



TF-IDF و TFDF در تحلیل مالی

میثم هاشمی^{*}، مهران رضایی^{*}، مرجان کاندی^آ

^آ گروه مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان، ایران.

چکیده

تحلیل متنی در زمینه بیزنس به تکنیک های بازیابی اطلاعات وابسته است. این تکنیک ها در حوزه پیش بینی مالی موفق نبوده و قابل توسعه نیستند. پیشنهاد ما توسعه تکنیک های درونزاد در حوزه مالی جهت پردازش متن است بویژه در مرحله وزن دهی کلمات چرا که روش های استاندارد نامناسب می باشند. بدین منظور ما دو مساله را پیگیری میکنیم. اول، هیوریستیک های برتر بازیابی اطلاعات را به دقت بررسی میکنیم و مشاهده میکنیم که TF-IDF یک روش سهل الوصول است که در پیش بینی گرایش تزلزل نشان میدهد و مقادیر بالای خطا بر روی اعداد بزرگ تولید میکند. دوم، بر روی یک هیوریستیک جدید کار میکنیم که نیازهای تحلیل مالی را مرتفع کند. بدین منظور، رابطه بین نرخ انتشار اطلاعات و اهمیت آنها را بعنوان هیوریستیک جدید در نظر میگیریم. این هیوریستیک نتایج بی همتا در پیش بینی گرایش و در معیارهای سنجش دقت به ارمغان می آورد. افزون بر این، روش پیشنهادی را با استفاده از الگوریتم ژنتیک بهبود داده و نتایج دقیق تری بدست می آوریم. در مقایسه با TF-IDF هیوریستیک پیشنهادی ۵.۳۸ درصد دقت بیشتری در معیارهای سنجش دقت نشان میدهد که این مقدار ۴۶.۱۶ درصد دیگر با استفاده از الگوریتم ژنتیک کاهش می یابد. در انتها، دستاورد این پژوهش نشان میدهد که تحلیل متن مالی و بازیابی اطلاعات باید هر کدام مسیر جداگانه خویش را طی کنند. © 2023. تمامی حقوق محفوظ است.

اطلاعات مقاله

تاریخچه مقاله:

دریافت: 10 May 2023

اصلاح: 18 Aug 2023

پذیرش: 21 Aug 2023

انتشار آنلاین: 09 Oct 2023

کلمات کلیدی:

تحلیل متن مالی، وزن دهی واژه ها، الگوریتم ژنتیک، بازار سهام.

* نویسنده مسئول.

آدرس های رایانامه: meisamhashemi@gmail.com (م. هاشمی)،
mhr.rezaei@gmail.com (م. رضایی)، kaedi@eng.ui.ac.ir (م. کاندی)

تمامی حقوق محفوظ است. © 2023 ISSN: 2322-4460

