

## مدلسازی قابل اطمینان کووید-۱۹ با استفاده از انواع تکنیک‌های سری زمانی و یادگیری ماشین و مقایسه کارایی آنها (مطالعه موردی: شهرستان بافق)

محمدحسین کریمی زارچی<sup>۱</sup> و داود شیشه‌بری<sup>۲</sup>

۱- کارشناسی ارشد، گروه مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه یزد، یزد، ایران، [karimizarchi.ie@yahoo.com](mailto:karimizarchi.ie@yahoo.com)

۲- دانشیار، گروه مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه یزد، یزد، ایران، [shishebori@yazd.ac.ir](mailto:shishebori@yazd.ac.ir)

### چکیده

بیماری کرونا ویروس ۲۰۱۹ یا کووید-۱۹ که به آن بیماری تنفسی حاد ان‌کاو-۲۰۱۹ یا به شکل عمومی به آن کرونا نیز می‌گویند، یک بیماری تنفسی است که در اثر سندرم تنفسی حاد کرونا ویروس-۲ ایجاد می‌شود. پیش بینی تعداد موارد جدید و مرگ و میر در این مدت می‌تواند گام مفیدی در پیش‌بینی هزینه‌ها و امکانات مورد نیاز در آینده باشد. هدف از این مطالعه مدلسازی و پیش بینی موارد جدید و مرگ و میر در آینده است. در این مقاله ۹ تکنیک متداول پیش‌بینی بر روی داده‌های کووید-۱۹ شهرستان بافق-یزد به عنوان یک 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 قدرمطلق خطا (MAPE) مدل‌ها باهم مقایسه شدند نتایج تحلیل نشان داد که بهترین مدل با توجه به معیارهای ارزیابی مذکور برای پیش‌بینی موارد تجمعی بستری کووید-۱۹ مدل BATS و برای موارد تجمعی فوت مدل هموارسازی اسپلاین مکعبی می‌باشد. همچنین مدل شبکه عصبی اتورگرسیو دارای بدترین عملکرد را در میان دیگر مدل‌ها برای موارد بستری و برای موارد فوت، مدل رگرسیون KNN بدترین عملکرد را دارا می‌باشد.

واژه‌های کلیدی: پاندمیک، کووید-۱۹، پیش‌بینی، سری زمانی، مدلسازی آماری.

### مقدمه

هستند[۲]. در ۱۱ مارس سال ۲۰۲۰، سازمان بهداشت جهانی شیوع ویروس کرونا (NCOV-2019) 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 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 مختلف 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[۳]. مدارکی که

بیماری کرونا ویروس ۲۰۱۹ یا کووید-۱۹ که به آن بیماری تنفسی حاد ان‌کاو-۲۰۱۹ 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 مشخص شد، یک سویه از ویروس‌های کرونا می‌باشد. این ویروس اولین بار در دسامبر ۲۰۱۹ شناسایی شد. گروهی از بیماران مبتلا به این ویروس شکل جدیدی از ذات‌الریه ویروسی از خود نشان دادند. همچنین همه بیماران، سابقه مشابه بازدید از بازار تره‌بار در ووهان چین را داشتند[۱]. علایم معمول آن تب، سرفه، تنگی نفس، و نابویایی هستند. درد عضلانی، تولید خلط، گلودرد، ناچشایی و سرخی چشم از جمله نشانه‌های کمتر معمول آن

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. غسل و همکاری‌ان [۱۵] 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. پاربات و چاکرابورتی [۱۶] برای پیش‌بینی موارد کووید-۱۹ در هند به مدت ۶۰ روز براساس داده‌های سری زمانی 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 SVR دارای دقت حدودی ۹۷ درصدی در پیش‌بینی موارد تلفات، موارد بهبود یافته و موارد تأیید شده تجمعی است. آنتونیو نامولی [۱۷] برای پیش‌بینی موارد کووید-۱۹ در ایتالیا براساس داده‌های سری زمانی 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.

کیبیریا و همکاری‌ان [۱۸] از مدل ARIMA برای تخمین تعداد مورد انتظار 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 [۱۹] از مدل سری زمانی تک متغیره برای پیش‌بینی تعداد موارد آلوده به کووید-۱۹ که می‌توان 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 پیاده‌سازی 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 هیچ گونه مداخله اضافی استفاده کردند. نتایج نشان‌دهنده، روند افزایشی در تعداد واقعی و پیش‌بینی شده موارد کووید-۱۹ با حدود ۱۵۰۰ مورد در روز، بر اساس داده‌های موجود تا ۴ آوریل ۲۰۲۰ بود. چاودوری و همکاری‌ان [۲۰] 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 از انسان به انسان دیگر منتقل می‌شود توسط مراکز کنترل و پیشگیری از بیماری تأیید شده است و همچنین گزارش شده است که کووید-۱۹ با لمس سطوح، تماس نزدیک با اشیاء و هوای حاوی ذرات ویروسی قابل انتقال است [۴].

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 کنار آمدن 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 و بهبود یافته و همچنین موارد مرگ و میر، نقش مهمی در برنامه‌ریزی برای کنترل شیوع کووید-۱۹ در جهان دارد.

