APPLICATION OF MACHINE LEARNING METHODS FOR TIME SERIES FORECASTING

Kapusta Darina, student of group 345
National Aerospace University "Kharkiv Aviation Institute"

Forecasting is the ability to predict what will happen in the future on the basis of analysis of past and current data. Time series forecasting is considered one of the most applied data science techniques that is used in different industries such as finance, supply chain management, production, and inventory planning. The predictive models based on machine learning found wide implementation in time series projects required by various businesses for facilitating predictive distribution of time and resources.

Machine learning forecasting methods have been used to predict weather conditions for a rather long time, relying on data from weather stations and satellites, while predicting the growth of epidemics is based solely on the human factor, and still requires a lot of effort to improve. Therefore, in the light of the events of the last year, we have developed a software product for predicting the dynamics of the growth of epidemics of various diseases.

Infectious disease is a leading threat to public health, economic stability, and other key social structures. Efforts to mitigate these impacts depend on accurate and timely monitoring to measure the risk and progress of disease. Traditional, biologically-focused monitoring techniques are accurate but costly and slow. That`s why, our goal is to create a simple, understandable and accessible application for everyone, which won`t only help medical workers make decisions to stop the spread of infection, but also show other people the possible alignment of events if they don`t adhere to the rules to eliminate the growth of the epidemic.

To automate the forecasting process in this work, the following forecasting methods were used: linear regression, polynomial regression, support vector machine, Holt's linear trend method and also a forecasting model called AUTO ARIMA. We have used several methods in order to make the forecast results more accurate, as none of the methods can give an absolutely accurate result. For implementation of our program we used Python, as it consistently ranks first in world research on working with data. Python contains a large number of built-in libraries such as SciKit-Learn, SciPy, pandas, which are widely used in machine learning while working with large datasets.

Epidemic forecasting is still in its infancy and is a growing field with great potential. Although predictive algorithms cannot eliminate medical uncertainty, they are already improving the allocation of scarce medical resources, helping to prevent patients from being hospitalized and prioritizing patients fairly. Early warning systems, which once took years to build, can now be quickly developed and optimized based on real data, just as deep learning neural networks regularly provide state-of-the-art image recognition capabilities previously considered impossible.