

PERBANDINGAN METODE *K-NEAREST NEIGHBOR* DAN *ADAPTIVE BOOSTING* PADA KASUS KLASIFIKASI MULTI KELAS

Ade Irma Prianti¹, Rukun Santoso², Arief Rachman Hakim³

^{1,2,3}Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

Email : irmapriantiade56@gmail.com

ABSTRACT

The company's financial health provides an indication of company's performance that is useful for knowing the company's position in industrial area. The company's performance needs to be predicted to knowing the company's progress. *K-Nearest Neighbor* (KNN) and *Adaptive Boosting* (AdaBoost) are classification methods that can be used to predict company's performance. KNN classifies data based on the proximity of the data distance while AdaBoost works with the concept of giving more weight to observations that include weak learners. The purpose of this study is to compare the KNN and AdaBoost methods to find out better methods for predicting company's performance in Indonesia. The dependent variable used in this study is the company's performance which is classified into four classes, namely unhealthy, less healthy, healthy, and very healthy. The independent variables used consist of seven financial ratios namely ROA, ROE, WCTA, TATO, DER, LDAR, and ROI. The data used are financial ratio data from 575 companies listed on the Indonesia Stock Exchange in 2019. The results of this study indicate that the prediction of company's performance in Indonesia should use the AdaBoost method because it has a classification accuracy of 0,84522 which is greater than the KNN method's accuracy of 0,82087.

Keywords: company's performance, classification, KNN and AdaBoost, classification accuracy.

1. PENDAHULUAN

Tingkat kesehatan keuangan perusahaan sangat penting diketahui oleh suatu perusahaan untuk meningkatkan efisiensi usaha supaya kemampuan memperoleh laba dapat ditingkatkan. Tingkat kesehatan keuangan perusahaan menjadi masalah penting bagi berbagai pihak baik pihak dalam (pemilik dan manajemen) maupun pihak luar seperti investor dan perbankan.

Kinerja perusahaan di Indonesia dapat dikategorikan menjadi empat tingkatan yaitu tidak sehat, kurang sehat, sehat, dan sehat sekali. Kategori tersebut dinilai berdasarkan rasio keuangan rentabilitas, likuiditas, dan solvabilitas serta berpedoman pada SK Menteri Keuangan RI Nomor 740/KMK.00/1989 dan lampiran perhitungan bobot kinerja yang direvisi dengan SK Menteri Keuangan RI Nomor 826/KMK.013/1992 tentang penilaian kinerja perusahaan

Pemerintah telah mengatur formula perhitungan klasifikasi kinerja perusahaan. Namun demikian, perhitungan secara statistik bisa dilakukan berdasarkan data-data historis. Salah satu metode statistika yang dapat digunakan yaitu metode klasifikasi yang dilakukan dengan memetakan data ke dalam suatu kelas berdasarkan data yang kelasnya sudah diketahui. Beberapa metode klasifikasi yang unggul dalam menganalisis klasifikasi dengan banyak variabel yaitu metode *K-Nearest Neighbor* (KNN) dan *Adaptive Boosting* (AdaBoost). *K-Nearest Neighbor* merupakan suatu metode yang mengelompokkan data berdasarkan kedekatan jarak suatu data dengan data lainnya (Prasetyo, 2012). Penelitian yang berkaitan dengan metode KNN dilakukan oleh Bagaskoro dkk. (2018) yang menerapkan KNN pada klasifikasi *tweets* dalam berita twitter. Penelitian tersebut menghasilkan tingkat akurasi tertinggi sebesar 90%.

Adaptive Boosting merupakan salah satu varian dari algoritma *boosting* yang dapat mengubah model lemah menjadi model kuat (Freund dan Schapire, 1999). Faizah (2017) menerapkan *boosting* dalam kasus klasifikasi dua kelas yaitu prediksi jenis kelamin dan menghasilkan akurasi terbaik sebesar 91,67%. Kasus multi kelas memiliki tingkat kesulitan yang lebih tinggi dibandingkan kasus biner sehingga diperlukan suatu kajian untuk mengetahui penerapan AdaBoost pada klasifikasi multi kelas di Indonesia.

Penelitian ini bertujuan melakukan perbandingan antara metode *K-Nearest Neighbor* dan *Adaptive Boosting* untuk mengetahui metode yang lebih baik dalam memprediksi kinerja perusahaan di Indonesia yang merupakan kasus klasifikasi multi kelas.