مدل‌های آماری مختلفی برای پیش‌بینی تعداد موارد آینده و پیش‌بینی شیوع بیماری‌های عفونی در آینده نزدیک و همچنین پیش‌بینی تعداد موارد کووید-۱۹ در آینده استفاده می‌شود [۵]. ژانگ و همکاری‌ان [۶] برای پیش‌بینی تب حصبه از مدل سری زمانی اتورگرسیون 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 همکاری‌ان [۷] با استفاده از مدل SARIMA شیوع آنفولانزا را در مناطق شهری و روستایی شیانگ 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 اتورگرسیون میانگین متحرک تلفیق شده (ARIMA) برای پیش‌بینی بیماری‌های عفونی مانند سل [۸]، تب دانگ [۹] و تب مالت [۱۰] استفاده کرده‌اند.

سیلان [۱۱] 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 کشورها انتخاب کرده‌اند. لوتفی و بورکو [۱۲] برای پیش‌بینی تعداد موارد تأیید شده و مرگ‌ومیر کووید-۱۹ از مدل ARIMA در مرکز کنترل و پیشگیری اروپا (ECDC) استفاده کردند. تاندن و همکاری‌ان [۱۳] مدل ARIMA را برای پیش‌بینی موارد کووید-۱۹ در هند بکار بردند. این مطالعه نشان دهنده روند صعودی موارد کووید-۱۹ در روزهای آینده است. تجزیه و تحلیل سری‌های زمانی نیز افزایش نمایی تعداد موارد را نشان می‌دهد. فرض بر این است که مدل‌های پیش‌بینی فعلی به دولت و پرسنل پزشکی کمک می‌کند تا برای شرایط آینده آماده باشند و در سیستم‌های مراقبت‌های بهداشتی آمادگی بیشتری داشته باشند. پرون [۱۴] در این مقاله یک مدل اتورگرسیون میانگین متحرک تلفیق

یکی از رایج‌ترین روش‌های پیش‌بینی سری‌های زمانی (ARIMA) بود.

در مقاله‌ای دیگر که در سال ۲۰۲۲ انجام شد لی و همکاران [۲۶] بریتانیا، آفریقای جنوبی، برزیل و نیجریه کشورهای هستند که به عنوان نمونه پژوهشی برای پژوهش حاضر 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 در بریتانیا، آفریقای جنوبی، برزیل و نیجریه استفاده کرده‌اند.

در مطالعه حاضر، قرار است تا موارد تجمعی کووید-۱۹ (موارد تأیید شده و مرگ‌ومیر) برای شهرستان بافق، 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)، هولت-وینترز (Holt-Winter)، 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.

### مدل‌های آماری

این مطالعه از مدل‌های آماری اتورگرسیو میانگین متحرک تلفیق شده (ARIMA)، هموارسازی نمایی (ETS)، هولت-وینترز (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 شهرستان بافق، یزد استفاده کرده‌ایم. برای داشتن یک پیش‌بینی معنی‌دار آماری از داده‌های سری زمانی، حداقل حجم نمونه ۳۰ مشاهده مورد نیاز است [۲۷]. 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 [۲۸]. 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

کووید-۱۹ در بنگلادش استفاده کردند. نتایج هر دو روش را مورد مقایسه قرار دادند که می‌توان گفت که LSTM نتایج رضایت‌بخش‌تری را نشان داده است. الاسافی و همکاران [۲۱] 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 مورد ارزیابی قرار گرفت. در این مطالعه، پیش‌بینی‌ها را با داده‌هایی که تا ۳ دسامبر ۲۰۲۰ در دسترس بود، انجام شد.

چایون و همکاران [۲۲] در مطالعه‌ای، مدل ARIMA را برای تحلیل زمانی گسترش جهانی کووید-۱۹ در بازه زمانی از ۲۲ ژانویه ۲۰۲۰ تا ۷ آوریل ۲۰۲۰ استفاده کردند. در مقاله‌ای دیگر، نایر و همکاران [۲۳] پیش‌بینی موارد مرگ‌ومیرهای تأیید شده تجمعی کووید-۱۹ برای هند بر اساس مدل‌های آماری مختلف مانند سری زمانی، یادگیری ماشین و یادگیری گروهی تحلیل می‌شود. مدل ARIMA و هموارسازی نمایی هولت-وینترز 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 آموزش دیده 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 آلفا [۲۴] از الگوریتم‌های پیش‌بینی کوتاه‌مدت برای پیش‌بینی تعداد مبتلایان و مرگ‌ومیرها در آینده کوتاه استفاده کردند. پیش‌بینی با استفاده از مدل ARIMA، ترکیبی، هولت-وینترز و هموارسازی نمایی با مقدار زمانی برابر با ۵۷ انجام شد. پس از ارزیابی الگوریتم‌های آزمایش شده با استفاده از MAPE، متوجه شدند که هموارسازی نمایی بهترین الگوریتم پیش‌بینی 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 هولت-وینترز بهترین الگوریتم برای پیش‌بینی موارد مرگ‌ومیر با MAPE برابر با ۵/۳۳ درصد می‌باشد. پراجاپاتی و همکاران [۲۵] از مدل‌های پیش‌بینی برای مجموع موارد تجمعی هر کشور با مقایسه داده‌های گزارش شده و مقدار پیش‌بینی شده و سپس رتبه‌بندی الگوریتم‌ها (Prophet, Holt-Winters, LSTM, ARIMA, ) براساس مقادیر MAPE و MAE، RMSE، میانگین مربعات خطا (RMSE) و میانگین درصد خطا مطلق (MAPE) آن‌ها انجام می‌شود. ترکیب مدل‌های ARIMA و شبکه عصبی اتورگرسیو غیرخطی (NARNN) بهترین نتیجه را در بین مدل‌های انتخابی با کاهش RMSE ارائه کرد که تقریباً ۳۵/۳ درصد بهتر از

می‌نامند و اغلب بصورت  $ARIMA(p,d,q)(P,D,Q)_s$  نشان می‌دهند که در آن اندیس  $s$  دوره فصلی را نشان می‌دهد.

