

Chapter 6

Using Online Data in Predicting Stock Price Movements: Methodological and Practical Aspects

František Dařena

Mendel University in Brno, Czech Republic

Jonáš Petrovský

Mendel University in Brno, Czech Republic

Jan Přichystal

Mendel University in Brno, Czech Republic

Jan Žižka

Mendel University in Brno, Czech Republic

ABSTRACT

A lot of research has been focusing on incorporating online data into models of various phenomena. The chapter focuses on one specific problem coming from the domain of capital markets where the information contained in online environments is quite topical. The presented experiments were designed to reveal the association between online texts (from Yahoo! Finance, Facebook, and Twitter) and changes in stock prices of the corresponding companies. As the method for quantifying the association, machine learning-based classification was chosen. The experiments showed that the data preparation procedure had a substantial impact on the results. Thus, different stock price smoothing, the lags between the release of documents and related stock price changes, levels of a minimal stock price change, different weighting schemes for structured document representation, and classifiers were studied. The chapter also shows how to use currently available open source technologies to implement a system for accomplishing the task.

DOI: 10.4018/978-1-5225-5586-5.ch006

INTRODUCTION

A lot of research has been focusing on incorporating the vast amount of data available online into models of various social and economic phenomena in miscellaneous fields of science. The data, which is generated not only by domain experts but also by regular people, can provide new perspectives and potentially complementary information to conventional quantitative and objective evidence (Kearney & Liu, 2014). The data often comes from non-traditional and contemporary information sources and environments, like social networks or microblogging sites. The data typically has a form of unstructured texts that are published by different types of subjects, without the time and spatial limits.

On one hand, there are many logical reasons for this trend: there is a lot of data freely available, many different sources can be combined together, the data is often accessible immediately, in real time, and the variety of data items, opinions, and perspectives can be great. On the other hand, collecting, storing, and processing such data relate to several problems: tools for collecting the data need to be available, it is necessary to handle multiple systems and data formats, it is necessary to consider differences in data quality, reliability, objectivity etc.

One of the domains, where using online data is very topical and plays an important role in analyzing the studied systems, is the field of capital markets. Here, the data provided by digital media can help, e.g., in explaining less rational factors such as investors' sentiment or public mood as influential for asset pricing and capital market volatility (Bukovina, 2016).

One of the greatest advantages of using online resources for decision making support in the domain of capital markets is the timeliness of the information, which is particularly important for investment decisions. The quality of the messages posted in online environments (such as microblogs or discussions in social networks) is, however, generally low. That is why Internet postings have been the least frequently studied source of textual sentiment (Kearney & Liu, 2014). Despite all difficulties, content generated by web users has become a widely accepted resource for determining sentiment or opinions related to different aspects of the public mood (Tumasjan et al., 2011). It has been also shown that a large number of people participating in a content generation process enables the creation of artifacts that are of equal or superior quality than those made by experts in the respective field (Gottschlich & Hinz, 2014). Messages from millions of people are also unlikely to be biased (Mostafa, 2013).

There exist numerous studies focusing on the usefulness of textual data for predictions related to stock prices. Majority of them focused on an aggregate level (e.g., the level of a stock market index), worked with a single source of texts (e.g., newspaper articles or financial reports), or required additional expert knowledge

33 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/using-online-data-in-predicting-stock-price-movements/208383

Related Content

Educating Competitive Professionals: A Requirement of the Modern Urban Economy

Olena Lvivna Ilienکوand Oksana Oleksiivna Rezvan (2023). *Considerations of Territorial Planning, Space, and Economic Activity in the Global Economy* (pp. 47-68). www.irma-international.org/chapter/educating-competitive-professionals/317813

Theory of Shocks About States in Static and Dynamic Economics

(2021). *Theory of Shocks, COVID-19, and Normative Fundamentals for Policy Responses* (pp. 202-240). www.irma-international.org/chapter/theory-of-shocks-about-states-in-static-and-dynamic-economics/278447

Forecasting Practices in Textile and Apparel Export Industry: A Systematic Review

Adeel Shah, Rizwan Matloob Ellahi, Urooj Nazirand Musawir Ali Soomro (2022). *International Journal of Circular Economy and Waste Management* (pp. 1-17). www.irma-international.org/article/forecasting-practices-in-textile-and-apparel-export-industry/288501

Reconstruction of the Old Institutions: Crisis, Governance, and the G20

Ugur Yasin Asaland Nazm Ekren (2020). *Comparative Approaches to Old and New Institutional Economics* (pp. 41-56). www.irma-international.org/chapter/reconstruction-of-the-old-institutions/245320

The Resource and Leagile Strategy Model for Apparel Export Enterprises: A Proposed Model to Mitigate COVID-19 Uncertainties

Adeel Shah, Che Rosmawati Binti Che Matand Alisa Ibrahim (2022). *International Journal of Circular Economy and Waste Management* (pp. 1-14). www.irma-international.org/article/the-resource-and-leagile-strategy-model-for-apparel-export-enterprises/288502