2. TINJAUAN PUSTAKA

Tingkat kesehatan keuangan perusahaan dapat diketahui dengan menghitung bobot kinerja perusahaan berdasarkan data laporan keuangan perusahaan. SK Menteri Keuangan RI Nomor 740/KMK.00/1989 yang direvisi dengan SK Menteri Keuangan RI Nomor 826/KMK.013/1992 menentukan perhitungan bobot kinerja perusahaan sebagai berikut :

1. Indikator utama sebesar 70% terdiri atas tiga indikator yaitu :
 - a. Rentabilitas sebesar 75% dengan target yang ditentukan sebesar 12% sehingga bobot kinerja menjadi $\frac{70\% \times 75\% \times \text{rentabilitas}}{12\%} = \frac{52,5\% \times \text{rentabilitas}}{12\%}$.
 - b. Likuiditas sebesar 12,5% dengan target yang ditentukan sebesar 150% sehingga bobot kinerja menjadi $\frac{70\% \times 12,5\% \times \text{rentabilitas}}{150\%} = \frac{8,75\% \times \text{likuiditas}}{150\%}$.
 - c. Solvabilitas sebesar 12,5% dengan target yang ditentukan sebesar 200% sehingga bobot kinerja menjadi $\frac{70\% \times 12,5\% \times \text{solvabilitas}}{200\%} = \frac{8,75\% \times \text{solvabilitas}}{200\%}$.
2. Indikator tambahan sebesar 30% meliputi profit margin, biaya produksi, dan biaya administrasi masing masing bernilai 10%.

Kinerja perusahaan dapat diketahui dari tingkat kesehatan keuangan perusahaan yang diukur berdasarkan bobot kinerja perusahaan. Berdasarkan SK Menteri Keuangan RI Nomor 826/KMK.013/1992, tingkat kesehatan keuangan perusahaan digolongkan menjadi:

1. Perusahaan sehat sekali, yaitu perusahaan dengan nilai bobot kinerja tahun terakhirnya di atas 110.
2. Perusahaan sehat, yaitu perusahaan dengan nilai bobot kinerja tahun terakhirnya di atas 100 sampai dengan 110.
3. Perusahaan kurang sehat, yaitu perusahaan dengan nilai bobot kinerja tahun terakhirnya di atas 90 sampai dengan 100.
4. Perusahaan tidak sehat, yaitu perusahaan dengan nilai bobot kinerja tahun terakhirnya di bawah atau sama dengan 90.

Klasifikasi merupakan salah satu teknik data mining yang dilakukan dengan memprediksi kelas suatu objek yang labelnya belum diketahui berdasarkan data yang labelnya telah diketahui (Han dkk., 2011). Klasifikasi termasuk metode prediksi yang membangun model dengan pelatihan (*learning*) dan menggunakan model tersebut untuk prediksi data baru. Proses klasifikasi memerlukan data latih yang berperan dalam pembentukan model untuk pengelompokan data uji.

2.1. *K-Nearest Neighbor*

K-Nearest Neighbor (KNN) adalah suatu metode yang mengelompokkan data berdasarkan kedekatan jarak suatu data dengan data lainnya (Prasetyo, 2012). Perhitungan jarak antara dua objek yang dapat digunakan adalah jarak *euclid*.

Misalkan objek A dengan nilai amatan $x = (x_1, x_2, \dots, x_p)$ dan objek B dengan nilai amatan $y = (y_1, y_2, \dots, y_3)$, jarak *euclid* antara objek A dan B (Han dkk., 2011) yaitu :

$$d_{AB} = d(x, y) = \sqrt{\sum_{i=1}^p (x_i - y_i)^2} \quad (1)$$

dengan d_{AB} adalah jarak *euclid* antara objek A dan B serta p adalah banyaknya variabel yang merepresentasikan objek A dan B .

Pemilihan nilai k yang tepat sangatlah penting pada metode KNN. Cara menentukan nilai k terbaik dapat dilakukan dengan *trial and error* seperti pada penelitian oleh Bagaskoro dkk. (2018) yaitu dengan melakukan percobaan beberapa nilai k sehingga diperoleh model dengan nilai k yang menghasilkan akurasi tertinggi.

