

Penyaringan Laporan Keuangan Berbasis Prioritas Risiko Menggunakan Entity Graph Matching, String Similarity, dan Branch and Bound

Nashiruddin Akram - 13524090¹

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia

¹akrambaasir@gmail.com, 13524090@std.stei.itb.ac.id

Abstract—Laporan keuangan emiten berisi banyak label, angka, catatan, dan istilah bilingual. Pembaca saham perlu menemukan bagian yang relevan sebelum menilai kualitas laba, arus kas, piutang, struktur modal, dan risiko pendanaan. Makalah ini mengusulkan sistem screening laporan keuangan berbasis prioritas risiko pada tiga laporan publik kuartal I 2026, yaitu PT Bank Jago Tbk, PT Telekomunikasi Indonesia Tbk, dan PT Adaro Andalan Indonesia Tbk. Metode yang digunakan meliputi regular expression untuk mengambil kandidat teks, Knuth-Morris-Pratt dan Boyer-Moore untuk pencarian kata kunci, Levenshtein Distance dan Jaro-Winkler untuk kemiripan label, entity graph untuk audit trail relasi dokumen, serta Branch and Bound untuk assignment field dan prioritas review. Hasil uji menunjukkan KMP dan Boyer-Moore menghasilkan jumlah kemunculan yang sama pada seluruh kata kunci, Branch and Bound memetakan delapan field utama per emiten dengan 37 simpul ekspansi, dan Branch and Bound knapsack memilih empat item review dengan skor total 261,02. Dataset tiga emiten diposisikan sebagai proof of concept, bukan dasar generalisasi statistik untuk seluruh emiten. Sistem ini tidak memberi rekomendasi investasi, tetapi membantu pembaca menentukan urutan baca laporan keuangan secara lebih terarah.

Index Terms—KMP, Boyer-Moore, string similarity, entity graph, Branch and Bound, laporan keuangan, screening saham

I. PENDAHULUAN

Laporan keuangan adalah sumber utama dalam riset saham. Dokumen tersebut memuat laporan posisi keuangan, laba rugi, arus kas, perubahan ekuitas, dan catatan yang menjelaskan kebijakan akuntansi serta rincian akun. Dari dokumen tersebut, pembaca dapat menilai pertumbuhan pendapatan, kualitas laba, kekuatan kas, eksposur piutang, struktur modal, dan perubahan risiko dari periode sebelumnya. Masalahnya, satu laporan keuangan dapat berisi lebih dari seratus halaman, bahkan pada emiten besar jumlahnya dapat mendekati ratusan halaman. Jika pembaca memeriksa beberapa emiten sekaligus, jumlah halaman yang harus dibaca menjadi sangat besar. Karena itu, output yang paling berguna bukan hanya tabel rasio, melainkan urutan baca yang menunjukkan bagian mana perlu dibaca lebih dahulu dan alasan mengapa bagian tersebut penting.

Tantangan lain muncul karena istilah laporan keuangan berbeda antar sektor. Telkom memakai label pendapatan dan piutang usaha. Adaro memakai pendapatan usaha dan trade

receivables. Bank Jago tidak cocok diperlakukan sama dengan perusahaan non-bank karena eksposur utamanya muncul pada kredit yang diberikan dan pembiayaan syariah. Jika sistem hanya mencari satu kata kunci, angka penting dapat terlewat atau dipetakan ke makna yang salah.

Makalah ini memandang persoalan tersebut sebagai gabungan pencocokan string, pemetaan label, dan optimisasi prioritas baca. Dokumen diubah menjadi teks. Kata kunci dicari memakai algoritma string matching. Label yang ditemukan dipetakan ke field kanonik memakai kemiripan string. Hasil pemetaan disimpan sebagai entity graph agar jejak keputusan dapat ditelusuri. Kandidat risiko kemudian dipilih memakai Branch and Bound dengan batas kapasitas review.

Studi kasus menggunakan tiga laporan keuangan kuartal I 2026 yang diberikan sebagai data masukan. Bank Jago mewakili bank digital, Telkom mewakili telekomunikasi, dan Adaro mewakili pertambangan batu bara. Ketiganya dipilih untuk menunjukkan bahwa sistem dapat bekerja pada istilah laporan yang berbeda sektor, tetapi jumlah tersebut tetap diperlakukan sebagai proof of concept. Tujuan penelitian bukan menentukan saham layak beli atau jual. Tujuan utamanya adalah membuat alat bantu awal agar pembaca tahu bagian laporan mana yang perlu dibaca terlebih dahulu.

Kontribusi makalah ini terdiri dari empat hal. Pertama, makalah menerapkan KMP dan Boyer-Moore pada teks laporan keuangan asli. Kedua, makalah memformulasikan pemetaan label keuangan sebagai assignment problem. Ketiga, makalah memakai entity graph sebagai audit trail dari dokumen ke indikator. Keempat, makalah memakai Branch and Bound 0/1 knapsack untuk memilih prioritas review berdasarkan skor dan biaya baca.

II. DASAR TEORI

A. Knuth-Morris-Pratt

Knuth-Morris-Pratt atau KMP adalah algoritma pencocokan string yang mencari pola P sepanjang m di dalam teks T sepanjang n . KMP membangun fungsi pinggir $b(k)$ yang menyimpan panjang prefiks terpanjang dari $P[0..k]$ yang juga menjadi sufiks dari bagian pola tersebut. Fungsi ini membuat

algoritma dapat melanjutkan pencarian tanpa mengulang perbandingan yang sudah pasti cocok.