### روش هموارسازی نمایی

یکی از روش‌های پیش‌بینی رایج و محبوب روش هموارسازی نمایی است. محبوبیت آن ناشی از انعطاف‌پذیری، سهولت اتوماسیون، محاسبات ساده، و عملکرد خوب است. روش هموارسازی نمایی به جای اینکه یک میانگین ساده را حساب کند میانگین وزنی مقادیر گذشته را می‌گیرد بطوریکه وزن‌ها بصورت نمایی متمایل به داده‌های گذشته کاهش می‌یابد بدین معنا که به داده‌های اخیر اهمیت بیشتری نسبت به داده‌های قدیمی‌تر می‌دهد. روش هموارسازی نمایی 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 که روش هموارسازی در زمان  $t+1$  به دست می‌دهد بصورت رابطه (۸) می‌باشد:

$$F_{t+1} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \dots \quad (8)$$

در رابطه بالا  $\alpha$  ثابت هموارسازی نامیده می‌شود و یک مقدار ثابت بین صفر و یک است. رابطه بالا هموارسازی نمایی را با استفاده از میانگین وزنی از تمام مشاهدات گذشته با وزن‌های نزولی نمایی نشان می‌دهد. ثابت هموارسازی  $\alpha$  توسط کاربر تعیین می‌شود. مقادیر پیش فرضی که برای  $\alpha$  در نظر گرفته شده و نشان‌دهنده‌ی عملکرد خوب آن می‌باشد، در محدوده  $0/1$  تا  $0/2$  است. همچنین برای انتخاب  $\alpha$  روش آزمون و خطا نیز می‌تواند راهکار مناسبی باشد. به هر حال باید در انتخاب  $\alpha$  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.

### روش هولت-وینترز

روش پیش‌بینی هولت-وینترز توسعه‌ای از روش هموارسازی نمایی می‌باشد. مدل هولت-وینترز، 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. چنین مدلی دارای چندین پارامتر است. این گروه از پارامترها در این مدل با  $\alpha, \beta, \gamma$  شناخته می‌شوند. شکل ساده (بدون روند و

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.

### فرآیند اتورگرسیو میانگین متحرک تلفیق شده (ARIMA)

مدل  $ARIMA(p,d,q)$  اولین بار توسط باکس و جنکینز در سال ۱۹۷۶ معرفی شد [۲۹]. 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، یک‌بار 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  $\varphi_p(B) = (1 - \varphi_1 B - \dots - \varphi_p B^p)$  و عملگر  $MA$  وارون‌پذیر  $\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$  می‌باشد. پارامتر  $\theta_0$  برای  $d=0$  و  $d>0$  نقش‌های بسیار متفاوتی را بازی می‌کند.

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

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

$$SARIMA: \Phi_p(B^s)\varphi_p(B)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\theta_q(B^s)a_t \quad (2)$$

$$ARIMA: \varphi_p(B)(1-B)^d Z_t = \theta_q(B)a_t \quad (3)$$

$$\varphi_p(B) = (1 - \varphi_1 B - \dots - \varphi_p B^p) \quad (4)$$

$$\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q) \quad (5)$$

$$\Phi_p(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_p B^{Ps} \quad (6)$$

$$\Theta_q(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_q B^{Qs} \quad (7)$$

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

شده است:  $f_1^i, f_2^i, \dots, f_n^i$ ، توصیف نمونه و یک بردار هدف مرتبط از ویژگی‌های  $t_1^i, t_2^i, \dots, t_m^i$ ؛ داده یک نمونه جدید، که ویژگی‌های آن شناخته شده است  $q_1, \dots, q_n$  اما هدف آن ناشناخته است. ویژگی‌های نمونه جدید برای یافتن  $k$  مشابه‌ترین نمونه‌های آموزشی آن با توجه به بردار ویژگی‌ها و شباهت یا 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،  $k$  مشابه‌ترین نمونه‌ها یا  $k$  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 آن احتمالاً مشابه هدف ناشناخته آن هستند. به این ترتیب، اهداف نزدیک‌ترین همسایه‌ها برای پیش‌بینی هدف نمونه جدید جمع می‌شوند. به عنوان مثال، با فرض اینکه اهداف  $k$  نزدیک‌ترین همسایه بردارها هستند:  $t^1, \dots, t^k$  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 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 به این امید که اهداف آن‌ها مشابه هدف ناشناخته‌اش باشد. حال، بیایید ببینیم که چگونه 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. به عنوان مثال، اجازه دهید با یک سری زمانی ماهانه شروع کنیم 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 کنید 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 آن 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 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

