

トレンドキーワードのテキストマイニングを用いた 今後株価上昇が見込まれる企業の予測

情報科学専門学校 実践AI科 1年 ©須崎涼、情報セキュリティ学科 1年 村田創信 横溝天城、
先端ITシステム科 1年 高木克海、ビジネス科 1年 井出愛琉

目標：テキストマイニングを使って伸びる株の情報をキャッチする

1月頃から新型コロナウイルスの影響により、コロナ禍では欠かせないマスク、ワクチンなどの衛生用品や、Stay Homeに伴いゲーム等が品薄になるなどし、トレンドによって大きく株価を伸ばした企業が多くみられる。
そこでトレンドキーワードに対してテキストマイニングを用いれば、このトレンドの初動や株価上昇の傾向をいち早くキャッチできるのではないかと考えた。



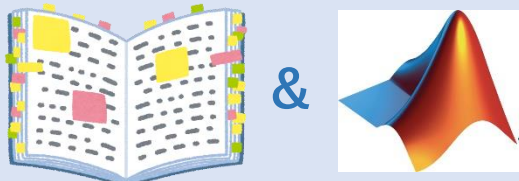
1. 各種媒体からトレンドが反映されそうなテキストデータを集める

即時性の高いTwitterやGoogle Trends、また、Google Newsからトレーダーが注目すると考えられる記事を、Twitter API、スクレイピングなどを用いて収集する。



2. Text Analytics Toolbox™を用いてテキストデータ前処理、解析を行う

Text Analytics Toolbox™を用いて、ドキュメントのトークン化、形態素解析を用いた品詞タグ付け、句読点の除去、ストップワードの除去、単語の正規化を行いテキストの分析を行えるように前処理を行う。



```
documents = normalizeWords(documents);
documents(1:10)

ans =
10x1 tokenizedDocument:
33 tokens: シャープ 97 長崎大学 協力 プラズマクラスター 技術 搭載 ウイルス
33 tokens: 同 試験 装置 長崎大学 感染 症 共同 研究 拠点 安田 朗 教授 素 熱
3 tokens: 試験 装置 イメージ
41 tokens: 同 実証 実験 約 3L 試験 空間 容器 ウイルス 感染 細胞 調製 ウィル
13 tokens: 実験 結果 プラズマクラスターイオン 照射 ない 比較 感染 症 ウィル
9 tokens: プラズマクラスターイオン 照射 による 新型コロナウイルス 感染 症
4 tokens: 噴霧 試験 フラワー 画像
34 tokens: 同社 世界 第三者 試験 機関 共同 2000 約 20 にわたり プラズマクラ
59 tokens: 安田 朗 教授 実証 実験 結果 受ける 付着 ウイルス 対策 として アノ
0 tokens:
```

図1)Text Analytics Toolbox™による前処理

3. 前処理、解析したデータを分析する

センチメント分析やLDA、単語の出現回数などを元に分析する。センチメント分析では単語ごとに感情スコアを計算し、それを用いて文全体がポジティブな内容なのか、ネガティブな内容なのか判別する。
LDAではテキストデータをトピックごとに分類する。



図2)LDAを使った分類

4. これらの分析から株価が上昇するであろう企業を予測する

これらの分析で分かったトレンドから株価の上昇が見込まれる企業を予測する。

まだ計画段階なのでアドバイスや過去の事例など教えていただけたら幸いです！
(右QRコードのGoogleフォームに是非ご意見下さい！)

