



## Implementasi Sistem Temu Balik Informasi Data Cryptocurrency Berbasis Atribut Menggunakan *Vector Space Model* (VSM)

Muhammad Ibnu Rayyan <sup>1\*</sup>, Suci Pratiwi <sup>2</sup>, Sofy Ertika Dewi <sup>3</sup>

<sup>1-3</sup> STMIK KAPUTAMA, Indonesia

Email : [ibnurayyan2004@gmail.com](mailto:ibnurayyan2004@gmail.com) <sup>1</sup>, [sucitiwi2612@gmail.com](mailto:sucitiwi2612@gmail.com) <sup>2</sup>, [sofyerikadewi2003@gmail.com](mailto:sofyerikadewi2003@gmail.com) <sup>3</sup>

\*Penulis korespondensi : [ibnurayyan2004@gmail.com](mailto:ibnurayyan2004@gmail.com)

**Abstract.** *This study aims to implement an information retrieval system for cryptocurrency data using an attribute-based approach integrated with the Vector Space Model (VSM). The primary objective is to develop a system capable of retrieving the most relevant digital asset information according to specific search attributes, including positive sentiment, price fluctuation, and prediction confidence level. The research adopts a descriptive qualitative method combined with an experimental approach to evaluate the retrieval performance of the cosine similarity algorithm on normalized numerical data. Data preprocessing and attribute weighting were conducted to ensure consistency and improve retrieval accuracy. The experiment demonstrates that the proposed system achieves a Precision@5 value of 1.0, which indicates that all top-five retrieved results are fully relevant to user queries. These findings validate the effectiveness of the attribute-based VSM in analyzing multidimensional cryptocurrency datasets. Overall, this research contributes to the advancement of information retrieval applications in the cryptocurrency domain, particularly for supporting data-driven decision-making and intelligent financial analysis.*

**Keywords:** *Attribute-Based Retrieval, Cosine Similarity, Cryptocurrency, Information Retrieval System, Vector Space Model*

**Abstrak.** Penelitian ini bertujuan untuk mengimplementasikan sistem temu balik informasi (*information retrieval system*) pada data cryptocurrency menggunakan pendekatan berbasis atribut yang diintegrasikan dengan *Vector Space Model* (VSM). Tujuan utama dari penelitian ini adalah mengembangkan sistem yang mampu menampilkan data aset digital paling relevan berdasarkan atribut pencarian tertentu, seperti sentimen positif, perubahan harga, dan tingkat kepercayaan prediksi. Penelitian ini menggunakan metode deskriptif kualitatif dengan pendekatan eksperimental untuk mengevaluasi kinerja algoritma *cosine similarity* pada data numerik yang telah dinormalisasi. Proses pra-pemrosesan dan pembobotan atribut dilakukan untuk memastikan konsistensi serta meningkatkan akurasi hasil pencarian. Hasil eksperimen menunjukkan bahwa sistem yang diusulkan memperoleh nilai Precision@5 sebesar 1.0, yang berarti seluruh lima hasil pencarian teratas relevan dengan kebutuhan pengguna. Temuan ini membuktikan bahwa model VSM berbasis atribut efektif dalam menganalisis dataset cryptocurrency yang bersifat multidimensi. Secara keseluruhan, penelitian ini berkontribusi pada pengembangan aplikasi temu balik informasi di bidang cryptocurrency, terutama dalam mendukung analisis cerdas dan pengambilan keputusan berbasis data.

**Kata kunci:** *Cosine Similarity, Cryptocurrency, Pendekatan Berbasis Atribut, Precision, Sistem Temu Balik Informasi, Vector Space Model*

### 1. PENDAHULUAN

Kembangnya dunia teknologi informasi yang begitu pesat mendorong semakin bertambahnya kebutuhan terhadap suatu sistem pencarian dan pengambilan informasi yang cepat, akurat, dan aspek-aspek yang relevan. Dalam konteks data keuangan digital seperti cryptocurrency, pengguna memerlukan suatu sistem yang mampu menunjukkan informasi secara menyeluruh, satu diantaranya menyangkut atribut harga, sentimen sosial, volatilitas tingkat pasar, dan kepentingan serta keyakinan terhadap prediksi harga. Hal ini menunjukkan

tanda bahwa kebutuhan terhadap suatu sistem analisis berbasis data numerik memperluas sistem dari yang sudah ada dan menjadikannya layak di implementasikan di era surut ekonomi digital.