Fungsi pinggiran dapat ditulis sebagai

$$b(k) = \max\{q \mid 0 < q \leq k, P[0..q-1] = P[k-q+1..k]\}.$$

Jika tidak ada prefiks yang memenuhi, maka $b(k) = 0$. Ketika terjadi mismatch pada posisi j , nilai j diganti menjadi $b(j-1)$. Pergeseran efektif pola menjadi

$$s = j - b(j-1).$$

Biaya pra-proses KMP adalah $O(m)$ untuk menghitung fungsi pinggiran. Biaya pencarian adalah $O(n)$ karena indeks teks tidak pernah mundur. Kompleksitas totalnya adalah

$$T_{KMP}(n, m) = O(n + m).$$

Pada laporan keuangan yang panjang, sifat linear ini penting karena teks hasil ekstraksi PDF dapat berisi ratusan ribu karakter.

Dalam implementasi ini, KMP menghitung kemunculan kata kunci seperti pendapatan, piutang usaha, arus kas, laba periode berjalan, jumlah aset, dan jumlah liabilitas. Hasil KMP menjadi bukti awal bahwa kata kunci benar-benar muncul pada teks laporan.

B. Boyer-Moore

Boyer-Moore mencocokkan pola dari kanan ke kiri. Pada makalah ini digunakan aturan bad character. Untuk setiap karakter x , dibuat fungsi kemunculan terakhir $L(x)$ yang didefinisikan sebagai

$$L(x) = \max\{i \mid P[i] = x\}.$$

Jika karakter x tidak muncul pada pola, maka $L(x) = -1$.

Misalkan mismatch terjadi ketika $P[j]$ dibandingkan dengan $T[i]$ dan $T[i] = x$. Pergeseran pola dihitung dengan

$$s = \max(1, j - L(x)).$$

Jika x tidak ada pada pola, pergeseran dapat menjadi lebih besar. Jika x ada di sebelah kiri posisi mismatch, pola digeser sampai kemunculan terakhir x sejajar dengan $T[i]$.

Worst case varian sederhana dapat mencapai $O(nm)$. Namun, pada teks bahasa alami dan dokumen bisnis, alfabet yang digunakan relatif besar sehingga mismatch sering menghasilkan pergeseran lebih dari satu karakter. Dalam makalah ini, Boyer-Moore dipakai sebagai pembanding exact matching terhadap KMP. Jika jumlah kemunculan dari KMP dan Boyer-Moore sama, tahap pencarian kata kunci dianggap konsisten.

C. Regular Expression dan String Similarity

Regular expression digunakan untuk menemukan pola semi-terstruktur. Laporan keuangan sering menulis angka sebagai 37.189, 1,044,192, atau angka negatif dalam tanda kurung. Regex membantu mengambil kandidat label dan angka sebelum dilakukan pemetaan field.

Exact matching tidak cukup untuk menyamakan label. Total aset dapat ditulis sebagai jumlah aset atau total assets. Untuk itu digunakan Levenshtein Distance dan Jaro-Winkler. Jika

$d(a, b)$ menyatakan jarak Levenshtein antara string a dan b , maka skor normalisasinya adalah

$$S_L(a, b) = 1 - \frac{d(a, b)}{\max(|a|, |b|)}.$$

Skor akhir label dihitung dengan kombinasi

$$S(a, b) = 0.65S_{JW}(a, b) + 0.35S_L(a, b).$$

Biaya pemetaan menjadi

$$C(a, b) = (1 - S(a, b)) \times 100.$$

Nilai biaya kecil berarti label kandidat semakin cocok dengan field kanonik.

D. Entity Graph

Entity graph memodelkan laporan sebagai graf $G = (V, E)$. Simpul merepresentasikan dokumen, perusahaan, label mentah, field, indikator, dan item review. Sisi merepresentasikan relasi seperti contains, mapped to, used by, dan selected as. Model graf membuat hasil screening dapat ditelusuri kembali ke dokumen asal. Pada makalah ini, entity graph tidak diposisikan sebagai algoritma optimisasi utama, melainkan sebagai audit trail. Hal ini penting karena pembaca perlu mengetahui mengapa suatu item masuk prioritas review, label mana yang dipakai, dan indikator apa yang terbentuk dari label tersebut.

E. Branch and Bound

Branch and Bound digunakan untuk menyelesaikan persoalan optimisasi dengan membentuk pohon ruang status. Setiap simpul menyimpan solusi parsial dan batas nilai terbaik yang mungkin dicapai dari cabang tersebut. Jika batas suatu cabang tidak dapat mengalahkan solusi terbaik sementara, cabang dipangkas.

Pada assignment, sistem memilih pasangan antara field kanonik dan label kandidat. Untuk setiap pasangan field i dan label j , biaya C_{ij} diperoleh dari string similarity. Tujuan assignment adalah

$$\min \sum_{i=1}^n \sum_{j=1}^n C_{ij} x_{ij}$$

dengan kendala

$$\sum_{j=1}^n x_{ij} = 1, \quad \sum_{i=1}^n x_{ij} = 1, \quad x_{ij} \in \{0, 1\}.$$

Batas bawah suatu simpul dihitung dari biaya yang sudah dipilih ditambah biaya minimum dari setiap field yang belum diproses.

Pada pemilihan prioritas review, sistem memakai model 0/1 knapsack. Setiap item risiko memiliki skor p_i dan biaya baca w_i . Dengan kapasitas review K , tujuan optimisasi adalah

$$\max \sum_{i=1}^n p_i x_i \quad \text{dengan} \quad \sum_{i=1}^n w_i x_i \leq K.$$

Sistem memakai batas atas fraksional untuk memangkas cabang yang tidak mungkin mengalahkan solusi terbaik.