تغییرات فصلی) مدل سری زمانی هولت-وینترز به صورت رابطه (۹) است:

$$S_t = \alpha \frac{y_t}{I_{t-1}} + (1 - \alpha)(S_{t-1} + b_{t-1}) \quad (9)$$

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  $\alpha$  حضور داشته و هر مشاهده به صورت ترکیب خطی از نقطه قبلی  $(S_{t-1})$  دیده می‌شود. از طرفی  $I$  نیز شاخص تغییرات فصلی نامیده شده و  $L$  نیز طول دوره تغییرات فصلی است. در صورتی که مدل دارای روند هم 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  $\gamma$  نیز در مدل اضافه شده است. در نهایت مدلی که به همراه تغییرات فصلی ارائه می‌شود به صورت زیر خواهد بود:

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

به منظور پیش‌بینی در مدل هولت-وینترز، از رابطه (۱۲) استفاده می‌کنیم.

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

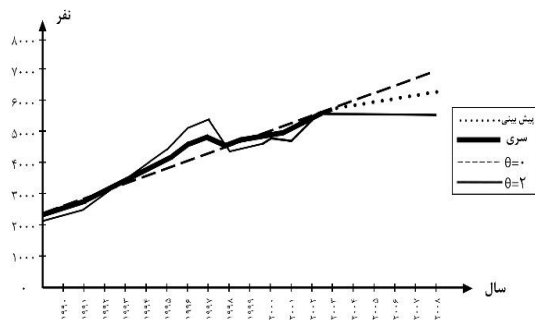
لازم به ذکر است که پارامترها در این مدل  $\alpha, \beta, \gamma$  ثابت هموارسازی و بین عدد صفر و یک می‌باشند.

### 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.

### 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 به سادگی مجموعه‌ای از نمونه‌های آموزشی را در اختیار دارد. نمونه آموزشی  $i$ ام از یک بردار با  $n$  ویژگی تشکیل



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 در حالیکه 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 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 [۳۱].

### روش STL

STL یک روش فیلتر برای 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. سپس:

$$Y_v = T_v + S_v + R_v \quad (17)$$

STL متشکل از دنباله‌ای از عملیات هموارسازی است که از رگرسیون وزن محلی یا Loess استفاده می‌کند. STL چندین پارامتر دارد که باید توسط تحلیلگر داده انتخاب شود. STL شش 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 /  $n_{(l)}$ : پارامتر

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 جمع می‌شود [۳۰].

**جدول ۱:** نمونه‌های آموزشی برای سری زمانی  $t$  برای پیش‌بینی یک گام جلوتر و تاخیرهای ۱:۲ به عنوان بردار ویژگی

ویژگی‌ها	هدف
$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}$	$X_{13}$
$X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}$	$X_{14}$
$X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}$	$X_{15}$
...	...
$X_{120}, X_{121}, X_{122}, \dots, X_{129}, X_{130}, X_{131}$	$X_{132}$

### مدلسازی تتا

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

$$X_{new}^t(\theta) = \theta, X_{data}^t \quad \text{where } X_{data}^t = X_t - 2X_{t-1} - X_{t-2} \quad \text{at time } t \quad (15)$$

اگر مقدار تتا به تدریج کاهش یابد، سری زمانی کاهش می‌یابد. هرچه مقدار ضریب تتا کوچکتر باشد، درجه قلعه‌ها کاهش پیدا می‌کند. در حالتی که  $\theta=0$  است، سری زمانی به یک خط رگرسیون خطی تبدیل می‌شود. ضریب  $\theta$  نیز می‌تواند مقادیر منفی داشته باشد. برعکس، اگر ضریب تتا افزایش یابد ( $\theta < 1$ ) آنگاه سری زمانی گشاد می‌شود. فرمول کلی روش تتا بشرح رابطه (۱۶) می‌باشد. سری زمانی اولیه به دو یا چند خط تتا 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  $\theta$  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  $L(\theta=0)$  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. ترکیب ساده این دو پیش‌بینی، پیش‌بینی نهایی مدل تتا را برای سری‌های زمانی خاص، همانطور که در شکل ۱ نشان داده شده است، به دست می‌آید.

مدل BATS واضح‌ترین تعمیم مدل‌های نوآوری فصلی سنتی است که برای دوره‌های فصلی متعدد امکان‌پذیر است. با این حال، نمی‌تواند فصلی غیر صحیح را در خود جای دهد، و 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 [۳۴] مراجعه شود.

### معیارهای ارزیابی

معیارهای ارزیابی بصورت زیر محاسبه می‌شوند:

میانگین مربعات خطا (MSE)

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

جزر میانگین مربعات خطا (RMSE)

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

میانگین قدر مطلق خطا (MAE)

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

میانگین درصد قدرمطلق خطا (MAPE)

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

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

### مدلسازی و نتایج

در این بخش نتایج برازش مدل و پیش‌بینی مقادیر تجمعی کرونا (مقادیر تجمعی بستری و مقادیر تجمعی فوت) را برای شهرستان بافق، با توجه به مدل‌های ذکر شده در بخش قبلی 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 می‌باشد که آمار موجود برای افرادی که در بیمارستان شهرستان بافق پذیرش و بستری شده، 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 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) برای شهرستان بافق پیش‌بینی می‌کنیم. و بعد بهترین مدل پیشنهادی (به‌وسیله معیارهای ارزیابی ذکر شده در بخش قبلی) را انتخاب می‌کنیم. لازم به ذکر است ما تاریخی را 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:

هموارسازی برای فیلتر پایین-گذر /  $\pi(t)$ : پارامتر هموارسازی برای مؤلفه روند /  $\pi(s)$ : پارامتر هموارسازی برای مؤلفه فصلی.