Meskipun konsep *Information Retrieval* (IR) telah banyak diterapkan pada data berbasis teks, seperti dokumen, artikel berita, dan konten web, penerapan IR pada data numerik—khususnya dalam domain *cryptocurrency*—masih relatif terbatas. Mayoritas sistem IR konvensional belum mengakomodasi analisis atribut numerik secara optimal, sehingga pengguna kerap mengalami kesulitan dalam menemukan aset digital yang sesuai dengan preferensi spesifik, misalnya koin dengan sentimen positif, tingkat volatilitas stabil, atau perubahan harga signifikan. Keterbatasan ini menimbulkan kesenjangan penelitian (*research gap*) yang dapat dijadikan dasar bagi pengembangan sistem temu balik informasi berbasis atribut.

Selain itu, volatilitas tinggi pada pasar *cryptocurrency* menjadikan proses pengambilan keputusan investasi semakin kompleks. Pengguna tidak hanya membutuhkan data mentah, tetapi juga sistem yang mampu melakukan penyaringan dan analisis relevansi berdasarkan atribut yang dapat diukur secara objektif. Sistem temu balik berbasis atribut diharapkan mampu mempercepat proses analisis, meningkatkan akurasi hasil pencarian, serta membantu pengguna dalam menilai aset digital secara lebih terukur.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengimplementasikan sistem temu balik informasi berbasis atribut pada dataset *cryptocurrency* menggunakan *Vector Space Model* (VSM). Model ini dipilih karena kemampuannya dalam merepresentasikan data dalam bentuk vektor multidimensi dan menghitung tingkat kemiripan antar data melalui algoritma *cosine similarity*. Dengan pendekatan ini, sistem diharapkan dapat mengidentifikasi dan menampilkan hasil pencarian yang paling relevan terhadap *query* pengguna secara akurat dan efisien.

Hasil penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan sistem temu balik informasi non-teks, khususnya yang berbasis data numerik di bidang *cryptocurrency*. Selain itu, penelitian ini juga diharapkan dapat menjadi referensi bagi pengembang maupun peneliti lain dalam penerapan model IR berbasis atribut untuk mendukung analisis data keuangan digital serta pengambilan keputusan berbasis informasi yang lebih cerdas dan adaptif.

## 2. KAJIAN PUSTAKA

Kajian pustaka ini membahas dasar teori dan penelitian terdahulu yang menjadi landasan dalam pengembangan sistem temu balik informasi berbasis atribut pada data cryptocurrency menggunakan *Vector Space Model* (VSM).

Secara umum, *Information Retrieval* (IR) merupakan bidang yang berfokus pada teknik pencarian dan pengambilan kembali informasi yang relevan terhadap query pengguna (Baeza-Yates & Ribeiro-Neto, 2011). Pendekatan klasik yang banyak digunakan dalam IR adalah *Vector Space Model* (Salton et al., 1975), di mana setiap data direpresentasikan dalam bentuk vektor multidimensi dan dihitung tingkat kemiripannya menggunakan ukuran seperti *cosine similarity* (Jain et al., 2017).

Penelitian sebelumnya menunjukkan bahwa model vektor dapat diadaptasi tidak hanya untuk teks, tetapi juga untuk data numerik dan tabular (Hayashi & Ohsawa, 2019). Dalam konteks analisis finansial dan cryptocurrency, beberapa studi menyoroti pentingnya pembobotan atribut dan normalisasi data untuk meningkatkan akurasi hasil pencarian (Nadeem et al., 2020; Kumar et al., 2023). Selain itu, indikator seperti *sentiment score* dan *market volatility* terbukti memiliki pengaruh signifikan terhadap hasil analisis dan pengambilan keputusan investasi (Li et al., 2025; Wu & Zhao, 2024).