III. METODOLOGI

A. Alur Sistem

Alur kerja sistem ditunjukkan pada Gambar 1. Laporan PDF diubah menjadi teks. Teks dibersihkan dengan normalisasi huruf, spasi, dan tanda baca. Kata kunci dicari dengan KMP dan Boyer-Moore. Label kandidat dipetakan ke field kanonik. Hasil pemetaan disimpan dalam entity graph. Terakhir, sistem menghitung skor risiko dan memilih prioritas review.

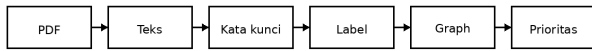


Fig. 1. Alur sistem screening laporan keuangan

B. Data Laporan Keuangan

Data yang digunakan berasal dari tiga laporan keuangan publik. Bank Jago melaporkan total aset Rp39.534.279 juta, total liabilitas Rp30.286.957 juta, total ekuitas Rp8.917.964 juta, dan laba bersih Rp85.776 juta pada 31 Maret 2026 [1]. Telkom melaporkan total aset Rp289.955 miliar, total liabilitas Rp134.145 miliar, total ekuitas Rp155.810 miliar, pendapatan Rp37.189 miliar, dan laba periode berjalan Rp6.054 miliar [2]. Adaro melaporkan total aset US\$5.780.540 ribu, total liabilitas US\$1.999.310 ribu, total ekuitas US\$3.781.230 ribu, pendapatan usaha US\$1.044.192 ribu, dan laba periode berjalan US\$153.768 ribu [3].

Fig. 2. Cuplikan sumber laporan keuangan yang digunakan

TABLE I. Dokumen studi kasus

Kode	Emiten	Sektor	Unit
ARTO	Bank Jago	Bank digital	juta Rp
TLKM	Telkom	Telekomunikasi	miliar Rp
AADI	Adaro	Pertambangan	ribu US\$

Tabel I memperlihatkan perbedaan sektor dan satuan. Perbandingan lintas sektor tidak dipakai untuk menyimpulkan emiten mana yang lebih baik. Setiap indikator dibaca sebagai sinyal untuk menentukan bagian laporan yang perlu diperiksa dalam emiten masing-masing.

C. Ground Truth dan Satuan Analisis

Ground truth pada makalah ini adalah daftar field utama yang dikurasi dari laporan keuangan. Daftar ini tidak dipakai

untuk menggantikan algoritma, tetapi dipakai sebagai acuan evaluasi. Sistem tetap harus mencari istilah di teks laporan, menghitung kemiripan label, dan memilih pasangan field melalui Branch and Bound. Dengan adanya ground truth, hasil pemetaan dapat diperiksa secara eksplisit.

Satuan analisis sengaja dipertahankan sesuai laporan masing-masing emiten. Bank Jago menggunakan juta Rupiah, Telkom menggunakan miliar Rupiah, dan Adaro menggunakan ribuan Dolar AS. Sistem tidak mengubah seluruh angka ke satu mata uang karena tujuan makalah bukan valuasi lintas emiten. Tujuan yang diuji adalah kemampuan mencari field, membentuk sinyal, dan memilih prioritas baca di dalam konteks tiap emiten.

Ada tiga prinsip yang dipakai saat membuat ground truth. Pertama, field harus muncul pada laporan utama atau catatan yang berhubungan langsung. Kedua, field harus dapat digunakan untuk membentuk indikator screening. Ketiga, field harus cukup umum agar dapat diterapkan pada sektor berbeda. Prinsip ketiga membuat beberapa field perlu didefinisikan sebagai proksi. Contohnya, receivable pada Bank Jago memakai kredit dan pembiayaan syariah, bukan piutang usaha.

TABLE II. Contoh pemetaan field lintas sektor

Kode	Field	Label yang dipakai
ARTO	revenue	pendapatan bunga dan syariah bersih
ARTO	receivable	kredit dan pembiayaan syariah
TLKM	revenue	pendapatan
TLKM	receivable	piutang usaha
AADI	revenue	pendapatan usaha
AADI	receivable	piutang usaha

Tabel II menunjukkan alasan mengapa sistem membutuhkan pemetaan label, bukan hanya pencarian kata. Label revenue pada tiga emiten memiliki bentuk yang berbeda. Label receivable juga berubah karena Bank Jago adalah bank, sedangkan Telkom dan Adaro adalah perusahaan non-bank. Tanpa pemetaan lintas sektor, sistem dapat membandingkan angka yang secara ekonomi tidak setara.

Pemetaan ini tetap dibatasi agar tidak terlalu bebas. Setiap field memiliki daftar alias yang ditentukan sebelum program dijalankan. Alias tersebut berfungsi sebagai kamus kecil yang menjadi acuan similarity. Dengan cara ini, sistem tidak mengarang field baru, tetapi memilih kandidat terbaik dari label yang sudah tersedia pada laporan.

D. Implementasi String Matching

Tahap pencarian kata kunci memakai KMP dan Boyer-Moore pada teks yang sudah dinormalisasi. KMP dipakai karena kompleksitasnya linear dan stabil. Boyer-Moore dipakai karena pergeseran bad character efisien pada teks bisnis yang memiliki alfabet besar. Keduanya sengaja dipakai bersamaan agar hasil exact matching dapat saling memeriksa.

```
text = bersih(baca_pdf(path))
pattern = bersih(term)
pos_kmp, cmp_kmp = kmp(text, pattern)
pos_bm, cmp_bm = boyer_moore(text, pattern)
```

Program menyimpan jumlah kemunculan dan jumlah perbandingan karakter. Jumlah kemunculan harus sama. Jumlah perbandingan karakter boleh berbeda karena mekanisme pergeseran KMP dan Boyer-Moore berbeda.

