**News impact on stock price return via sentiment analysis**

Dampak berita terhadap pengembalian harga saham melalui analisis sentimen

Xiaodong Lia, Haoran Xiea,\*, Li Chenb, Jianping Wanga, and Xiaotie Dengc

1,2Department of Statistics, IPB University, Indonesia

naffiliation, country

‡corresponding author: agusms@apps.ipb.ac.id

Copyright © 20xx Agus M Soleh, Author2, ..., and Authorn. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract**

Artikel berita keuangan diyakini memiliki dampak pada perubahan harga saham. Pekerjaan sebelumnya memodelkan potongan berita dalam ruang kata-kantung (bag-of-words), yang menganalisis hubungan laten antara pola statistik kata dan perubahan harga saham. Namun, sentimen berita, yang merupakan elemen penting dalam hubungan antara pola kata dan pergerakan harga, jarang dipertimbangkan. Dalam makalah ini, kami pertama-tama menerapkan kerangka kerja prediksi harga saham generik, dan menyematkan enam model berbeda dengan pendekatan analisis yang berbeda. Untuk melangkah lebih jauh, kami menggunakan kamus psikologis Harvard dan kamus sentimen keuangan Loughran-McDonald untuk membangun ruang sentimen. Artikel berita teks kemudian diukur secara kuantitatif dan diproyeksikan ke ruang sentimen. Metode pelabelan instan dibahas dan diuji secara ketat. Kami mengevaluasi akurasi prediksi model dan membandingkan kinerja mereka secara empiris pada berbagai tingkat klasifikasi pasar yang berbeda. Percobaan dilakukan pada harga Bursa Saham Hong Kong historis selama lima tahun dan artikel berita. Hasil menunjukkan bahwa (1) pada tingkat saham individu, sektor, dan indeks, model dengan analisis sentimen lebih unggul daripada model kata-kantung baik dalam set validasi maupun set pengujian independen; (2) model yang menggunakan polaritas sentimen tidak dapat memberikan prediksi yang berguna; (3) ada perbedaan minor antara model yang menggunakan dua kamus sentimen yang berbeda.

**Keywords**: Dampak berita, Perubahan harga saham, Sentimen, Prediksi, Eksperimen.

1. **Pendahuluan**

* Penjelasan umum tentang saham, bisa dilengkapi dengan data-data.
* Ada apa di berita/hongkong? Kenapa pilih berita/hongkong? Kenapa pilih berita bahasa inggris dari tahun Januari 2003 sampe maret 2008?
* Intinya cerita dulu dengan singkat dan jelas permasalahannya apa, lalu dikorelasikan dengan metode atau solusinya
* Metode apa yang biasanya dipakai? Kenapa metode itu umum dipakai? Apa kekurangan metode yang biasa dipakai? Apa kelebihan metode yang dipakai di penelitian ini? Jelaskan mengapa metode xxx meodel layak dipakai, sebisa mungkin alasannya merujuk pada jurnal-jurnal atau buku
* Penelitian terdahulu yang sudah pernah dilakukan bagaimana? Apakah ada yang sudah memakai metode itu juga? Atau di yang memilih hongkong? Bagaimana hasilnya?
* Apa yang membedakan antara penelitian ini dengan penelitian2 yang sudah ada?
* Apa yang sebenarnya jadi tujuan di peneltian ini? Apakah memang penelitian ini perlu dilakukan karena belum ada yang meneliti? Atau bagaimana?
* Cerita dulu permasalahannya apa, lalu korelasikan dengan metodenya.
* Merujuk pada sitasi yang memang merujuk pada peramalan
* Yang kita rujuk adalah hasilnya
* Pendahuluan 4-5 paragraf
* Maksimal 5 halaman+1cover untuk proposal
* Tidak dibahas di tinpus tapi disinggung di metode
* Memabuat metodologi dulu, tinjauan pustaka, pendahuluan
* Rujuk jurnal2 yang dibuat departemen stk
* Pernyataan2 tertentu cari jurnal/textbook yang sesuai
* Buku yang lain ketika merujuk pada sebuah rumus

**Paragraf 1** (Penjelasan umum tentang saham dan pentingnya analisis berita):

Pasar saham merupakan bagian penting dan aktif dari pasar keuangan saat ini. Baik investor maupun spekulator di pasar ingin memperoleh keuntungan yang lebih besar dengan menganalisis informasi pasar. Artikel berita keuangan, yang dikenal sebagai salah satu sumber utama informasi pasar, banyak digunakan dan dianalisis oleh investor. Di era Big Data saat ini, jumlah artikel berita terus meningkat secara signifikan. Menghadapi volume berita yang sangat besar, semakin banyak institusi yang mengandalkan kekuatan pemrosesan komputer modern untuk analisis. Prediksi yang diberikan oleh sistem pendukung keputusan dapat membantu investor untuk menyaring noise dan membuat keputusan yang lebih bijak.