انتخاب پنج مورد اول ساده است. با این حال، آخرین پارامتر،  $\pi(s)$ ، باید به دقت برای هر برنامه تنظیم شود. برای اطلاعات بیشتر در مورد این روش به رفرنس [۳۲] مراجعه شود.

### 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 مشاهده شده 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 که بر روی همه توابع دو بار قابل تمایز  $f$  در  $S$  که در آن  $[1, n] \subseteq S \subseteq R$  به حداقل می‌رسد.

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

راه حل اول شامل چند جمله‌ای‌های مکعبی تکه‌تکه‌ای است که در زمان مشاهده،  $t=1, 2, \dots, n$  به هم وصل شده‌اند. علاوه‌براین، راه حل دارای مشتق دوم صفر در  $t=n$  است. بنابراین، برون‌یابی  $f(t)$  برای  $t > n$  خطی است. برون‌یابی خطی  $f(t)$  پیش‌بینی‌های نقطه‌ای ما را ارائه 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  $\lambda$  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 [۳۳] مراجعه شود.

### روش BATS

شناسه BATS مخفف ویژگی‌های کلیدی مدل است: تبدیل Box-Cox، خطاهای ARMA، روند و اجزای فصلی. برای نشان دادن پارامتر Box-Cox، 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.

هولت-وینترز	۱/۸۲۷۶	۱/۳۵۱۹	۱/۰۸۵۱	۰/۷۶۹۸
شبکه عصبی اتورگرسیو	۰/۲۳۵۳	۰/۴۸۵۱	۰/۴۱۰۷	۰/۲۹۱۵
رگرسیون KNN	۶/۴۱۱۵	۲/۵۳۲۱	۲/۴۶۵۸	۱/۷۴۹۸
تتا	۰/۱۳۴۸	۰/۳۶۷۲	۰/۳۱۲۴	۰/۲۲۱۷
STL	۰/۷۷۲۴	۰/۸۷۸۹	۰/۷۶۷۷	۰/۵۴۵۵
اسپلاین مکعبی	۰/۱۱۰۶	۰/۳۳۲۶	۰/۳۹۱۸	۰/۲۰۷۲
BATS	۱/۸۱۹۵	۱/۳۴۸۹	۱/۰۸۲۵	۰/۷۶۸۰

برای بررسی بهتر عملکرد این ۹ روش شکل‌های ۲ الی ۷ ترسیم شده‌اند. بر اساس بررسی شکل‌های ۲ الی ۴، به نظر می‌رسد که مدل‌های هموارسازی اسپلاین مکعبی، BATS، ARIMA، هموارسازی نمایی و هولت-وینترز برای پیش‌بینی موارد جدید بستری کووید-۱۹ دارای خطای کمتری هستند به عبارتی این مدل‌ها را برای پیش‌بینی موارد تجمعی بستری شهرستان بافق در یک سطح می‌توان در نظر گرفت که از این میان مدل BATS برای پیش‌بینی موارد تجمعی بستری کووید-۱۹ دارای کمترین خطا می‌باشد. همچنین با بررسی و تحلیل شکل‌های ۵ الی ۷ به این نتیجه رسیدیم که روش‌های مدلسازی تتا و هموارسازی اسپلاین مکعبی دارای خطاهای کمتری نسبت به بقیه روش‌ها هستند که روش هموارسازی اسپلاین مکعبی دارای کمترین خطا برای پیش‌بینی موارد تجمعی فوت می‌باشند. همچنین مدل شبکه‌های عصبی اتورگرسیو دارای بدترین عملکرد را در میان دیگر مدل‌ها برای موارد بستری و برای موارد فوت، مدل رگرسیون KNN بدترین عملکرد را دارا می‌باشد.

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 BATS و برای موارد تجمعی فوت مدل هموارسازی اسپلاین مکعبی می‌باشد. همچنین مدل شبکه‌های عصبی اتورگرسیو دارای بدترین عملکرد را در میان دیگر مدل‌ها برای موارد بستری و برای موارد فوت، مدل رگرسیون KNN بدترین عملکرد را دارا می‌باشد.

۲- همچنین مهم‌ترین نکته‌ای که می‌توان از نتایج تحلیل برداشت 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 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.

- 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 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.

با توجه به تجزیه و تحلیل داده‌ها و نتایج بدست آمده مطابق جدول ۲، بهترین روش و مدل برای موارد تجمعی بستری کووید-۱۹ شهرستان بافق با توجه به معیارهای ارزیابی یا همان خطاهای پیش‌بینی، مدل BATS با معیارهای ارزیابی  $MSE=7/2027$ ،  $RMSE=2/6838$ ،  $MAE=2/1587$  و  $MAPE=0/1890$  می‌باشد که برای آینده از این روش برای پیش‌بینی موارد تجمعی بستری کووید-۱۹ در شهرستان بافق استفاده می‌کنیم.

