

استفاده از روش های علم داده جهت فراخوانی داده های ثبتي و رسمي، مدل سازی و پیش بینی

مهدی جباری نوقابی¹، دانشیار گروه آمار دانشگاه فردوسی مشهد

چکیده:

در این مطالعه ضمن مروری بر روش های آماری برای جمع آوری داده ها و آمارگیری، به اهمیت و ضرورت استفاده از روش های علم داده برای فراخوانی داده های رسمي و ثبتي پرداخته می شود. سپس به کمک **Application Programming Interface (API)** به بررسی چگونگی فراخوانی داده های ویروس کرونا و نیز داده های رمز ارزها از پایگاه های مختلف داده می پردازیم. در ادامه با استفاده از روش های مختلف آماری سری زمانی شامل مدل های **ARIMA**، **NNETAR**، **ARFIMA**، **Holt-Winter**، **Hybrid**، **BSTS**، **Prophet**، **TBATS** و مدل های شبکه عصبی مانند **MLP** و **ELM** به مدل سازی برخط داده ها پرداخته و پس از آزمایش و آزمون نتایج مدل سازی و مقایسه آن ها با یکدیگر به کمک شاخص های نیکویی برازش مدل و شاخص های مبتنی بر خطای مدل، بهترین مدل انتخاب می گردد. سپس بر مبنای مدل منتخب، پیش بینی برای دوره های زمانی بعدی انجام گرفته و در مورد اقدام بعدی شامل خرید یا فروش یا نگهداری رمز ارز در کیف پول دیجیتال، تصمیم گیری صورت گرفته و به صورت برخط اقدام می گردد. در واقع، تمام موارد فوق تحت عنوان یک ربات عمل لازم و مناسب را در زمان مقتضی انجام می دهد. لازم به ذکر است که برخی از موارد بیان شده به کمک برنامه نویسی در نرم افزار **R** و برخی به کمک زبان برنامه نویسی **Python** انجام شده است. برای استفاده مناسب تر و آسان تر از موارد فوق، کدهای مربوطه با استفاده از بسته شایینی در نرم افزار **R** به اپلیکیشن های مناسب تبدیل می گردد.

کلمات کلیدی: فراخوانی برخط داده ها، مدل سازی، پیش بینی، ربات، اپلیکیشن

¹ E-mail: jabbarinm@um.ac.ir; jabbarinm@yahoo.com

Using data science methods to call online data, modeling and forecasting

Mehdi Jabbari Nooghabi*,

Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran

Abstract:

In this study, while reviewing statistical methods for data collection and statistical surveys, the importance and necessity of using data science methods to call official and registered data/online data is discussed. Then, with the help of Application Programming Interface (API), we will examine how to call corona virus (COVID-19) data as well as crypto-currency data from different databases. Then using various time series statistical methods including ARIMA, NNETAR, ARFIMA, Holt-Winter, Hybrid, BSTS, TBATS, Prophet and neural network models such as MLP and ELM try to fit online models on the data and after training and testing the modeling results and comparing them with each other by using goodness of fit measures as well as residual measures, the best model is selected. Then, based on the selected model, forecasts are made for the next time periods, and decisions are made about the next step, including buying or selling or keeping the crypto-currency in the digital wallet and the action is taken online. In fact, the entire above are prepared as a robot to perform the necessary and appropriate actions at the same time. It should be noted that some of the above cases have been done by R software and some with the help of Python programming language. For more appropriate and easier use of the above, the relevant codes are converted into appropriate applications using the well-known Shiny package in R software.

Keywords: Calling online data, Modeling, Forecasting, Robot, Application.

* E-mail: jabbarinm@um.ac.ir; jabbarinm@yahoo.com