

Accepted Manuscript

Title: Assessing the usefulness of online message board mining in automatic stock prediction systems

Author: id="aut0005" orcid="0000-0002-9611-7371"
author-id="S1877750317300091-
88f60665c4634a0dc4cc12e0829199ed"
biographyid="vt0005"> Ramiro H. Gálvez id="aut0010"
author-id="S1877750317300091-
b98fdd172fa9fe0cb043e6292ab4ac32"
biographyid="vt0010"> Agustín Gravano



PII: S1877-7503(17)30009-1
DOI: <http://dx.doi.org/doi:10.1016/j.jocs.2017.01.001>
Reference: JOCS 601

To appear in:

Received date: 23-8-2016
Revised date: 7-11-2016
Accepted date: 4-1-2017

Please cite this article as: Ramiro H. Gálvez, Agustín Gravano, Assessing the usefulness of online message board mining in automatic stock prediction systems, *Journal of Computational Science* (2017), <http://dx.doi.org/10.1016/j.jocs.2017.01.001>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Assessing the usefulness of online message board mining in automatic stock prediction systems

Ramiro H. Gálvez^{a,*}, Agustín Gravano^{a,b}

^a*Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina*

^b*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina*

Abstract

We provide evidence of the usefulness of exploiting online text data in stock prediction systems. We do this by mining a popular Argentinian stock message board and empirically answering two questions. First, is there information in the online stock message board useful for predicting stock returns? Second, if useful information is found, is it novel or it is simply a different way of expressing information already available in the past behavior of stock prices?

To address these questions, we build and validate a series of predictive models using state-of-the-art machine learning and topic discovery techniques. Running experiments in which the models are trained with different combinations of features extracted from the past behavior of stock prices, or mined from the online message boards.

Evidence suggests that it is possible to extract predictive information from stock message boards. Furthermore, we find that adding this information improves the performance of classification systems trained solely on technical indicators. Our results suggest that information from online text data is complementary to the one available in the past evolution of stock prices. Additionally, we find that highly predictive features derived from the message board data seem to have an important and relevant semantic content.

Keywords: Stock market, Text mining, Latent semantic analysis, Ridge regression, Random forest

*Corresponding author

Email addresses: rgalvez@dc.uba.ar (Ramiro H. Gálvez), gravano@dc.uba.ar (Agustín Gravano)

Download English Version:

<https://daneshyari.com/en/article/4951066>

Download Persian Version:

<https://daneshyari.com/article/4951066>

[Daneshyari.com](https://daneshyari.com)