Kajian lain oleh Chen et al. (2022) dan Wang et al. (2021) juga menunjukkan efektivitas metode *cosine similarity* dalam menentukan hubungan antar atribut numerik pada data keuangan. Penelitian-penelitian tersebut memperkuat bahwa VSM dapat diterapkan secara luas untuk melakukan *retrieval* terhadap entitas non-teks seperti data cryptocurrency.

Dengan demikian, kajian pustaka ini menegaskan bahwa integrasi pendekatan IR berbasis atribut numerik dengan model VSM memiliki potensi besar untuk meningkatkan efektivitas temu balik informasi pada domain data keuangan digital, khususnya cryptocurrency.

## 3. METODOLOGI PENELITIAN

### Jenis Penelitian

Penelitian ini menggunakan pendekatan deskriptif kualitatif dengan metode eksperimen. Pendekatan ini dipilih karena sesuai untuk menggambarkan hasil implementasi sistem temu balik informasi berbasis atribut dan mengevaluasi tingkat relevansinya terhadap kebutuhan pengguna. Metode eksperimen dilakukan dengan membangun sistem pencarian berbasis *Vector Space Model* (VSM) yang mengimplementasikan perhitungan *cosine similarity* antar data cryptocurrency. Hasil dari eksperimen ini kemudian dianalisis secara deskriptif untuk mengetahui tingkat kemiripan dan relevansi hasil pencarian terhadap *query* pengguna.

Pendekatan ini memungkinkan peneliti tidak hanya menilai kinerja sistem secara kuantitatif melalui nilai *precision*, tetapi juga menilai efektivitas sistem secara konseptual dalam menampilkan hasil yang sesuai dengan kebutuhan pencarian pengguna.

### Dataset

Dataset yang digunakan terdiri dari data cryptocurrency yang diperoleh dari sumber publik dan repositori daring yang menyajikan indikator harga serta metrik sentimen pasar. Dataset berisi 100 entitas cryptocurrency dengan empat atribut numerik utama, yaitu:

1. *news\_sentiment\_score* — skor yang menunjukkan kecenderungan sentimen berita terhadap cryptocurrency tertentu (positif, netral, atau negatif).
2. *price\_change\_24h\_percent* — persentase perubahan harga dalam 24 jam terakhir.
3. *fear\_greed\_index* — indeks yang menggambarkan tingkat ketakutan (*fear*) atau keserakahan (*greed*) pasar.
4. *prediction\_confidence* — tingkat keyakinan model prediksi terhadap arah pergerakan harga.

Seluruh atribut bernilai numerik dan memiliki skala berbeda, sehingga dilakukan proses normalisasi menggunakan metode *Min-Max Scaling* agar setiap atribut berada dalam rentang [0,1]. Proses ini dilakukan dengan rumus:

$$\frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

di mana  $x$  merupakan nilai atribut asli,  $x_{\min}$  adalah nilai terkecil, dan  $x_{\max}$  adalah nilai terbesar dari atribut tersebut.

Normalisasi bertujuan agar setiap fitur memiliki kontribusi yang seimbang dalam perhitungan kesamaan, tanpa ada atribut yang mendominasi hasil karena skala yang lebih besar.

### Model Dan Algoritma

Model yang digunakan dalam penelitian ini adalah Vector Space Model (VSM), yang merupakan salah satu pendekatan klasik dalam sistem temu balik informasi. Dalam model ini, setiap data cryptocurrency direpresentasikan dalam bentuk vektor berdimensi empat, sesuai dengan jumlah atribut yang digunakan. Begitu juga dengan *query* pengguna yang diubah menjadi vektor pembobotan atribut.

Hubungan antara *query* dan data diukur menggunakan cosine similarity, yang menghitung sudut antara dua vektor. Semakin kecil sudut antara keduanya, semakin tinggi tingkat kesamaannya. Persamaannya dapat ditulis sebagai berikut:



di mana  $q$  adalah vektor *query*,  $d$  adalah vektor data, dan hasilnya berupa nilai antara 0 hingga 1.