Pseudocode KMP yang digunakan terdiri dari dua tahap. Tahap pertama menghitung fungsi pinggiran. Tahap kedua membaca teks dari kiri ke kanan dan memakai nilai border ketika mismatch terjadi.

```

hitung border function untuk pola P
j ← 0
for setiap karakter T[i] do
  while j > 0 dan T[i] ≠ P[j] do
    j ← b(j - 1)
  end while
  if T[i] = P[j] then
    majukan indeks pola
  end if
  if seluruh pola cocok then
    simpan posisi kemunculan
  end if
end for

```

Pseudocode Boyer-Moore dimulai dari posisi paling kanan pola. Ketika mismatch terjadi, pola digeser memakai fungsi kemunculan terakhir.

```

bangun last occurrence L(x) untuk pola P
sejajarkan pola dengan teks
while pola masih berada dalam teks do
  cocokkan karakter dari kanan ke kiri
  if seluruh pola cocok then
    simpan posisi kemunculan
  else
    geser pola sejauh max(1, j - L(x))
  end if
end while
end while

```

E. Pemetaan Field dan Prioritas Review

Setelah label ditemukan, setiap field kanonik dipasangkan dengan label kandidat. Field yang digunakan adalah revenue, assets, liabilities, equity, income, cfo, receivable, dan cash. Pada Bank Jago, field receivable diproksikan dengan kredit yang diberikan dan pembiayaan syariah. Pada Telkom dan Adaro, field tersebut diproksikan dengan piutang usaha.

Branch and Bound assignment dipakai agar pemetaan tidak hanya memilih label terbaik secara lokal. Satu label hanya boleh dipakai oleh satu field. Setelah field dipetakan, sistem membentuk kandidat risiko. Skor risiko pada penelitian ini adalah skor prioritas baca, bukan ukuran risiko investasi absolut. Skor tinggi berarti bagian laporan lebih layak dibaca terlebih dahulu.

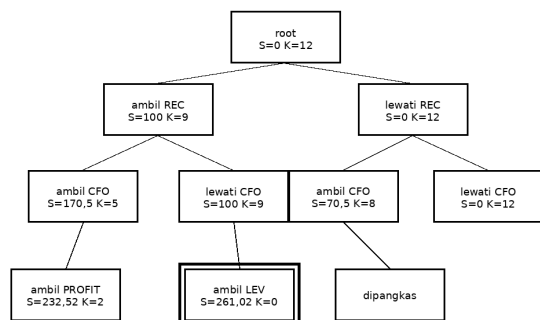


Fig. 3. Visualisasi Branch and Bound pada pemilihan prioritas review

Gambar 3 memperlihatkan cabang ambil dan lewati pada knapsack. Setiap simpul menyimpan skor sementara dan sisa kapasitas. Cabang dengan batas atas rendah dipangkas sehingga sistem tidak perlu memeriksa seluruh subset.

F. Bentuk Keluaran Sistem

Keluaran akhir sistem adalah berkas hasil/prioritas_baca.csv. Berkas ini berisi rank, emiten, kode item, skor, biaya baca, dan informasi jumlah simpul Branch and Bound. Format keluaran ditunjukkan pada Gambar 4. Pada laporan yang panjang, keluaran ini berfungsi sebagai peta baca awal. Pembaca tidak perlu membuka seluruh laporan dari halaman pertama, tetapi dapat langsung memulai dari item dengan rank tertinggi.

hasil/prioritas_baca.csv				
\$ python src/jalankan_screening.py				
selesai				
Rank	Emiten	Item	Skor	Biaya
1	ARTO	REC	100	3
2	AADI	PROFIT	62.02	3
3	AADI	CFO	70.5	4
4	ARTO	LEV	28.5	2
Total skor = 261.02				
Kapasitas baca = 12				
Simpul BnB = 18				

Fig. 4. Cuplikan keluaran akhir prioritas baca

Kode item pada keluaran dibaca sebagai kategori review. REC mengarah ke piutang, kredit, pembiayaan, dan cadangan kerugian penurunan nilai. CFO mengarah ke laporan arus kas. PROFIT mengarah ke laba, margin, dan beban. LEV mengarah ke struktur pendanaan. Dengan format ini, keluaran sistem dapat dipakai sebagai checklist ringkas sebelum pembaca melakukan analisis fundamental yang lebih lengkap.

Pseudocode Branch and Bound assignment yang dipakai adalah sebagai berikut. Fungsi lower bound menghitung biaya solusi parsial ditambah biaya minimum dari field yang belum dipasangkan.

```

best ← ∞
mulai dari field pertama
for setiap label yang belum dipakai do
  pilih pasangan sementara
  hitung lower bound cabang
  if lower bound lebih kecil dari best then
    lanjutkan ke field berikutnya
  else
    pangkas cabang
  end if
end for

```

Pada knapsack, urutan kandidat disusun berdasarkan densitas skor per biaya. Batas atas dihitung dengan relaksasi fraksional. Jika batas atas tidak lebih baik dari solusi terbaik, cabang tidak diperluas.