**Paragraf 2** (Metode yang biasa digunakan dan kekurangannya):

Pendekatan berbasis bag-of-words memodelkan artikel berita dengan model ruang vektor yang menerjemahkan setiap artikel berita menjadi vektor dari pengukuran statistik kata, seperti jumlah kemunculan, dan sebagainya. Model pembelajaran mesin kemudian digunakan untuk menangkap hubungan antara pola statistik kata dan pergerakan harga saham. Meskipun banyak pendekatan berbasis bag-of-words dilaporkan memiliki kekuatan prediksi dalam beberapa penelitian sebelumnya [14, 41, 42], mereka kehilangan satu mata rantai penting dalam pemetaan dari artikel berita tekstual ke prediksi arah akhir, yaitu sentimen berita. Seperti yang ditunjukkan dalam Gambar 1, satu langkah penting dalam alur adalah bahwa artikel berita terlebih dahulu diinterpretasikan oleh investor dan diterjemahkan menjadi sentimen pasar; investor kemudian membuat keputusan mereka berdasarkan interpretasi sentimen; dan harga pasar mengagregasi tindakan setiap investor dan mencerminkannya dalam pergerakan harga akhir.

**Paragraf 3** (Metode yang diusulkan dan kelebihannya):

Analisis sentimen memodelkan dokumen dari dimensi sentimen. Alih-alih hanya mengukur frekuensi kata, setiap kata (terutama yang berwarna yang memiliki polaritas sentimen) diuraikan dan direpresentasikan oleh vektor fitur sentimen. Membuat prediksi berdasarkan representasi sentimen harus memiliki banyak keuntungan dibandingkan bag-of-words: (1) Dimensi reduktif. Dibandingkan dengan puluhan ribu kata yang biasanya diamati saat menggunakan kata sebagai fitur, representasi sentimen sangat mengurangi dimensi menjadi ratusan, misalnya, kamus psikologi Harvard IV-4 memiliki 182 dimensi dan kamus sentimen keuangan Loughran-McDonald memiliki 6 dimensi; (2) Interpretasi afektif. Biasanya sulit untuk menafsirkan pemetaan yang dihasilkan dalam ruang bag-of-words. Sebaliknya, skor sentimen di berbagai dimensi terkadang dapat memberikan orang perasaan yang lebih langsung tentang dokumen.

**Paragraf 4** (Penelitian terdahulu dan perbedaan dengan penelitian ini):

Meskipun analisis sentimen dokumen telah diusulkan dan digunakan dalam banyak aplikasi, misalnya, sistem rekomendasi [9], belum banyak penelitian yang dilaporkan dalam domain silang ilmu komputer dan perdagangan algoritmik. Dalam penelitian ini, kami pertama kali membangun kerangka kerja generik yang dapat mengambil kumpulan informasi pasar dan memasukkan berbagai model prediksi khusus. Kami kemudian mengimplementasikan model prediksi menggunakan pendekatan analisis sentimen, pendekatan bag-of-words, atau pendekatan polaritas sentimen, dan membandingkan akurasi prediksi pengembalian harga saham harian mereka pada data pasar Hong Kong Stock Exchange selama lima tahun. Untuk menghindari bias yang diperkenalkan oleh kerangka kerja, seperti kamus sentimen, metode pelabelan, dan perbandingan pada tingkat klasifikasi pasar yang berbeda, kami (1) menggunakan dua kamus sentimen untuk perbandingan; (2) mendiskusikan dan menguji secara ketat berbagai metode pelabelan instance; dan (3) membandingkan kinerja model pada tingkat saham individu, sektor, dan indeks dalam eksperimen.

**Paragraf 5** (Tujuan penelitian):

Penelitian ini bertujuan untuk mengembangkan model prediksi harga saham yang lebih akurat dengan mengintegrasikan analisis sentimen ke dalam kerangka kerja prediksi. Dengan memanfaatkan informasi sentimen dari artikel berita keuangan, model yang diusulkan diharapkan dapat menangkap hubungan antara sentimen pasar dan pergerakan harga saham dengan lebih baik dibandingkan pendekatan bag-of-words tradisional. Penelitian ini memberikan kontribusi penting dalam memahami peran analisis sentimen dalam prediksi pasar saham dan menawarkan solusi yang potensial untuk meningkatkan akurasi prediksi harga saham.

1. **Tinjauan Pustaka**

**2.1 Analisis Sentimen**

Analisis sentimen mengacu pada penggunaan pemrosesan bahasa alami, analisis teks, dan linguistik komputasi untuk mengidentifikasi dan mengekstrak informasi subjektif dari sumber data [16]. Dalam analisis sentimen, terdapat beberapa pendekatan utama:

**2.1.1 Pendekatan Berbasis Kamus**

Pendekatan ini menggunakan kamus sentimen yang berisi kata-kata yang dianalisis dan ditandai dengan polaritas atau dimensi sentimen. Contohnya adalah Harvard IV-4 Psychological Dictionary dengan 182 dimensi sentimen [15, 26] dan Loughran-McDonald Financial Sentiment Dictionary untuk domain keuangan [23].