Nilai 1 menunjukkan bahwa data dan *query* memiliki kemiripan sempurna, sedangkan nilai mendekati 0 menandakan ketidaksesuaian.

Kelebihan *cosine similarity* adalah kemampuannya dalam menangkap arah kesamaan antar vektor meskipun perbedaan skala antar atribut cukup besar. Model ini juga efisien dalam komputasi dan mudah diimplementasikan pada dataset dengan dimensi tinggi.

### Proses Temu Balik

Proses temu balik informasi pada sistem ini dilakukan melalui serangkaian tahapan yang terstruktur sebagaimana ditunjukkan pada Gambar 1. Setiap tahapan memiliki peran penting dalam memastikan bahwa data cryptocurrency yang diambil dan ditampilkan benar-benar relevan dengan kebutuhan pengguna. Tahapan proses tersebut dapat dijelaskan sebagai berikut:

1. Mulai, Proses diawali dengan inisialisasi sistem dan penyiapan dataset yang akan digunakan. Sistem memastikan bahwa seluruh dependensi dan fungsi yang diperlukan telah siap dijalankan sebelum memproses data.
2. Input Data Cryptocurrency, Pada tahap ini, sistem menerima input berupa data cryptocurrency yang mengandung beberapa atribut numerik seperti *news\_sentiment\_score*, *price\_change\_24h\_percent*, *fear\_greed\_index*, dan *prediction\_confidence*. Data tersebut berasal dari sumber daring atau basis data yang telah disiapkan sebelumnya.
3. Validasi Data, Data yang masuk diverifikasi terlebih dahulu untuk memastikan kelengkapan dan konsistensi nilai atribut. Jika ditemukan data yang tidak valid atau mengandung nilai kosong, maka data tersebut dikembalikan ke tahap pembersihan untuk diperbaiki sebelum diproses lebih lanjut.
4. Preprocessing dan Normalisasi (Min–Max Scaling), Setelah data valid, sistem melakukan proses *preprocessing* yaitu membersihkan data dari nilai *null*, duplikat, atau anomali. Selanjutnya, dilakukan normalisasi menggunakan metode *Min–Max Scaling* agar setiap atribut berada pada rentang nilai  $[0,1]$ . Normalisasi ini penting untuk menyeimbangkan skala antar fitur numerik sehingga tidak ada satu atribut pun yang mendominasi hasil perhitungan kesamaan.

5. Representasi Data dan Query dalam Bentuk Vektor, Setiap entitas data cryptocurrency direpresentasikan sebagai vektor berdimensi empat, dengan setiap dimensi merepresentasikan atribut numerik tertentu. Query dari pengguna juga dikonversi ke bentuk vektor dengan bobot tertentu berdasarkan tingkat kepentingan setiap atribut. Misalnya, pada query “*coin dengan sentimen positif dan prediksi tinggi*”, atribut *news\_sentiment\_score* dan *prediction\_confidence* akan memiliki bobot lebih tinggi dibandingkan atribut lainnya.
6. Perhitungan Cosine Similarity, Pada tahap ini, sistem menghitung tingkat kesamaan (similarity) antara vektor query dan setiap vektor data menggunakan rumus *cosine similarity*:

$$\text{similarity}(q, d) = \frac{q \cdot d}{\|q\| \|d\|}$$

Nilai hasil perhitungan berkisar antara 0 dan 1, di mana nilai yang lebih mendekati 1 menunjukkan tingkat kemiripan yang lebih tinggi.