IV. HASIL DAN PEMBAHASAN

A. Hasil Pencarian Kata Kunci

Tabel III menunjukkan contoh hasil pencarian kata kunci. Kolom KMP dan BM selalu sama. Kolom perbandingan memperlihatkan bahwa Boyer-Moore membutuhkan lebih sedikit perbandingan karakter pada contoh ini karena pergeseran bad character menghindari banyak posisi teks.

TABLE III. Hasil pencarian kata kunci

Kode	Kata kunci	KMP	BM	Cmp KMP	Cmp BM
ARTO	pendapatan	95	95	438339	62774
ARTO	kredit	101	101	438738	49376
TLKM	pendapatan	134	134	309067	45770
TLKM	piutang	45	45	310053	33691
AADI	pendapatan	66	66	459118	65571
AADI	piutang	25	25	458155	50722

Tabel III menunjukkan bahwa exact matching berhasil diterapkan pada laporan nyata. Perbedaan jumlah kemunculan antar emiten tidak langsung menjadi risiko. Angka tersebut hanya menunjukkan seberapa sering istilah tertentu dibahas dalam dokumen.

B. Validasi Pemetaan Field

Tabel IV memperlihatkan validasi pemetaan field terhadap ground truth manual dari laporan. Validasi dilakukan pada delapan field utama per emiten. Seluruh field pada studi kasus berhasil dipetakan ke label yang sesuai.

TABLE IV. Akurasi pemetaan field

Kode	Benar	Total field	Akurasi
AADI	8	8	100%
ARTO	8	8	100%
TLKM	8	8	100%

Tabel IV tidak berarti sistem sudah pasti akurat pada semua emiten. Dataset uji masih kecil. Namun, tabel ini menunjukkan bahwa pada tiga dokumen yang digunakan, pemetaan label oleh Branch and Bound sesuai dengan ground truth manual.

C. Branch and Bound dan Baseline

Tabel V membandingkan greedy assignment dengan Branch and Bound. Pada dataset ini, greedy menghasilkan biaya yang sama dengan Branch and Bound. Walaupun demikian, Branch and Bound tetap penting karena memberikan jaminan optimal global. Greedy dapat gagal pada kasus ketika pilihan lokal terbaik mengambil label yang lebih dibutuhkan field lain.

TABLE V. Perbandingan assignment dan jumlah simpul

Kode	Greedy	BnB	Simpul BnB
AADI	94,729	94,729	37
ARTO	123,878	123,878	37
TLKM	95,800	95,800	37

Assignment delapan field memiliki $8! = 40320$ kemungkinan jika diperiksa secara exhaustive. Pada tahap knapsack, seluruh subset berjumlah $2^8 = 256$ untuk delapan kandidat

positif. Tabel VI menunjukkan perbandingan jumlah ruang solusi dengan simpul yang benar-benar diekspansi.

TABLE VI. Ukuran ruang solusi dan simpul BnB

Tugas	Exhaustive	Simpul BnB
Assignment	40320	37
Knapsack	256	18

Tabel VI menunjukkan bahwa Branch and Bound tetap relevan walaupun studi kasusnya kecil. Pada dataset yang lebih besar, selisih antara exhaustive dan simpul yang diekspansi akan semakin penting.

D. Ringkasan Rasio Screening

Tabel VII menunjukkan rasio yang dibentuk dari hasil pemetaan field. Gap persamaan akuntansi bernilai nol pada semua emiten. Cash ratio Adaro sebesar 0,295 menjadi sinyal bahwa arus kas operasi kuartal berjalan lebih rendah daripada laba. Receivable exposure Adaro sebesar 46,01 persen juga lebih tinggi daripada Telkom.

TABLE VII. Rasio screening utama

Kode	Gap	Cash	RecRev	Lev
ARTO	0,000	27,734	3464,13	3,40
TLKM	0,000	2,856	31,42	0,86
AADI	0,000	0,295	46,01	0,53

Tabel VII dibaca dalam konteks masing-masing sektor. RecRev Bank Jago adalah eksposur kredit dan pembiayaan terhadap pendapatan bunga dan syariah bersih. RecRev Telkom dan Adaro adalah piutang usaha terhadap pendapatan. Oleh karena itu, angka Bank Jago tidak boleh dibandingkan langsung dengan angka perusahaan non-bank.

Skor kandidat risiko dibentuk dari aturan heuristik sederhana. Cash flow risk dihitung ketika rasio arus kas operasi terhadap laba berada di bawah satu. Receivable risk dihitung dari kelebihan rasio receivable terhadap pendapatan di atas ambang 25 persen. Profit risk dan revenue risk dihitung ketika terjadi penurunan terhadap periode pembandingan. Leverage risk dihitung ketika rasio liabilitas terhadap ekuitas melewati ambang 1,5. Seluruh skor dibatasi pada rentang 0 sampai 100.

Rumus yang dipakai untuk cash flow risk adalah

$$score_{CFO} = \max(0, (1 - \text{cash ratio}) \times 100).$$

Rumus receivable risk adalah

$$score_{REC} = \max(0, \min(100, (\text{RecRev} - 25) \times 2)).$$

Kedua rumus tersebut sengaja dibuat sederhana agar fokus makalah tetap berada pada pencocokan string, pemetaan label, dan optimisasi prioritas.

E. Prioritas Review

Tabel VIII menunjukkan hasil akhir Branch and Bound knapsack dengan kapasitas review $K = 12$. Sistem memilih empat item dengan skor total 261,02. Kapasitas habis tepat karena biaya item terpilih adalah 3, 3, 4, dan 2.

