Comparison of the predictive accuracy of hybrid models of NNAR, TBATS, ETS-EANN, and ARIMA for COVID-19 incidences in Saudi Arabia

Ahmad Ghaleb Al-Rabbasi [(1)]; Faisal A.M Ali [(1)];
Fahmy Al–Salwi [(1)]; Ahmed Abdel -Razzaq Alansari [(1)]

[(1)] Department of Data Science and Information Technology, University of Taiz, College of Administrative Sciences, Taiz, Yemen.

Abstract:

COVID-19 is a strain of the coronavirus that causes sickness. It first showed up in December 2019 and caused a world pandemic that is still going on. There is a vital need to monitor the spread of COVID-19 and to be able to predict where the pandemic will go so as to control its spread in a more effective way. COVID-19 data is an example of time-series data that can be used in different ways to make predictions. Even though there are many time-series forecasting models, it is hard to make broad general statements about which ones are better. This paper gives an empirical review of several time-series models for predicting COVID-19 cases, in Saudi Arabia, where Arab countries like Arab Gulf countries, lack studies in diagnosing this aspect. In particular, this research contributes to the literature on Time-Series Data forecasting in a variety of ways through outcomes of a variety of forecasting strategies using series forecasting techniques. These include hybrid combinations of an exponential smoothing state-space model ETS, EANN, autoregressive -integrated moving-average (ARIMA), a neural network autoregressive model (NNAR), and TBATS. Thus, they are eleven top combination architectures of the: (E-N, E-N-F, N-F, N-T, A-N, A-N-F, N-T-F, A-E-N, E-N-T, A-N-T, and A-N-E-F). The models were made with daily COVID-19 data from Saudi Arabia from Jan 2022 to May 30, 2022, which was available to the public. The best forecasting model was selected based on the high accuracy measure of root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE). The testing results show that based on the current data findings of this investigation, the E-N, E-N-F, and N-F models had the lowest forecasting error.

Keywords: Hybrid models forecasting; ARIMA; ETS; TBATS NNAR F, COVID-19

مقارنة الدقة التنبؤية للنماذج الهجينة من NNAR وTBATS وETS-EANN و ARIMA لانتشار حالات COVID -19 في المملكة العربية السعودية

أ.د أحمد غالب الرباصي، د. فيصل علي مُحَمَّد علي، د. فهمي الصلوي، د. أحمد عبد الرزاق الأنصاري

الملخص:

في ديسمبر ٢٠١٩، ظهر مرض فيروس كورونا (COVID-19) وتسبب في جائحة عالمية لا تزال خطورتها مستمرة. هناك حاجة ملحة لمراقبة انتشار COVID-19 والتنبؤ به للسيطرة على هذا الانتشار بشكل أكثر فعالية. تعد نماذج السلاسل الزمنية مهمة في التنبؤ بتأثير تفشي COVID-19 واتخاذ التدابير اللازمة للاستجابة لهذه الأزمة. على الرغم من وجود عدد من نماذج التنبؤ بالسلاسل الزمنية، فإنه من الصعب إصدار بيانات عامة واسعة عن أيها الأفضل. تقدم هذه الورقة مراجعة تجريبية لعدد من نماذج السلاسل الزمنية للتنبؤ بحالات انتشار COVID-19 في المملكة العربية السعودية، إذ تفتقر الدول العربية إلى دراسات في تشخيص هذا الجانب، وعلى وجه الخصوص يسهم هذا البحث في الأدبيات المتعلقة بالتنبؤ ببيانات السلاسل الزمنية بطرق متعددة، تتضح من خلال مقارنة نتائج مجموعة متنوعة من أساليب التنبؤ باستخدام تقنيات التنبؤ المتسلسل. وتشمل هذه المجموعات نماذج هجينة من النماذج الأُسِيَّة ETS و TBATS ونموذج ARIMA المتوسط المتحرك الانحدار الذاتي ونموذج NNAR الانحدار الذاتي للشبكة العصبية، وبالتالي تم صياغة أحد عشر نموذجًا هجينًا للجمع بين النماذج المتمثلة في (E-N ، و E-N-F، و N-F، و N-T، و A-N، و A-N-F، و N-T-F ، و A-E-N، و E-N-T ، و A-N-T ، وA-N-E-F). كما تم جمع البيانات اليومية عن انتشار حالات COVID-19 في المملكة العربية السعودية من يناير ٢٠٢٢ إلى ٣٠ مايو ٢٠٢٢ من موقع منظمة الصحة العالمية. وتم اختيار أفضل نموذج تنبؤ بناءً على معايير المقارنة المتمثلة في جذر متوسط الخطأ التربيعي (RMSE) ومتوسط النسبة المئوية للخطأ المطلق (MAPE) ، ومتوسط الخطأ المطلق .(MAE) تظهر نتائج هذه الدراسة أن أفضل النتائج هي عن النماذج المتمثلة في E-N و E-N-F و N-F إذ كان لها أدنى خطأ في التنبؤ.

الكلمات المفتاحية: تنبؤات، النماذج الهجينة-ARIMA; ETS; TBATS NNAR F, COVID