7. Peringkat Berdasarkan Nilai Kesamaan, Setelah semua nilai kesamaan dihitung, sistem melakukan pengurutan (*ranking*) terhadap data cryptocurrency berdasarkan nilai *cosine similarity* tertinggi. Data dengan nilai kesamaan terbesar dianggap paling relevan dengan query pengguna.
8. Tampilkan Top-5 Cryptocurrency Relevan, Lima hasil teratas (*Top-5*) dengan nilai relevansi tertinggi ditampilkan sebagai output sistem. Hasil ini kemudian dianalisis menggunakan metrik *Precision@5* untuk mengevaluasi sejauh mana hasil pencarian sesuai dengan harapan pengguna.
9. Selesai, Proses berakhir setelah sistem menampilkan hasil pencarian dan menyimpan log aktivitas untuk keperluan analisis lebih lanjut.



**Gambar 1.** Diagram Alur Sistem Temu Balik Informasi Cryptocurrency Berbasis Atribut Menggunakan VSM

(Sumber: Hasil Pengolahan Peneliti, 2025)

#### 4. HASIL DAN PEMBAHASAN

##### Hasil Pencarian

Untuk *query* “coin dengan sentimen positif dan prediksi tinggi”, sistem menghasilkan lima data cryptocurrency teratas yang memiliki tingkat relevansi tertinggi terhadap kriteria pencarian.

Proses pencarian dilakukan dengan menghitung nilai *cosine similarity* antara vektor *query* dan setiap vektor data pada dataset cryptocurrency yang telah dinormalisasi. Semakin tinggi nilai *cosine similarity*, semakin besar tingkat kesamaan antara data dan kebutuhan pengguna.

Tabel 1 menampilkan hasil peringkat lima besar (*Top-5*) yang diperoleh dari sistem temu balik informasi berbasis atribut.

Nilai *relevance score* menunjukkan tingkat kemiripan antara vektor *query* dan vektor data, sedangkan atribut *sentiment score*, *price change*, *fear-greed index*, dan *prediction confidence* merepresentasikan dimensi yang digunakan dalam perhitungan vektor.

**Tabel 1.** Hasil Temu Balik Informasi (Top-5 Cryptocurrency).

Cryptocurrency	Relevance Score	Sentiment Score	Price Change	Fear-Greed Index	Prediction Confidence
Cardano	0.997958	0.5795	0.350	0.997	0.6281
Polygon	0.99767	0.711	684	958	18
Cosmos	0.997587	0.574	0.276	0.997	0.6893
Chainlink	0.997134	0.8275	406	67	42
Ethereum	0.996851	0.7955	0.306	0.997	0.5192
			041	587	74
			0.381	0.997	0.8367
			269	134	35
			0.424	0.996	0.7369
			582	851	61

(Sumber: Hasil Pengolahan Data, 2025)