TABLE VIII. Prioritas review hasil knapsack

Rank	Kode	Item	Skor	Biaya
1	ARTO	REC	100,00	3
2	AADI	PROFIT	62,02	3
3	AADI	CFO	70,50	4
4	ARTO	LEV	28,50	2

Tabel VIII adalah output yang langsung dapat dipakai sebagai checklist baca. ARTO REC mengarahkan pembaca ke catatan kredit, pembiayaan, dan cadangan kerugian penurunan nilai. AADI PROFIT mengarahkan pembaca ke penurunan laba. AADI CFO mengarahkan pembaca ke laporan arus kas. ARTO LEV mengarahkan pembaca ke struktur pendanaan bank.

F. Kandidat Risiko Positif

Tabel IX memperlihatkan kandidat risiko yang memiliki skor positif. Tidak semua kandidat dipilih oleh knapsack karena kapasitas review dibatasi. Jika pembaca memiliki waktu lebih banyak, kandidat yang tidak terpilih dapat dibaca sebagai daftar lanjutan.

TABLE IX. Kandidat risiko positif

Kode	Item	Skor	Biaya
ARTO	REC	100,00	3
ARTO	LEV	28,50	2
TLKM	REC	12,84	3
TLKM	PROFIT	34,96	3
AADI	CFO	70,50	4
AADI	REC	42,02	3
AADI	PROFIT	62,02	3
AADI	REV	20,66	2

Tabel IX menunjukkan bahwa Telkom tetap memiliki sinyal penurunan laba, walaupun tidak masuk prioritas utama pada kapasitas $K = 12$. Adaro juga memiliki sinyal piutang dan penurunan pendapatan. Hasil ini membuat sistem fleksibel karena pembaca dapat menaikkan kapasitas review jika ingin daftar yang lebih panjang.

G. Cara Membaca Output Sistem

Output sistem dibaca sebagai urutan kerja, bukan sebagai vonis. Jika item ARTO REC berada pada rank pertama, pembaca tidak langsung menyimpulkan bahwa Bank Jago buruk. Pembaca diarahkan untuk membuka bagian kredit, pembiayaan, dan cadangan kerugian penurunan nilai. Setelah itu, pembaca dapat menilai apakah pertumbuhan pembiayaan masih diikuti kualitas aset yang baik.

Jika item AADI CFO berada pada prioritas tinggi, pembaca diarahkan ke laporan arus kas. Fokus baca bukan hanya nilai arus kas operasi, tetapi juga komponen pembentuknya seperti penerimaan pelanggan, pembayaran pemasok, royalti, pajak, biaya karyawan, dan biaya keuangan. Jika arus kas operasi turun tajam ketika laba juga turun, pembaca memperoleh alasan untuk membaca catatan pendapatan dan modal kerja dengan lebih teliti.

Jika item TLKM PROFIT tidak masuk daftar utama, bukan berarti sinyal tersebut diabaikan. Item itu tetap ada pada kandidat risiko. Dalam kondisi kapasitas review yang lebih

longgar, item tersebut dapat dipilih setelah item dengan skor lebih tinggi selesai dibaca. Dengan demikian, sistem tidak membuang informasi, tetapi mengatur urutan baca berdasarkan keterbatasan waktu.

H. Audit Trail Entity Graph

Tabel X menunjukkan contoh relasi graf dari dokumen menuju prioritas review. Relasi ini menjaga output tetap dapat dijelaskan. Jika pembaca melihat item AADI CFO, pembaca dapat menelusuri bahwa item tersebut berasal dari label arus kas operasi pada laporan Adaro dan dibandingkan dengan laba periode berjalan.

TABLE X. Contoh relasi entity graph

Sumber	Relasi	Tujuan
PDF AADI	contains	Pendapatan usaha
Pendapatan usaha	mapped to	revenue
revenue	used by	revenue growth
CFO AADI	compared to	income AADI
AADI CFO	selected as	review priority

Tabel X memperlihatkan bahwa skor risiko bukan angka yang berdiri sendiri. Setiap skor dapat dikaitkan dengan dokumen, label, field, dan indikator yang membentuknya.

V. STUDI KASUS PER EMITEN

A. PT Bank Jago Tbk

Bank Jago memiliki struktur laporan khas bank. Total aset, total liabilitas, ekuitas, dan laba dapat dipetakan langsung. Revenue diprosikan dengan pendapatan bunga dan syariah bersih. Receivable diprosikan dengan kredit yang diberikan dan pembiayaan syariah. Keputusan ini penting karena menggunakan piutang usaha biasa pada bank akan menghasilkan interpretasi yang salah.

Hasil sistem menempatkan ARTO REC sebagai prioritas tertinggi. Skor 100 bukan berarti laporan mengandung kesalahan. Skor tersebut berarti eksposur kredit dan pembiayaan sangat besar terhadap pendapatan yang digunakan sebagai pembanding. Dalam analisis saham, sinyal ini mengarahkan pembaca untuk membuka catatan kredit, kualitas aset, cadangan kerugian penurunan nilai, dan konsentrasi pembiayaan.

Item ARTO LEV juga terpilih. Leverage tinggi adalah kondisi umum pada bank karena bisnis bank memang menghimpun dana dan menyalurkan pembiayaan. Namun, indikator ini tetap berguna sebagai pengingat bahwa pembaca perlu melihat struktur dana pihak ketiga, pinjaman yang diterima, efek repo, serta dana syirkah temporer.