2.2. Adaptive Boosting

Adaptive Boosting (AdaBoost) merupakan salah satu varian dari algoritma *boosting*. Menurut Freund dan Schapire (1999), algoritma *boosting* dapat mengubah model lemah (*weak learner*) menjadi model kuat (*strong learner*). *Boosting* secara umum berfokus membuat deret pohon klasifikasi dengan suatu *base learner*. Inti dari algoritma AdaBoost adalah memberikan suatu bobot lebih pada *missclassified observation*.

Metode *Classification and Regression Trees* (CART) merupakan salah satu metode yang dapat digunakan sebagai *base learner* metode AdaBoost. CART merupakan metode nonparametrik untuk melakukan analisis regresi maupun klasifikasi melalui pembentukan pohon keputusan. CART dikenal sebagai metode pemilahan rekursif biner (Lewis, 2000), artinya sekelompok data dalam simpul pohon dipilah menjadi dua simpul anak kemudian masing-masing simpul anak dipilah lagi menjadi dua simpul anak, dan seterusnya sampai berhenti karena kriteria tertentu.

Zhu dkk. (2009), memperluas algoritma AdaBoost untuk klasifikasi multi kelas yaitu :

1. Menginisialisasi bobot awal amatan yaitu $w_n = \frac{1}{N}$, dengan N adalah banyaknya amatan dalam data latih dan $n = 1, 2, \dots, N$.
2. Untuk $m = 1, 2, \dots, M$
 - a. Menetapkan fungsi classifier $y^{(m)}(x)$ pada data latih dengan bobot w_n .
 - b. Menghitung :

$$err^{(m)} = \frac{\sum_{n=1}^N w_n^{(m)} I(y^{(m)}(x_n) \neq t_n)}{\sum_{n=1}^N w_n^{(m)}} \quad (2)$$

dengan $err^{(m)}$ adalah nilai *error* iterasi ke- m , $I(y^{(m)}(x_n) \neq t_n)$ adalah fungsi indikator yang akan bernilai 1 jika amatan *missclassified* dan 0 untuk lainnya.

- c. Apabila nilai $err^{(m)} > 1 - \frac{1}{K}$, iterasi dihentikan. Apabila nilai $err^{(m)} \leq 1 - \frac{1}{K}$, maka menghitung pembobot klasifikasi yaitu :

$$\alpha^{(m)} = \ln \left[\frac{(1 - err^{(m)})}{err^{(m)}} \right] + \ln(K - 1) \quad (3)$$

dengan K adalah banyaknya kelas pada variabel dependen.

- d. Memperbarui nilai bobot amatan dengan rumus :

$$w_n^{(m+1)} = \frac{w_n^{(m)}}{z^{(m)}} \exp(\alpha^{(m)} I(y^{(m)}(x_n) \neq t_n)) \quad (4)$$

dengan $z^{(m)} = \sum_{n=1}^N w_n^{(m)}$

3. Output :

$$T(x) = \underset{k}{\operatorname{argmax}} \sum_{m=1}^M \alpha^{(m)} I(y^{(m)}(x_n) = k) \quad (5)$$

dengan k merupakan kelas data uji.

2.3. Ketepatan Hasil Klasifikasi

Metode klasifikasi yang baik menghasilkan sedikit kesalahan klasifikasi (Johnson dan Wichern, 2007). Perhitungan kesalahan klasifikasi maupun ketepatan klasifikasi dapat didasarkan pada matriks konfusi. Ukuran ketepatan klasifikasi yang dapat digunakan untuk mengetahui kebaikan suatu model yaitu *accuracy*, *precision*, dan *recall*. Rumus *accuracy* (Johnson dan Wichern, 2007), rumus *precision* dan *recall* (Han dkk., 2011) sebagai berikut :

$$Accuracy = \frac{\text{Banyaknya amatan yang diprediksi dengan benar}}{\text{Banyaknya amatan yang diprediksi}} \quad (6)$$

$$Precision(j) = \frac{\text{Banyaknya amatan kelas } j \text{ yang diprediksi dengan benar}}{\text{Banyaknya amatan kelas } j \text{ yang diprediksikan}} \quad (7)$$

$$Recall(i) = \frac{\text{Banyaknya amatan kelas } i \text{ yang diprediksi dengan benar}}{\text{Banyaknya amatan kelas } i \text{ sebenarnya}} \quad (8)$$