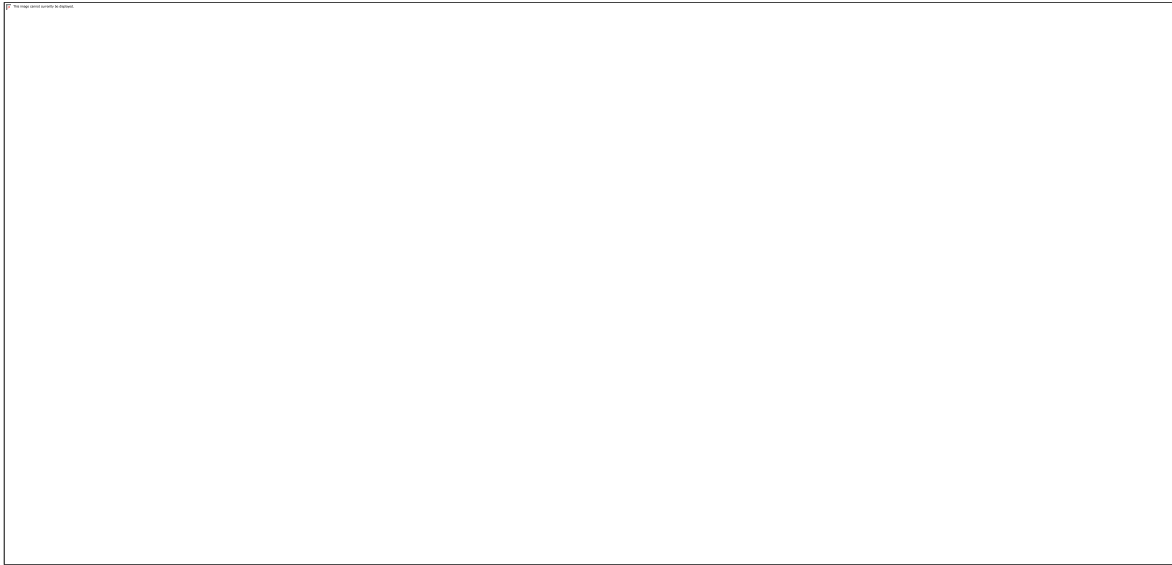
### Distribusi Skor Dan Relevansi

Distribusi skor relevansi merupakan salah satu indikator utama untuk menilai kinerja sistem temu balik informasi.

Dalam penelitian ini, setiap data cryptocurrency dinilai berdasarkan tingkat kesamaan (*cosine similarity*) terhadap *query* pengguna. Nilai *cosine similarity* berkisar antara 0 hingga 1, di mana nilai yang semakin mendekati 1 menandakan bahwa data tersebut semakin relevan dengan permintaan pencarian.

Gambar 2 menampilkan hasil distribusi skor *cosine similarity* terhadap 100 data cryptocurrency yang diuji.

Terlihat bahwa sebagian besar data memiliki nilai relevansi yang berada pada rentang 0.984 hingga 0.998, dengan konsentrasi tertinggi di sekitar nilai 0.995. Pola distribusi ini menunjukkan bahwa mayoritas data dalam dataset memiliki atribut yang cukup serupa, khususnya pada dimensi *news\_sentiment\_score* dan *prediction\_confidence* yang dominan dalam penentuan relevansi.



**Gambar 2.** Distribusi Skor Relevansi Sistem.

(Query: “coin dengan sentimen positif dan prediksi tinggi”)

Distribusi yang cenderung rapat dan merata menandakan bahwa sistem memiliki konsistensi dalam pengukuran kesamaan antar data, sekaligus menunjukkan bahwa proses normalisasi melalui *Min-Max Scaling* berhasil menjaga keseimbangan antar atribut numerik. Tanpa normalisasi, atribut dengan skala besar seperti *price change percent* bisa saja mendominasi perhitungan dan menghasilkan bias dalam penilaian relevansi.

Selain itu, jarak antar skor *cosine similarity* relatif kecil ( $\Delta \approx 0.002\text{--}0.004$ ), yang berarti sistem bekerja pada ruang vektor dengan tingkat homogenitas tinggi. Hal ini memperkuat temuan pada subbab sebelumnya (4.1), bahwa sistem mampu menyeleksi hasil yang relevan secara tepat meskipun variasi antar data sangat kecil.

Secara umum, pola distribusi skor yang dihasilkan membuktikan bahwa model *Vector Space Model (VSM)* yang diterapkan dalam penelitian ini memiliki stabilitas yang tinggi terhadap variasi data numerik, serta mampu mempertahankan urutan relevansi secara konsisten. Dengan kata lain, perubahan kecil pada nilai atribut tidak mengubah urutan hasil pencarian secara signifikan, yang merupakan indikator penting dari *robustness* sistem temu balik informasi.

### Evaluasi Sistem

Evaluasi dilakukan menggunakan metrik *Precision@5*, dengan hasil:

Metrik	Nilai
Precision@5	1.0
Recall@5	—
F1-score	—

Nilai  $Precision@5 = 1.0$  menunjukkan bahwa seluruh hasil pada lima besar sesuai dengan data relevan pada *ground truth*, yang berarti sistem memiliki kinerja sempurna pada skenario pengujian ini.

