۱۸/۸۲۵۲ | ۰/۰۹۹۹۶ |
| اسپلاین مکعبی | ۱۴۲۲/۶۵۵ | ۳۷/۷۱۸۱ | ۳۲/۴۱۹۷ | ۰/۱۷۷۷ |
| روش BATS | ۴۵۵/۷۳۷۱ | ۲۱/۳۴۸۰ | ۱۸/۷۳۱۸ | ۰/۱۰۲۸ |

جدول ۳. معیارهای ارزیابی مربوط به موارد تجمعی فوت کووید ۱۹ در شهرستان یزد

| مدل | MSE | RMSE | MAE |
|---------------------|------------|----------|----------|
| ARIMA | ۱۲۱۵/۱۰۸۰ | ۳۴/۸۵۸۴ | ۳۴/۷۹۲۶ |
| هموارسازی نمایی | ۱۲۲۰/۷۵۴۶ | ۳۴/۹۳۹۳ | ۳۴/۸۶۸۱ |
| Holt-Winters | ۱۲۲۱/۴۸۱۵ | ۳۴/۹۴۹۷ | ۳۴/۸۷۸۵ |
| شبکه عصبی اتورگرسیو | ۲۸۳۶۸/۱۲۵۹ | ۱۶۸/۴۲۸۴ | ۱۵۹/۳۴۴۲ |
| رگرسیون KNN | ۳۹۰۸/۴۲۵۳ | ۶۲/۵۱۷۴ | ۶۱/۳۷۷۲ |
| تتا | ۹۹۷/۸۶۴۹ | ۳۱/۵۸۹۰ | ۳۱/۴۷۴۵ |
| روش تجزیه STL | ۱۰۰۱/۹۸۸۳ | ۳۱/۶۵۴۲ | ۳۱/۵۴۸۹ |
| اسپلاین مکعبی | ۱۰۱۷/۹۶۷۳ | ۳۱/۹۰۵۶ | ۳۱/۸۱۶۷ |
| روش BATS | ۱۲۲۲/۰۱۲۸ | ۳۱/۹۵۷۳ | ۳۴/۸۸۶۰ |



شکل ۳. نمودار میله‌ای پیش‌بینی ۱۴ روزه میزان موارد تجمعی بستری کووید-۱۹ در شهرستان یزد



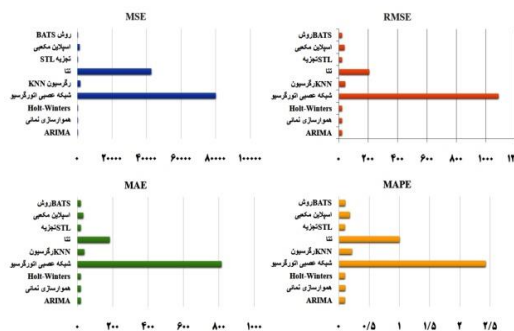
شکل ۲. نمودار پیش‌بینی ۱۴ روزه میزان موارد تجمعی بستری کووید-۱۹ در شهرستان یزد

برای بررسی بهتر عملکرد این ۹ روش شکل‌های (۲) الی (۷) براساس جداول ۲ و ۳ ترسیم شده‌اند. این نمودارها همان مقادیر معیارهای ارزیابی یا خطاهای پیش‌بینی می‌باشد. براساس بررسی شکل‌های (۲) الی (۴)، به نظر می‌رسد که مدل‌های ARIMA و هموارسازی نمایی و Holt-Winters برای پیش‌بینی موارد جدید بستری کووید-۱۹ در شهرستان یزد دارای خطای کمتری هستند به عبارتی این مدل‌ها برای پیش‌بینی موارد تجمعی بستری شهرستان یزد، در یک سطح می‌توان در نظر گرفت، که از این میان مدل ARIMA برای پیش‌بینی موارد تجمعی بستری کووید-۱۹ دارای کمترین خطا (کمترین مقادیر معیارهای ارزیابی) می‌باشد. همچنین با بررسی و تحلیل شکل‌های ۵ الی ۷ به این نتیجه رسیدیم که روش‌های تتا و روش STL و روش اسپلاین مکعبی دارای خطاهای کمتری نسبت به بقیه روش‌ها هستند که روش مدلسازی تتا دارای کمترین خطا (کمترین مقادیر معیارهای ارزیابی) برای پیش‌بینی موارد تجمعی فوت می‌باشند. همچنین از نظر معیارهای ارزیابی، بدترین عملکرد در پیش‌بینی موارد بستری و فوت (بیشترین مقادیر معیارهای ارزیابی)، مدل شبکه‌های عصبی اتورگرسیو دارد.