B. PT Telekomunikasi Indonesia Tbk

Telkom memiliki label yang lebih mudah dipetakan. Istilah pendapatan, piutang usaha, jumlah aset, jumlah liabilitas, dan jumlah ekuitas muncul eksplisit pada laporan. Biaya assignment Telkom rendah. Cash ratio Telkom juga berada di atas satu, yang berarti arus kas operasi lebih besar daripada laba periode berjalan.

Sinyal yang tetap muncul pada Telkom adalah penurunan laba. Pendapatan naik 1,50 persen, sedangkan laba turun

17,48 persen. Sistem menyimpan item TLKM PROFIT sebagai kandidat, tetapi tidak memilihnya pada kapasitas review yang sempit. Bagi pembaca saham, Telkom tetap layak dibaca pada bagian beban operasi, penyusutan, biaya pendanaan, dan laba yang dapat diatribusikan kepada pemilik entitas induk.

C. PT Adaro Andalan Indonesia Tbk

Adaro menghasilkan dua item prioritas utama. Laba periode berjalan turun 31,01 persen dibanding periode pembandingan. Arus kas operasi hanya 0,295 kali laba periode berjalan. Dua sinyal ini saling memperkuat karena penurunan laba disertai tekanan pada konversi kas.

Perhitungan cash ratio Adaro adalah

$$\frac{45.289}{153.768} = 0,295.$$

Nilai tersebut membuat skor CFO menjadi

$$(1 - 0,295) \times 100 = 70,50.$$

Hasil ini mengarahkan pembaca ke laporan arus kas Adaro. Bagian yang perlu diperiksa adalah penerimaan dari pelanggan, pembayaran kepada pemasok, royalti, pajak, biaya karyawan, dan biaya keuangan.

VI. EVALUASI ALGORITMA

Penerapan algoritma pada makalah ini memiliki peran yang terpisah. KMP dan Boyer-Moore dipakai untuk exact string matching pada teks laporan. Levenshtein dan Jaro-Winkler mengubah variasi label menjadi matriks biaya. Branch and Bound assignment memilih pasangan field dan label dengan biaya minimum. Branch and Bound knapsack memilih item risiko dengan skor maksimum dalam kapasitas review terbatas.

Pemisahan ini membuat metodologi tidak hanya berupa kumpulan rumus. KMP dan Boyer-Moore menjawab apakah kata kunci muncul. Similarity menjawab seberapa dekat label mentah dengan field kanonik. Assignment menjawab pasangan label yang paling konsisten secara global. Knapsack menjawab bagian yang harus dibaca lebih dahulu ketika waktu terbatas.

KMP memiliki kompleksitas $O(n+m)$. Boyer-Moore varian bad character memiliki worst case $O(nm)$, tetapi efisien secara praktis pada teks laporan. Pada eksperimen ini, keduanya digunakan sebagai validasi konsistensi exact matching karena hasil jumlah kemunculan harus sama untuk kata kunci yang sama. Assignment dengan delapan field membutuhkan 40320 kemungkinan jika exhaustive. Branch and Bound hanya mengekskansi 37 simpul per emiten. Knapsack delapan kandidat positif memiliki 256 subset, sedangkan Branch and Bound mengekskansi 18 simpul. Angka ini memperlihatkan manfaat pemangkasan, walaupun ukuran studi kasus masih kecil.

Sistem masih memiliki keterbatasan. Nilai line item utama dimasukkan ke CSV hasil kurasi dari laporan, sehingga ekstraksi tabel PDF otomatis belum menjadi fokus. Rasio Bank Jago tidak boleh dibaca sama dengan rasio piutang perusahaan non-bank. Skor risiko bersifat heuristik dan perlu dikalibrasi jika ingin digunakan pada banyak emiten. Sistem juga tidak mengeluarkan rekomendasi investasi. Sistem hanya membantu pembaca menentukan urutan baca dokumen.

Dalam workflow riset saham, sistem ini paling tepat ditempatkan sebelum analisis valuasi. Pembaca memasukkan laporan terbaru, menjalankan program, lalu membaca bagian laporan berdasarkan urutan prioritas. Setelah itu, pembaca tetap perlu melakukan analisis fundamental normal seperti memahami model bisnis, kualitas manajemen, prospek industri, valuasi, dan risiko makro. Dengan posisi tersebut, sistem berperan sebagai filter awal, bukan pengganti keputusan investasi.

Kelebihan utama pendekatan ini adalah transparansi. Setiap output dapat dijelaskan melalui algoritma yang dipakai. Jika suatu item masuk prioritas, pembaca dapat menelusuri kata kunci yang ditemukan, label yang dipetakan, rasio yang terbentuk, dan proses pemilihan knapsack. Transparansi ini penting karena screening laporan keuangan tidak boleh menjadi kotak hitam.

VII. IMPLIKASI PRAKTIS

Sistem ini ditempatkan sebagai alat bantu baca sebelum analisis fundamental mendalam. Output yang dihasilkan bukan rekomendasi investasi, melainkan urutan bagian laporan yang perlu dibuka terlebih dahulu. Manfaat praktisnya dapat diringkas sebagai berikut.