بهترین روش و مدل برای موارد تجمعی فوت کووید-۱۹ شهرستان بافق با توجه به معیارهای ارزیابی یا همان خطاهای پیش‌بینی، مدل هموارسازی اسپلاین مکعبی با معیارهای ارزیابی  $MSE=0/1106$ ،  $RMSE=0/3326$ ،  $MAE=0/2918$  و  $MAPE=0/2072$  می‌باشد که برای آینده از این روش برای پیش‌بینی موارد تجمعی فوت کووید-۱۹ در شهرستان بافق استفاده می‌کنیم.

**جدول ۲:** معیارهای ارزیابی مدل‌های مربوط به موارد تجمعی بستری کووید-۱۹ در شهرستان بافق

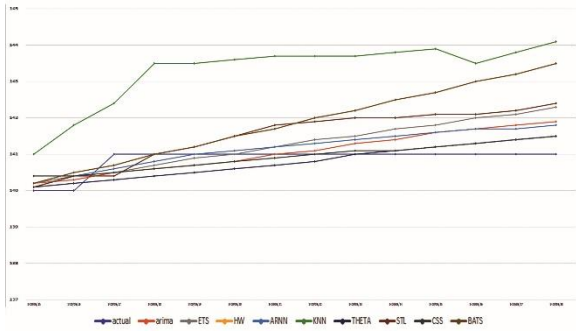
مدل	MSE	RMSE	MAE	MAPE
ARIMA	۸/۷۰۸۴	۲/۹۵۱۰	۲/۴۳۴۹	۰/۲۱۳۲
هموارسازی نمایی	۸/۷۸۰۵	۲/۹۶۳۲	۲/۴۴۷۰	۰/۲۱۴۲
هولت-وینترز	۸/۸۸۵۷	۲/۹۸۰۹	۲/۴۶۴۰	۰/۲۱۵۷
شبکه عصبی اتورگرسیو	۱۳۳/۱۳۱۸	۱۸/۲۵۱۹	۱۶/۷۷۸۶	۱/۴۶۹۳
رگرسیون KNN	۱۹/۳۶۶۱	۴/۴۰۰۷	۳/۸۵۶۹	۰/۳۳۷۶
تتا	۱۸۱/۷۱۸۴	۱۳/۴۸۰۳	۱۲/۴۲۹۵	۱/۰۸۸۵
STL	۸/۴۹۳۷	۲/۹۱۴۴	۲/۴۷۸۱	۰/۲۱۷۴
اسپلاین مکعبی	۷/۲۰۹۲	۲/۶۸۵۰	۲/۳۱۸۴	۰/۲۰۳۰
BATS	۷/۲۰۲۷	۲/۶۸۳۸	۲/۱۵۸۷	۰/۱۸۹۰

**جدول ۳:** معیارهای ارزیابی مدل‌های مربوط به موارد تجمعی فوت کووید-۱۹ در شهرستان بافق

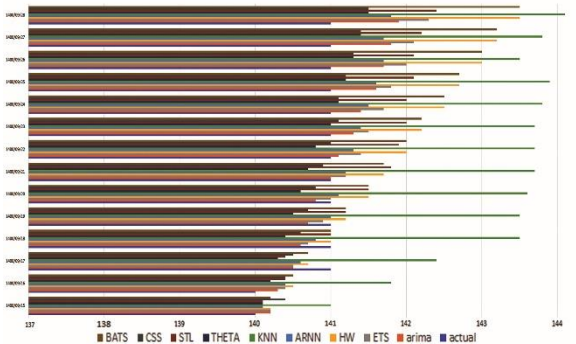
مدل	MSE	RMSE	MAE	MAPE
ARIMA	۰/۲۶۱۳	۰/۵۱۱۲	۰/۴۲۹۸	۰/۳۰۵۰
هموارسازی نمایی	۰/۴۶۴۵	۰/۶۸۱۶	۰/۵۵۹۸	۰/۳۹۷۲



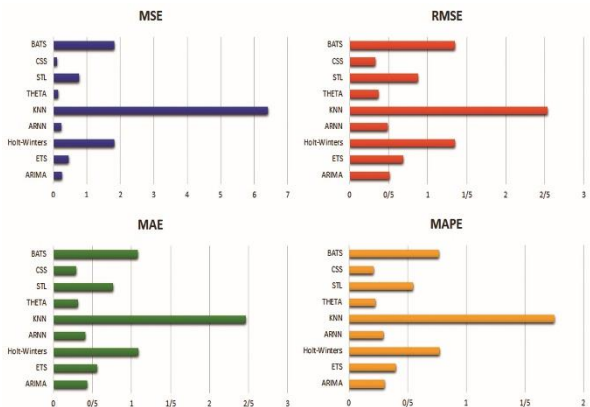
۳- و می‌توان از نتایج حاصله برای پیش‌بینی و به کنترل درآوردن بیماری‌های ویروسی دیگر که ممکن است در آینده به عنوان یک پاندمی و یا حتی اپیدمی و اندمیک ظهور کند، استفاده کرد.