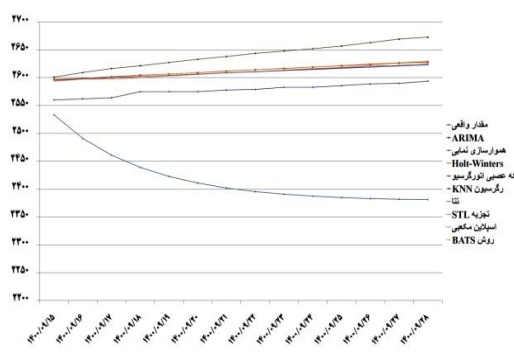
با توجه به ناشناخته بودن رفتار این ویروس و پژوهش‌های صورت گرفته در پیش‌بینی کووید-۱۹ مهم‌ترین جنبه‌ی نوآوری این مطالعه استفاده از ۹ روش متفاوت در زمینه‌های مختلف برای موارد جدید بستری و مرگ‌ومیر کرونا شهرستان یزد، برخلاف پژوهش‌های انجام شده که از ۲ الی ۳ روش در زمینه خاصی بهره گرفته‌اند، می‌باشد. در 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.

نتایج کاربردی و دیدگاه‌های مدیریتی حاصل از مطالعه حاضر به قرار زیر می‌باشد:

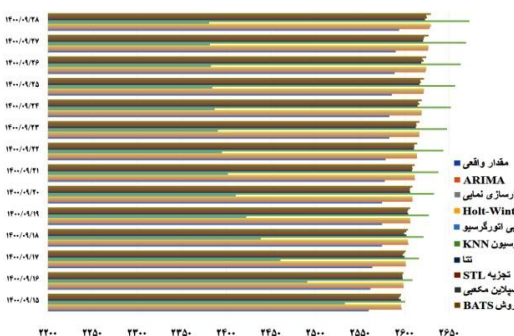
- ۱- نتایج تحلیل نشان داد که بهترین مدل با توجه به معیارهای ارزیابی مذکور برای پیش‌بینی موارد تجمعی بستری کووید-۱۹ مدل ARIMA و برای موارد تجمعی فوت مدل تتا می‌باشد. همچنین مدل شبکه‌های عصبی اتورگرسیو دارای بدترین عملکرد را در میان دیگر مدل‌ها هم برای موارد بستری و هم موارد فوت دارا می‌باشد.
- ۲- مهم‌ترین نکته‌ای که می‌توان از نتایج تحلیل برداشت



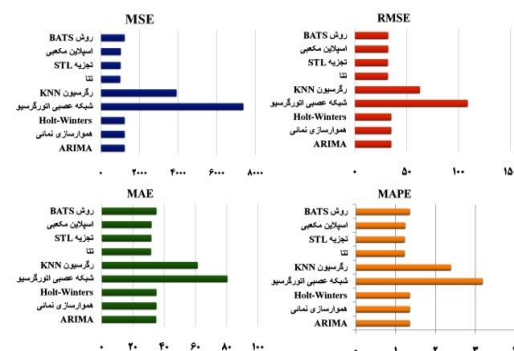
شکل ۴. نمودار معیارهای ارزیابی موارد تجمعی بستری کووید-۱۹ در شهرستان یزد



شکل ۵. نمودار پیش‌بینی ۱۴ روزه میزان موارد تجمعی فوت کووید-۱۹ در شهرستان یزد



شکل ۶. نمودار میله‌ای پیش‌بینی ۱۴ روزه میزان موارد تجمعی فوت کووید-۱۹ در شهرستان یزد



شکل ۷. نمودار معیارهای ارزیابی موارد تجمعی فوت کووید-۱۹ در شهرستان یزد

موردی، مورد استفاده و با استفاده از معیارهای ارزیابی MAPE، RMSE، MSE و MAE با یکدیگر مقایسه شدند. در این میان ۱۴ داده آخر را به عنوان داده تست در نظر گرفتیم. بهترین مدل برای پیش‌بینی موارد تجمعی بستری و مرگ‌ومیر کووید-۱۹ در شهرستان یزد متفاوت می‌باشند، برای موارد تجمعی بستری در یزد بهترین مدل، ARIMA فصلی یا همان SARIMA هست درحالیکه برای موارد تجمعی فوت مدل تتا به عنوان بهترین مدل از نظر معیارهای ارزیابی برگزیده شد. با توجه به نتایج این مطالعه به محققین آتی پیشنهاد می‌شود:

- پیش‌بینی کووید-۱۹ با در نظر گرفتن اثر استفاده از ماسک و رعایت فاصله اجتماعی بررسی شود.
- پیش‌بینی کووید-۱۹ با ترکیب روش‌ها مثل روش تتا و STL بررسی شود.
- پیش‌بینی کووید-۱۹ با توجه به اثر واکسیناسیون علیه کووید-۱۹ بررسی شود.
- از روش‌های استفاده شده برای دیگر بیماری‌های واگیردار یا همه‌گیری برای پیش‌بینی استفاده شود.
- این روش‌ها را می‌توان در مکانی دیگر مدل‌سازی و نتیجه‌گیری کرد.