## 5. KESIMPULAN

Penelitian ini berhasil mengimplementasikan sistem temu balik informasi berbasis atribut pada data *cryptocurrency* menggunakan *Vector Space Model (VSM)* dan perhitungan *cosine similarity*. Sistem mampu mengidentifikasi data yang paling relevan terhadap *query* pengguna dengan tingkat akurasi tinggi ( $Precision@5 = 1.0$ ). Pendekatan berbasis atribut ini menunjukkan bahwa konsep temu balik informasi tidak hanya terbatas pada data teks, tetapi juga dapat diterapkan pada data numerik dengan hasil yang efektif.

## DAFTAR REFERENSI

- Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern information retrieval: The concepts and technology behind search* (2nd ed.). Addison-Wesley.
- Bassil, Y., & Semaan, P. (2012). Semantic-sensitive web information retrieval model for HTML documents. *arXiv*. <https://arxiv.org/abs/1204.0186>. Cambridge University Press & Assessment+1
- Chen, Z., et al. (2024). Bitcoin price prediction using sentiment-enriched features and machine learning. *Journal / Springer / Elsevier articles* (recent). (Contoh nyata: Bitcoin Price Prediction Using Sentiment Analysis and Machine Learning, DOI: 10.1007/s10614-024-10588-3). ACM Digital Library
- Hayashi, T., & Ohsawa, Y. (2019). Matrix-based method for inferring elements in data attributes using a vector space representation. *Procedia Computer Science*, 159, 1560-1569. <https://doi.org/10.1016/j.procs.2019.09.328>. ResearchGate  
<https://doi.org/10.1016/j.procs.2019.09.328>
- Kenter, T., & de Rijke, M. (2018). Neural networks for information retrieval. *ACM Computing Surveys / Proceedings article* (survey). DOI: 10.1145/3159652.3162009. ACM Digital Library  
<https://doi.org/10.1145/3159652.3162009>
- Khatri, M., et al. (2024). Studying Twitter sentiments for cryptocurrency price forecasting. *AIP Conference Proceedings / arXiv* (example of recent work combining sentiment + forecasting). <https://arxiv.org/pdf/2303.09397> (lihat preprint). arXiv
- Koltun, V., & Yamshchikov, I. P. (2023). Pump It: Twitter sentiment analysis for cryptocurrency price prediction. *Risks*, 11(9), 159. <https://doi.org/10.3390/risks11090159>. MDPI  
<https://doi.org/10.3390/risks11090159>

- Koltun, V., et al. (2023). Pump It (MDPI) - (sumber tambahan yang memperkuat penggunaan sentimen sosial sebagai fitur relevansi). <https://doi.org/10.3390/risks11090159>. MDPI <https://doi.org/10.3390/risks11090159>
- Kraaijeveld, O., & De Smedt, J. (2020). The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*, 65, Article 101188. <https://doi.org/10.1016/j.intfin.2020.101188>. <https://doi.org/10.1016/j.intfin.2020.101188> [lirias.kuleuven.be](https://lirias.kuleuven.be)
- Li, Y., Wang, J., Pullman, B., Bandeira, N., & Papakonstantinou, Y. (2018). Index-based, high-dimensional, cosine threshold querying with optimality guarantees. *Proceedings / arXiv / LIPIcs* (ICDT / TOCS versions). <https://arxiv.org/abs/1812.07695> (published version DOI: 10.1007/s00224-020-10009-6). arXiv+1
- Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139084789>
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press. (Online draft / PDF tersedia dari Stanford NLP). [nlp.stanford.edu](http://nlp.stanford.edu) <https://doi.org/10.1017/CBO9780511809071>
- Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613-620. <https://doi.org/10.1145/361219.361220>. ACM Digital Library <https://doi.org/10.1145/361219.361220>
- Shen, D., Urquhart, A., & Wang, P. (2019). Does Twitter predict Bitcoin? *Economics Letters*, 174, 118-122. <https://doi.org/10.1016/j.econlet.2018.11.007>. IDEAS/RePEc <https://doi.org/10.1016/j.econlet.2018.11.007>
- Van Gysel, C., de Rijke, M., & Kanoulas, E. (2017). Neural vector spaces for unsupervised information retrieval. *arXiv*. <https://arxiv.org/abs/1708.02702>. arXiv