شکل ۵: نمودار پیش‌بینی ۱۴ روزه میزان موارد تجمعی فوت کووید-۱۹ در شهرستان بافق



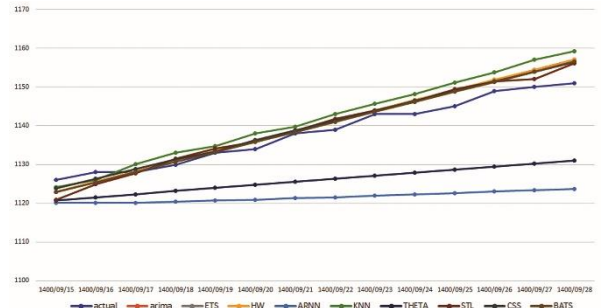
شکل ۶: نمودار میله‌ای پیش‌بینی ۱۴ روزه میزان موارد تجمعی فوت کووید-۱۹ در شهرستان بافق



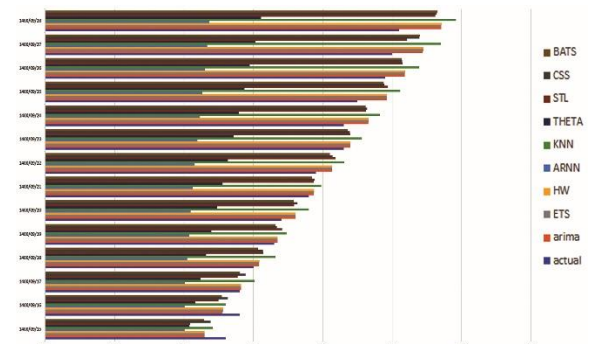
شکل ۷: نمودار معیارهای ارزیابی موارد تجمعی فوت کووید-۱۹ در شهرستان بافق

### نتیجه‌گیری و جمع‌بندی

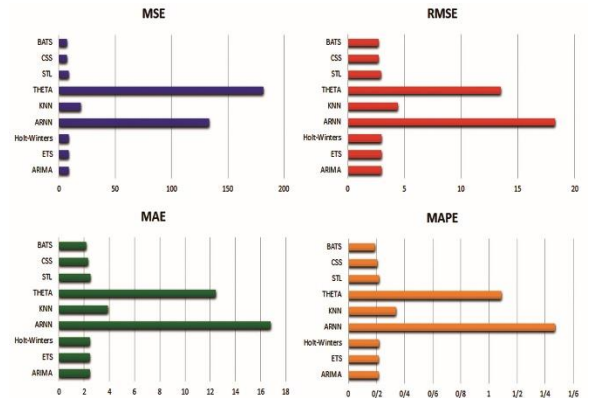
در مطالعه حاضر، از ۹ روش ARIMA، تئاه هموارسازی نمائی، هولت-وینترز، شبکه عصبی اتورگرسیون، رگرسیون KNN، روش BATS، روش هموارسازی اسپلاین مکعبی و روش STL برای پیش‌بینی تعداد موارد تجمعی جدید و مرگ‌ومیر کووید-۱۹ در شهرستان بافق به عنوان یک مطالعه موردی، مورد استفاده و با استفاده از معیارهای ارزیابی MSE، RMSE، MAE و MAPE 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 بافق متفاوت می‌باشند، برای موارد



شکل ۲: نمودار پیش‌بینی ۱۴ روزه میزان موارد تجمعی بستری کووید-۱۹ در شهرستان بافق



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



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

- [11] Ceylan, Z. (2020). Estimation of COVID-19 prevalence in Italy, Spain, and France. *Science of The Total Environment*, 729, 138817.
- [12] Bayyurt, L., & Bayyurt, B. (2020). Forecasting of COVID-19 cases and deaths using ARIMA models. *medrxiv*. <https://doi.org/10.1101/2020.04.17.20069237>
- [13] Tandon, H., Ranjan, P., Chakraborty, T., & Suhag, V. (2020). Coronavirus (COVID-19): ARIMA based time-series analysis to forecast near future. *arXiv preprint arXiv:2004.07859*.
- [14] Perone, G. (2020). An ARIMA model to forecast the spread of COVID-2019 epidemic in Italy. *arXiv preprint arXiv:2004.00382*.
- [15] Ghosal, S., Sengupta, S., Majumder, M., & Sinha, B. (2020). 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(4), 311-315.
- [16] Parbat, D., & Chakraborty, M. (2020). A python based support vector regression model for prediction of COVID19 cases in India. *Chaos, Solitons & Fractals*, 138, 109942.
- [17] Naimoli, A. (2022). Modelling the persistence of Covid-19 positivity rate in Italy. *Socio-Economic Planning Sciences*, 101225.
- [18] Kibria, H. B., Jyoti, O., & Matin, A. (2022). Forecasting the spread of the third wave of COVID-19 pandemic using time series analysis in Bangladesh. *Informatics in Medicine Unlocked*, 28, 100815.
- [19] Khan, F. M., & Gupta, R. (2020). ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. *Journal of Safety Science and Resilience*, 1(1), 12-18.
- [20] Chowdhury, A. A., Hasan, K. T., & Hoque, K. K. S. (2021). Analysis and prediction of COVID-19 pandemic in Bangladesh by using ANFIS and LSTM network. *Cognitive Computation*, 13(3), 761-770.
- [21] Alassafi, M. O., Jarrah, M., & Alotaibi, R. (2022). Time series predicting of COVID-19 based on deep learning. *Neurocomputing*, 468, 335-344.
- [22] Chyon, F. A., Suman, M. N. H., Fahim, M. R. I., & Ahmed, M. S. (2022). Time series analysis and predicting COVID-19 affected patients by ARIMA model using machine learning. *Journal of Virological Methods*, 301, 114433.
- [23] Nair, S., Ckm, G., Varsha, R., Ghosal, S., Vergin, M., & Anbarasi, L. J. (2022). Intelligent Forecasting Strategy for COVID-19 Pandemic Trend in India: A Statistical Approach. In *Artificial Intelligence and Technologies* (pp. 553-560). Springer, Singapore.
- [24] Mukhairez, H. H., & Alaff, A. J. (2022). Short-term Forecasting of COVID-19. In *Computational Intelligence for COVID-19 and Future Pandemics* (pp. 257-266). Springer, Singapore.
- [25] Prajapati, S., Swaraj, A., Lalwani, R., Narwal, A., Verma, K., Singh, G., & Kumar, A. (2021). Comparison of Traditional and Hybrid Time Series Models for Forecasting COVID-19 Cases. *arXiv preprint arXiv:2105.03266*.
- [26] Li, C., Sampene, A. K., Agyeman, F. O., Robert, B., & Ayisi, A. L. (2022). Forecasting the Severity of COVID-19 Pandemic Amidst the Emerging SARS-CoV-2 Variants: Adoption of ARIMA Model.