3. METODOLOGI PENELITIAN

Data yang digunakan dalam penelitian ini merupakan data sekunder yaitu laporan keuangan periode ke-2 tahun 2019 dari 575 perusahaan yang tercatat di Bursa Efek Indonesia yang diperoleh dari website resmi milik Bursa Efek Indonesia yaitu www.idx.co.id. Variabel yang digunakan terdiri dari variabel dependen (Y) yaitu kinerja perusahaan yang dikategorikan menjadi empat kelas yaitu Tidak Sehat (TS), Kurang Sehat (KS), Sehat (S), dan Sehat Sekali (SS), sedangkan variabel independen (X) terdiri atas tujuh rasio keuangan yaitu *Return on Assets* (ROA), *Return on Equity* (ROE), *Working Capital to Total Assets* (WCTA), *Total Asset Turn Over* (TATO), *Debt to Equity Ratio* (DER), *Longterm Debt to Asset Ratio* (LDAR), dan *Return on Investment* (ROI). *Software* yang digunakan yaitu Ms. Excel 2007 dan R 3.6.1.

Langkah-langkah yang dilakukan pada metode KNN adalah sebagai berikut :

1. Membagi data menjadi dua bagian yaitu data latih dan data uji dengan perbandingan 60%:40%, 70%:30%, dan 80%:20%.
2. Menghitung jarak *euclid* tiap data uji terhadap semua data latih.
3. Mengurutkan jarak *euclid* dari yang nilainya terkecil sampai terbesar.
4. Menentukan nilai k (banyaknya tetangga terdekat) dengan $k = 1, 2, \dots, 15$.
5. Menentukan jarak tetangga terdekat sebanyak nilai k yang telah ditentukan.
6. Menetapkan kelas mayoritas dari k tetangga terdekat sebagai kelas data uji.
7. Menghitung ketepatan hasil klasifikasi.

Langkah-langkah yang dilakukan pada metode AdaBoost sebagai berikut :

1. Membagi data menjadi dua bagian yaitu data latih dan data uji dengan perbandingan 60%:40%, 70%:30%, dan 80%:20%.
2. Menginisialisasi bobot pada data latih yaitu $w_n = \frac{1}{N}$.
3. Menentukan jumlah iterasi maksimal yaitu $M = 5, 10, 15, 20, 25, 50$.
4. Mengklasifikasikan seluruh data latih (metode CART) dan menghitung $err^{(m)}$.
5. Apabila $err^{(m)} > 1 - \frac{1}{K}$, maka iterasi dihentikan. Apabila $err^{(m)} \leq 1 - \frac{1}{K}$, maka menghitung pembobot klasifikasi $\alpha^{(m)}$.
6. Memperbarui nilai bobot amatan.
7. Menentukan prediksi kelas data uji dan menghitung ketepatan hasil klasifikasi.

Langkah yang dilakukan setelah diperoleh ketepatan hasil klasifikasi dari dua metode tersebut yaitu memilih model dengan ketepatan hasil klasifikasi terbesar dari masing-masing metode untuk dibandingkan antara dua metode tersebut dan ditarik kesimpulan.

4. HASIL DAN PEMBAHASAN

Kinerja perusahaan ditentukan dengan bobot kinerja yang diatur dalam SK Menteri Keuangan RI Nomor 826/KMK.013/1992. Perhitungan bobot kinerja perusahaan AALI :

1. Nilai rentabilitas = $\frac{\text{Laba sebelum pajak}}{\text{Aset lancar} + \text{Aset tetap}} \times 100\% = \frac{135081 \times 10^6}{5083151 \times 10^6 + 9999585 \times 10^6} \times 100\% = 0,896\%$ dikonversi menjadi $\frac{52,5 \times 0,896}{12} = 3,918$.
2. Nilai likuiditas = $\frac{\text{Aset lancar}}{\text{Hutang jangka pendek}} \times 100\% = \frac{5083151 \times 10^6}{3939961 \times 10^6} \times 100\% = 129,015\%$ dikonversi menjadi $\frac{8,75 \times 129,015}{150} = 7,526$.
3. Nilai solvabilitas = $\frac{\text{Total aset}}{\text{Total hutang}} \times 100\% = \frac{27161831 \times 10^6}{8200728 \times 10^6} \times 100\% = 331,212\%$ dikonversi menjadi $\frac{8,75 \times 331,212}{200} = 14,491$.
4. Indikator tambahan yaitu profit margin 10%, biaya produksi 10%, dan biaya administrasi 10% masing-masing dikonversi menjadi 10.

