Methodological Approaches to Sentiment Classification and Their Impact on Modeling the Relationship Between Twitter (X) and the Stock Market

Joanna Michalak

Nicolaus Copernicus University in Toruń – joanna.michalak@umk.pl

Abstract

In recent years, there has been a development in the research and modelling of the relationship between social media and the stock market. To conduct an analysis of the interdependence between them, it is necessary to build a system of indicators representing sentiment. Methods of sentiment analysis used for this purpose include supervised machine learning approaches, dictionary-based approaches, and neural networks. These various methodologies exhibit different levels of effectiveness, resulting primarily in distinct patterns in time series data. As a result, varying conclusions about the causality between social media and the stock market can be drawn. The aim of this paper is to summarize these three approaches in terms of their strengths and limitations and demonstrate their impact on the process of modelling the interdependence between Twitter and the stock market. The relationship between the systems was established by using VAR and Granger Causality. Sentiment analysis was conducted using the following methods: supervised machine learning (MNB), dictionary-based approaches (NRC, VADER), and neural networks (LSTM). The results unequivocally lead to different conclusions regarding the short-term dynamics analysis between the systems. The study presented in the article is of a preliminary nature.