کرد، این است که نمی‌توان مدل یکتایی برای موارد بستری و فوت در نظر گرفت و همچنین با استفاده از مدل‌هایی که در سال‌های خیلی قبل معرفی شده‌اند و یا به اصطلاح مدل‌های قدیمی، می‌توان در عمل از آن‌ها نتایج بهتری کسب نمود، درحالیکه نتایج به دست آمده از مدل‌های جدیدتر مثل مدل شبکه‌های عصبی اتورگرسیو نتایج بدتری در پیش‌بینی کووید-۱۹ در مطالعه موردی 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.

۳- و می‌توان از نتایج حاصله برای پیش‌بینی و به کنترل درآوردن بیماری‌های ویروسی دیگر که ممکن است در آینده به عنوان یک بیماری بومی، منطقه‌ای و یا حتی همه‌گیری ظهور کند، استفاده کرد.

۴- نتیجه‌گیری

در مطالعه حاضر، از ۹ روش ARIMA، تتا، هموارسازی نمائی، Holt-Winters، شبکه عصبی اتورگرسیو، رگرسیون KNN، روش BATS، روش هموارسازی اسپلاین مکعبی و روش STL برای پیش‌بینی تعداد موارد تجمعی جدید و مرگ‌ومیر کووید-۱۹ در شهرستان یزد به عنوان یک مطالعه

مراجع

- [1] Cucinotta, Domenico, and Maurizio Vanelli. "WHO declares COVID-19 a pandemic." *Acta bio medica: Atenei parmensis* 91, no. 1 (2020): 157.
- [2] Cheng, Zhangkai J., and Jing Shan. "2019 Novel coronavirus: where we are and what we know." *Infection* 48 (2020): 155-163.
- [3] Tran, Thai T., L. T. Pham, and Q. X. Ngo. "Forecasting epidemic spread of SARS-CoV-2 using ARIMA model (Case study: Iran)." *Global Journal of Environmental Science and Management* 6, no. Special Issue (Covid-19) (2020): 1-10.
- [4] Zhang, Xingyu, Yuanyuan Liu, Min Yang, Tao Zhang, Alistair A. Young, and Xiaosong Li. "Comparative study of four time series methods in forecasting typhoid fever incidence in China." *PloS one* 8, no. 5 (2013): e63116.
- [5] Chen, Ye, Kunkun Leng, Ying Lu, Lihai Wen, Ying Qi, Wei Gao, Huijie Chen et al. "Epidemiological features and time-series analysis of influenza incidence in urban and rural areas of Shenyang, China, 2010–2018." *Epidemiology & Infection* 148 (2020): e29.
- [6] Olanrewaju, Samuel Olayemi, Emmanuel Oluwatosin Ojo, and Emmanuel Segun Oguntade. "Time Series Analysis on Reported Cases of Tuberculosis in Minna Niger State Nigeria." *Open Journal of Statistics* 10, no. 3 (2020): 412-430.
- [7] Nayak, M. Siva Durga Prasad, and K. A. Narayan. "Forecasting dengue fever incidence using ARIMA analysis." *International Journal of Collaborative Research on Internal Medicine & Public Health* 11, no. 3 (2019): 924-932.