Total bobot kinerja = nilai konversi rentabilitas + nilai konversi likuiditas + nilai konversi solvabilitas + nilai indikator tambahan = $3,918 + 7,526 + 14,491 + 30 = 55,935$. Berdasarkan SK Menteri Keuangan RI Nomor 826/KMK.013/1992, perusahaan AALI masuk dalam kelas perusahaan tidak sehat karena nilai total bobot kinerjanya di bawah 90.

Analisis KNN dan AdaBoost dimulai dengan membagi data menjadi data latih dan data uji. Pembagian data pada penelitian ini dengan sepuluh kali penyampelan acak pada setiap model untuk menguji kekonsistenan penyampelan menggunakan *software* R 3.6.1 dengan proporsi pembagian data 60%:40%, 70%:30%, dan 80%:20%. Ketiga proporsi ini dipilih karena memberikan *error* klasifikasi yang kecil (Fitriyaningsih dan Basani, 2019).

4.1. Metode *K-Nearest Neighbor*

Langkah yang diperlukan setelah membagi data yaitu menghitung jarak *euclid*. Ilustrasi perhitungan akan disajikan menggunakan data uji ke-1 sebagai berikut :

$$X^*_{1(1)} = -0,03733; \quad X^*_{2(1)} = -0,3923; \quad X^*_{3(1)} = -0,49948; \quad X^*_{4(1)} = 0,14422; \\ X^*_{5(1)} = 0,50874; \quad X^*_{6(1)} = 0,40504; \quad \text{dan} \quad X^*_{7(1)} = -0,00376.$$

Data latih yang digunakan yaitu seluruh data latih dengan nilai sebagai berikut :

$$X_{1(1)} = 0,09661; \quad X_{2(1)} = 0,19233; \quad X_{3(1)} = -0,0178; \quad X_{4(1)} = 0,32149; \quad X_{5(1)} = 0,99085; \quad X_{6(1)} = 0,2539; \quad \text{dan} \quad X_{7(1)} = 0,07185. \\ X_{1(2)} = 0,02931; \quad X_{2(2)} = 0,04201; \quad X_{3(2)} = 0,27766; \quad X_{4(2)} = 0,31133; \quad X_{5(2)} = 0,43348; \quad X_{6(2)} = 0,17588; \quad \text{dan} \quad X_{7(2)} = 0,02214. \\ \vdots \\ X_{1(345)} = 0,01096; \quad X_{2(345)} = 0,0189; \quad X_{3(345)} = 0,02658; \quad X_{4(345)} = 0,35368; \\ X_{5(345)} = 0,72516; \quad X_{6(345)} = 0,08213; \quad \text{dan} \quad X_{7(345)} = 0,00812.$$

Perhitungan jarak *euclid* menggunakan persamaan (1) sebagai berikut :

$$d_{1(1)} = \sqrt{(-0,03733 - 0,09661)^2 + (-0,3923 - 0,19233)^2 + \dots + (-0,00376 - 0,07185)^2} = 73,20618 \\ d_{1(2)} = \sqrt{(-0,03733 - 0,02931)^2 + (-0,3923 - 0,04201)^2 + \dots + (-0,00376 - 0,02214)^2} = 83,23847 \\ \vdots \\ d_{1((345))} = \sqrt{(-0,03733 - 0,01096)^2 + (-0,3923 - 0,0189)^2 + \dots + (-0,00376 - 0,00812)^2} = 77,7477$$

Tabel 2. Jarak *Euclid* yang telah diurutkan untuk data uji ke-1 dengan $k = 5$

Amatan ke-	Jarak <i>Euclid</i>	Kelas
59	0,8621	1
65	2,19925	1
515	3,45605	1
378	4,2665	1
527	4,2845	1

Prediksi kelas data uji ke-1 menggunakan $k = 5$ menghasilkan jarak *euclid* dari lima tetangga terdekat seperti pada Tabel 2. Diperoleh informasi bahwa data uji ke-1 diprediksi masuk kelas 1 (TS) karena kelas mayoritas dari lima tetangganya adalah kelas 1.