تجمعی بستری در بافق بهترین مدل، BATS می‌باشد، درحالیکه برای موارد تجمعی فوت مدل هموارسازی اسپلاین مکعبی 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 بررسی شود.
- پیش‌بینی کووید-۱۹ با توجه به اثر رعایت فاصله اجتماعی بررسی شود
- پیش‌بینی کووید-۱۹ با توجه به اثر واکسیناسیون علیه کووید-۱۹ بررسی شود
- از روش‌های استفاده شده برای دیگر بیماری‌های واگیردار یا پاندمیک برای پیش‌بینی استفاده شود.
- این روش‌ها را می‌توان در مکانی دیگر مدلسازی و نتیجه‌گیری کرد.

## مراجع

- [1] Cucinotta, D., & Vanelli, M. (2020). WHO declares COVID-19 a pandemic. *Acta Bio Medica: AteneiParmensis*, 91(1), 157
- [2] "Q&A on coronaviruses (COVID-19)". World Health Organization. Retrieved 11 March 2020.
- [3] W.H. Organization, WHO announces COVID-19 outbreak a pandemic, 2020, Available from: <https://www.euro.who.int/en/health-topics/healthemergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic>
- [4] Cheng, Z. J., & Shan, J. (2020). 2019 Novel coronavirus: where we are and what we know. *Infection*, 48(2), 155-163.
- [5] T. Tran, L. Pham, Q. Ngo, Forecasting epidemic spread of SARS-CoV-2 using ARIMA model (case study: Iran), *Glob. J. Environ. Sci. Manag.* 6 (2020) 1-10, (Special Issue (Covid-19))
- [6] Zhang, X., Liu, Y., Yang, M., Zhang, T., Young, A. A., & Li, X. (2013). Comparative study of four time series methods in forecasting typhoid fever incidence in China. *PloS one*, 8(5), e63116.
- [7] Chen, Y., Leng, K., Lu, Y., Wen, L., Qi, Y., Gao, W., ... & Dong, J. (2020). Epidemiological features and time-series analysis of influenza incidence in urban and rural areas of Shenyang, China, 2010-2018. *Epidemiology & Infection*, 148.
- [8] Olayemi, O. S., Oluwatosin, O. E., & Segun, O. E. (2020). Time Series Analysis on Reported Cases of Tuberculosis in Minna Niger State Nigeria. *Open Journal of Statistics*, 10(3), 412-430.
- [9] Nayak, M. S. D. P., & Narayan, K. A. (2019). Forecasting dengue fever incidence using ARIMA analysis. *International Journal of Collaborative Research on Internal Medicine & Public Health*, 11(3), 924-932.
- [10] Wu, W., An, S. Y., Guan, P., Huang, D. S., & Zhou, B. S. (2019). Time series analysis of human brucellosis in mainland China by using Elman and Jordan recurrent neural networks. *BMC infectious diseases*, 19(1), 1-11.

- [31] Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: a decomposition approach to forecasting. *International journal of forecasting*, 16(4), 521-530.
- [32] Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition. *J. Off. Stat.*, 6(1), 3-73.
- [33] Hyndman, R. J., King, M. L., Pitrun, I., & Billah, B. (2005). Local linear forecasts using cubic smoothing splines. *Australian & New Zealand Journal of Statistics*, 47(1), 87-99.
- [34] De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American statistical association*, 106(496), 1513-1527.
- Computational and Mathematical Methods in Medicine*, 2022.
- [27] R.A. Yaffee, M. McGee, An Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS , Elsevier, 2000
- [28] B. Fanoodi, B. Malmir, F.F. Jahantigh, Reducing demand uncertainty in the platelet supply chain through artificial neural networks and ARIMA models, *Comput. Biol. Med.* 113 (2019) 103415.
- [29] Kufel, T. (2020). ARIMA-based forecasting of the dynamics of confirmed Covid-19 cases for selected European countries. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 15(2), 181-204.
- [30] Martínez, F., Frías, M. P., Charte, F., & Rivera, A. J. (2019). Time Series Forecasting with KNN in R: the tsfknn Package. *R J.*, 11(2), 229.