**2.1.2 Pendekatan Statistik Korpus**

Pendekatan ini menggunakan supervised learning pada korpus data berlabel untuk mengekstrak pola statistik dan melatih model klasifikasi sentimen [30, 39].

**2.1.3 Pendekatan Berbasis Konsep**

Pendekatan terbaru ini mengeksploitasi pengetahuan konseptual seperti ConceptNet untuk melakukan analisis sentimen pada level konsep, bukan hanya pada level kata [38, 43, 8, 31].

**2.2 Aplikasi Analisis Sentimen**

Analisis sentimen telah banyak diterapkan pada berbagai aplikasi seperti:

* Rekomendasi produk dengan mengklasifikasi ulasan [39, 30, 24]
* Analisis tren sentimen pada arsip teks [13]
* Memprediksi dampak berita keuangan terhadap pasar [11, 44]
* Klasifikasi sentimen pada media sosial seperti Twitter [28]
* Perbaikan kinerja klasifikasi dengan menyaring kata-kata objektif [20]

**2.3 Analisis Sentimen untuk Prediksi Harga Saham**

Penelitian sebelumnya telah mengeksplorasi penggunaan analisis sentimen berita untuk memprediksi pergerakan harga saham [11, 44]. Namun, kebanyakan hanya menggunakan dimensi sentimen positif dan negatif. Mengintegrasikan representasi multi-dimensi sentimen diharapkan dapat memberikan prediksi yang lebih akurat [penelitian ini]. Selain itu, penelitian terdahulu juga menggunakan pendekatan bag-of-words yang tidak merepresentasikan makna dan sentimen kata secara langsung.

* 1. **Bahan dan Data**

Untuk memasukkan gambar dapat disajikan dan diberikan nomor seperti pada Gambar 1. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Gambar 1: This is sample picture - ShareLaTeX lion.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

* 1. **Metode Penelitian**

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

1. **Hasil dan Pembahasan**

Untuk menyajikan tabel dan penomorannya dilakukan seperti terlihat pada Tabel 1 berikut. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.

Tabel 1: Nilai yang dihasilkan ditampilkan pada tabel.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | A | B | C | D | E |
|  |  |  | F |  | G |
| 77 | 121.58 | 117.19 | 114.05 | 115.05 | 114.23 |
| 78 | 114.22 | 110.12 | 105.88 | 105.20 | 105.66 |
| 79 | 81.10 | 79.04 | 75.94 | 77.86 | 75.94 |

Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

* 1. **Preliminary Notes**

**Definition 3.***1 This is a text of a definition.*

**Theorem 3.2** *This is a text of a theorem.*

|  |  |
| --- | --- |
|  | (1) |

* 1. **Lemma**

**Lemma 3.3** *This is a text of a lemma.*

1. **Simpulan dan Saran**

Pada bagian akhir penulis diwajibkan menulis simpulan yang merupakan simpulan hasil dan pembahasan. Pada bab ini, dapat memasukkan saran untuk pengembangan metode penelitian dengan menyebutkan metodenya secara jelas. Oleh karena itu, bagian saran dalam suatu artikel tidak wajib ada, jika belum diketahui metode apa yang akan disarankan. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

**Ucapan Terima Kasih**. Bagian ucapan terima kasih adalah opsional.Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

**Daftar Pustaka**

SEO SH. 2021. Forecasting Korean LNG import price using ARIMAX, VECM, LSTM and hybrid models [tesis]. Ulsan (KR): Ulsan National Institute of Science and Technology (UNIST).

IGU. (2020). 2020 World LNG Report. International Gas Union.

<https://igu.org/resources/2020-world-lng-report/>

KEMCO. (2019). Korea Energy Outlook 2019. Korea Energy Management Corporation.

<https://policy.asiapacificenergy.org/sites/default/files/KoreaEnergyStandards%26Labeling.pdf>

Lee, S., et al. (2020). Impact of IMO 2020 Sulphur Regulations on Bunker Demand and Prices. Journal of Shipping and Trade, 5(1), 1-15.

<https://www.sciencedirect.com/science/article/pii/S2666822X22000144>

Moon, J., & Jung, W. (2020). Long-run relationship between LNG import price and system marginal price in South Korea: An empirical analysis. Energy Policy, 137, 111123.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.

<https://books.google.co.id/books?hl=en&lr=&id=_bBhDwAAQBAJ&oi=fnd&pg=PA7&dq=Forecasting:+principles+and+practice&ots=TjfVvnYNIK&sig=MDl0DQDr47hl5BwLnliGI1dVST8&redir_esc=y#v=onepage&q=Forecasting%3A%20principles%20and%20practice&f=false>

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)

Johansen, S. (1995). Likelihood-based inference in cointegrated vector autoregressive models. Oxford Bulletin of Economics and statistics.

<https://doi.org/10.1017/S0266466698144067>

Pankratz, A. (2012). Forecasting with dynamic regression models. John Wiley & Sons.

Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. MIS quarterly, 553-572.

<https://www.jstor.org/stable/23042796>

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.

<https://doi.org/10.1016/S0925-2312(01)00702-0>