- [8] Wu, Wei, Shu-Yi An, Peng Guan, De-Sheng Huang, and Bao-Sen Zhou. "Time series analysis of human brucellosis in mainland China by using Elman and Jordan recurrent neural networks." *BMC infectious diseases* 19 (2019): 1-11.
- [9] Ceylan, Zeynep. "Estimation of COVID-19 prevalence in Italy, Spain, and France." *Science of The Total Environment* 729 (2020): 138817.
- [10] Bayyurt, Lutfi, and Burcu Bayyurt. "Forecasting of COVID-19 cases and deaths using ARIMA models." *medrxiv* (2020): 2020-04.
- [11] Tandon, Hiteshi, Prabhat Ranjan, Tanmoy Chakraborty, and Vandana Suhag. "Coronavirus (COVID-19): ARIMA-based time-series analysis to forecast near future and the effect of school reopening in India." *Journal of Health Management* 24, no. 3 (2022): 373-388.
- [12] Perone, Gaetano. "An ARIMA model to forecast the spread and the final size of COVID-2019 epidemic in Italy." *MedRxiv* (2020): 2020-04.
- [13] Ghosal, Samit, Sumit Sengupta, Milan Majumder, and Binayak Sinha. "Linear Regression Analysis to predict the number of deaths in India due to SARS-CoV-2 at 6 weeks from day 0 (100 cases-March 14th 2020)." *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 14, no. 4 (2020): 311-315.
- [14] Parbat, Debanjan, and Monisha Chakraborty. "A python based support vector regression model for prediction of COVID19 cases in India." *Chaos, Solitons & Fractals* 138 (2020): 109942.
- [15] Naimoli, Antonio. "Modelling the persistence of Covid-19 positivity rate in Italy." *Socio-Economic Planning Sciences* 82 (2022): 101225.
- [16] Kibria, Hafsa Binte, Oishi Jyoti, and Abdul Matin. "Forecasting the spread of the third wave of COVID-19 pandemic using time series analysis in Bangladesh." *Informatics in medicine unlocked* 28 (2022): 100815.
- [17] Khan, Farhan Mohammad, and Rajiv Gupta. "ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India." *Journal of Safety Science and Resilience* 1, no. 1 (2020): 12-18.
- [18] Chowdhury, Anjir Ahmed, Khandaker Tabin Hasan, and Khadija Kubra Shahjalal Hoque. "Analysis and prediction of COVID-19 pandemic in Bangladesh by using ANFIS and LSTM network." *Cognitive Computation* 13, no. 3 (2021): 761-770.
- [19] Alassafi, Madini O., Mutasem Jarrah, and Reem Alotaibi. "Time series predicting of COVID-19 based on deep learning." *Neurocomputing* 468 (2022): 335-344.
- [20] Chyon, Fuad Ahmed, Md Nazmul Hasan Suman, Md Rafiul Islam Fahim, and Md Sazol Ahmmed. "Time series analysis and predicting COVID-19 affected patients by ARIMA model using machine learning." *Journal of virological methods* 301 (2022): 114433.
- [21] Nair, Siddharth, Ganesan Ckm, R. Varsha, Sankhasubhra Ghosal, M. Vergin, and L. Jani Anbarasi. "Intelligent Forecasting Strategy for COVID-19 Pandemic Trend in India: A Statistical Approach." In *Artificial Intelligence and Technologies: Select Proceedings of ICRTAC-AIT 2020*, pp. 553-560. Springer Singapore, 2022.
- [22] Mukhairez, Hosam HA, and Ahmed JI Alaff. "Short-term Forecasting of COVID-19." *Computational Intelligence for COVID-19 and Future Pandemics: Emerging Applications and Strategies* (2022): 257-266.
- [23] Prajapati, Samyak, Aman Swaraj, Ronak Lalwani, Akhil Narwal, and Karan Verma. "Comparison of traditional and hybrid time series models for forecasting COVID-19 cases." *arXiv preprint arXiv:2105.03266* (2021).
- [24] Li, Cai, Agyemang Kwasi Sampene, Fredrick Oteng Agyeman, Brenya Robert, and Abraham Lincoln Ayisi. "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 (2022).
- [25] Yaffee, Robert Alan, and Monnie McGee. *An introduction to time series analysis and forecasting: with applications of SAS® and SPSS®*. Elsevier, 2000.
- [26] Fanoodi, Bahareh, Behnam Malmir, and Farzad Firouzi Jahantigh. "Reducing demand uncertainty in the platelet supply chain through artificial neural networks and ARIMA models." *Computers in biology and medicine* 113 (2019): 103415.
- [27] Kufel, Tadeusz. "ARIMA-based forecasting of the dynamics of confirmed Covid-19 cases for selected European countries." *Equilibrium. Quarterly Journal of Economics and Economic Policy* 15, no. 2 (2020): 181-204.

- [28] Martínez, Francisco, María Pilar Frías, Francisco Charte, and Antonio J. Rivera. "Time Series Forecasting with KNN in R: the tsfknn Package." *R J.* 11, no. 2 (2019): 229.
- [29] Assimakopoulos, Vassilis, and Konstantinos Nikolopoulos. "The theta model: a decomposition approach to forecasting." *International journal of forecasting* 16, no. 4 (2000): 521-530.
- [30] Cleveland, Robert B., William S. Cleveland, Jean E. McRae, and Irma Terpenning. "STL: A seasonal-trend decomposition." *J. Off. Stat* 6, no. 1 (1990): 3-73.
- [31] Hyndman, Rob J., Maxwell L. King, Ivet Pitrun, and Baki Billah. "Local linear forecasts using cubic smoothing splines." *Australian & New Zealand Journal of Statistics* 47, no. 1 (2005): 87-99.
- [32] De Livera, Alysha M., Rob J. Hyndman, and Ralph D. Snyder. "Forecasting time series with complex seasonal patterns using exponential smoothing." *Journal of the American statistical association* 106, no. 496 (2011): 1513-1527.