- **Screening awal.** Pembaca tidak perlu memulai dari halaman pertama. Sistem langsung mengarahkan perhatian ke akun dan catatan yang memiliki sinyal paling kuat.
- **Checklist riset.** Jika item CFO muncul, pembaca memeriksa laporan arus kas. Jika item REC muncul, pembaca memeriksa piutang, kredit, pembiayaan, dan cadangan kerugian penurunan nilai. Jika item PROFIT muncul, pembaca memeriksa margin, beban, dan pos non-operasional.
- **Audit trail.** Setiap prioritas dapat ditelusuri dari kata kunci, label laporan, field kanonik, rasio, sampai keputusan Branch and Bound. Hal ini membuat output lebih transparan dibanding daftar risiko yang tidak dapat dijelaskan.
- **Lintas sektor.** Sistem tidak memaksa seluruh emiten memakai label yang sama. Bank tetap dibaca dengan konteks kredit dan pembiayaan, sedangkan perusahaan non-bank dibaca dengan konteks piutang usaha dan pendapatan.
- **Efisiensi waktu.** Ketika banyak laporan dirilis bersamaan, sistem dapat menjadi filter awal untuk menentukan laporan dan catatan mana yang perlu dibaca lebih dahulu.

Batasan pentingnya adalah skor dan kapasitas review masih berupa rancangan awal. Untuk penggunaan nyata, bobot perlu diuji terhadap data historis, misalnya apakah item yang sering muncul berkorelasi dengan revisi estimasi analisis, penurunan kualitas laba, atau reaksi pasar.

Dalam konteks Strategi Algoritma, studi kasus ini menunjukkan bahwa algoritma klasik masih relevan untuk dokumen bisnis. KMP dan Boyer-Moore dipakai untuk menelusuri istilah penting dalam laporan panjang. Branch and Bound dipakai untuk memilih pasangan field-label serta daftar review yang paling bernilai terhadap kapasitas baca.

VIII. KESIMPULAN

Makalah ini membangun sistem screening laporan keuangan publik menggunakan algoritma Strategi Algoritma. KMP dan Boyer-Moore menghasilkan jumlah kemunculan kata kunci yang konsisten. String similarity mampu mengatasi variasi label antar sektor. Branch and Bound assignment memetakan delapan field utama per emiten dengan simpul ekspansi terbatas. Branch and Bound knapsack menghasilkan prioritas review yang konkret bagi pembaca saham.

Hasil studi kasus menunjukkan prioritas utama berada pada eksposur pembiayaan Bank Jago, penurunan laba Adaro, kualitas arus kas Adaro, dan leverage Bank Jago. Telkom tetap memiliki sinyal penurunan laba, tetapi tidak masuk prioritas utama pada kapasitas review yang sempit. Dengan demikian, sistem ini berguna sebagai alat bantu awal dalam membaca laporan keuangan, terutama ketika pembaca perlu menyaring banyak emiten dalam waktu terbatas.

Interpretasi hasil perlu ditempatkan secara proporsional. Dataset tiga emiten adalah proof of concept, CSV masih dikurasi manual, dan skor risiko masih bersifat heuristik. Namun, keterbatasan tersebut tidak mengubah kontribusi utama makalah, yaitu menunjukkan alur yang dapat dijelaskan dari pencarian kata kunci sampai prioritas baca.

Pengembangan berikutnya dapat diarahkan pada ekstraksi tabel PDF yang lebih otomatis, perluasan dataset ke lebih banyak emiten, kalibrasi skor risiko berbasis historis, dan integrasi dengan data harga saham untuk menguji hubungan sinyal dokumen dengan reaksi pasar.

IX. UCAPAN TERIMA KASIH

Penulis mengucapkan syukur kepada Allah SWT atas segala rahmat dan karunia-Nya sehingga makalah ini dapat diselesaikan dengan baik. Ucapan terima kasih juga disampaikan kepada Bapak Rila Mandala dan tim pengajar IF2211 Strategi Algoritma yang telah memberikan fondasi teoritis yang kuat dalam bidang strategi algoritma. Apresiasi juga disampaikan kepada rekan-rekan mahasiswa Teknik Informatika ITB atas diskusi dan masukan konstruktif yang sangat membantu selama penulisan makalah ini.

REFERENCES

- [1] PT Bank Jago Tbk, *Laporan Keuangan 31 Maret 2026 dan 31 Desember 2025, serta untuk Periode Tiga Bulan yang Berakhir 31 Maret 2026 dan 2025*, 2026.
- [2] PT Telekomunikasi Indonesia Tbk, *Laporan Keuangan Konsolidasian Tanggal 31 Maret 2026 dan untuk Periode Tiga Bulan yang Berakhir pada Tanggal Tersebut*, 2026.
- [3] PT Adaro Andalan Indonesia Tbk, *Laporan Keuangan Konsolidasian Interim 31 Maret 2026 dan 31 Desember 2025*, 2026.
- [4] R. Munir, N. U. Maulidevi, and M. L. Khodra, *Bahan Kuliah IF2211 Strategi Algoritma, Pencocokan String, Regular Expression, dan Branch and Bound*. Institut Teknologi Bandung, 2026.
- [5] D. E. Knuth, J. H. Morris, and V. R. Pratt, "Fast pattern matching in strings," *SIAM Journal on Computing*, vol. 6, no. 2, pp. 323–350, 1977.
- [6] R. S. Boyer and J. S. Moore, "A fast string searching algorithm," *Communications of the ACM*, vol. 20, no. 10, pp. 762–772, 1977.
- [7] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Soviet Physics Doklady*, vol. 10, no. 8, pp. 707–710, 1966.

- [8] W. E. Winkler, "String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage," U.S. Bureau of the Census, 1990.
- [9] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 3rd ed. MIT Press, 2009.

LAMPIRAN PROGRAM DAN VIDEO

- 1) <https://github.com/Akram17t/MakalahSTIMA/>

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiat.

Jakarta, 19 Juni 2026



Nashiruddin Akram, 13524090