Model terbaik KNN diperoleh dengan cara *trial and error*, diperoleh rata-rata ketepatan klasifikasi model (masing-masing sepuluh kali perulangan) pada Tabel 3 berikut:

Tabel 3. Ketepatan Hasil Klasifikasi Metode KNN (*Output R 3.6.1*)

Prop	k	Acc	Prec1	Prec2	Prec3	Prec4	Rec1	Rec2	Rec3	Rec4
60%:40%	1	0,75522	0,85553	0,12351	0,10067	0,60628	0,89128	0,11986	0,11354	0,48287
	2	0,73174	0,84668	0,07389	0,06361	0,54946	0,86849	0,09521	0,05354	0,46368
	3	0,78435	0,83607	0,14111	0,12500	0,67597	0,93820	0,07309	0,10000	0,46928
	4	0,78913	0,83206	0,14000	-	0,65987	0,94065	0,06817	0,02576	0,49466
	5	0,79261	0,82554	0,07000	-	0,68241	0,95378	0,02778	0,00000	0,47659
	6	0,79870	0,82806	0,18667	-	0,70391	0,95968	0,03686	0,00000	0,48379
	7	0,79478	0,82211	-	-	0,68281	0,96085	0,00833	0,00000	0,46625
	8	0,80391	0,82624	-	-	0,70111	0,96846	0,01667	0,00000	0,48079
	9	0,80565	0,82555	-	-	0,70563	0,96907	0,03409	0,00000	0,48313
	10	0,80087	0,82112	-	-	0,68198	0,96741	0,02576	0,00000	0,46646
	11	0,80087	0,82018	-	-	0,68367	0,96739	0,01742	0,00000	0,46883
	12	0,79696	0,81852	-	-	0,67041	0,96388	0,00909	0,00000	0,46481
	13	0,79696	0,81916	-	-	0,66750	0,96444	0,00909	0,00000	0,46307
	14	0,79348	0,81807	-	-	0,64985	0,95970	0,00909	0,00000	0,46345
	15	0,78913	0,81633	-	-	0,62636	0,95492	0,00000	0,00000	0,46073
70%:30%	1	0,76821	0,86161	0,12790	0,06524	0,62018	0,89549	0,12350	0,13095	0,52431
	2	0,75318	0,84956	0,13667	0,07397	0,56272	0,88990	0,07628	0,13667	0,47369
	3	0,80289	0,84524	0,12667	-	0,69876	0,94973	0,05608	0,06667	0,51665
	4	0,80751	0,84260	-	-	0,69882	0,95520	0,02020	0,06429	0,53095
	5	0,80809	0,83908	-	-	0,68732	0,95823	0,03449	0,04762	0,52026
	6	0,80867	0,83982	-	-	0,70153	0,95902	0,01429	0,00000	0,53183
	7	0,81040	0,83676	-	-	0,70580	0,96358	0,01429	0,00000	0,52267
	8	0,81214	0,83484	-	-	0,72063	0,96590	0,02338	0,00000	0,51910
	9	0,81387	0,83859	-	-	0,70547	0,96748	0,00000	0,00000	0,53031
	10	0,81156	0,83512	-	-	0,69328	0,96595	0,00000	0,00000	0,52399
	11	0,81387	0,83705	-	-	0,69427	0,96824	0,01429	0,00000	0,52389
	12	0,81214	0,83415	-	-	0,68930	0,96673	0,02857	0,00000	0,51741
	13	0,80578	0,83209	-	-	0,66980	0,96289	0,00000	0,00000	0,50429
	14	0,80751	0,83257	-	-	0,67149	0,96289	0,00000	0,00000	0,51439
	15	0,80694	0,83345	-	-	0,66653	0,96205	0,00000	0,00000	0,51549
80%:20%	1	0,75565	0,87079	0,11658	0,13667	0,61534	0,88110	0,11456	0,11762	0,52668
	2	0,75739	0,85016	0,05750	-	0,56432	0,89881	0,08095	0,14167	0,48674
	3	0,79739	0,84636	0,08333	-	0,68506	0,95070	0,07000	0,01667	0,51248
	4	0,81652	0,83766	-	-	0,75863	0,96856	0,04500	0,10000	0,54784
	5	0,82087	0,83715	-	-	0,76082	0,97664	0,02500	0,15833	0,53481
	6	0,81304	0,83000	-	-	0,73583	0,97299	0,00000	0,00000	0,52284
	7	0,81130	0,82682	-	-	0,73159	0,97040	0,00000	0,00000	0,52057
	8	0,80870	0,82321	-	-	0,71194	0,97402	0,00000	0,00000	0,49168
	9	0,81043	0,82834	-	-	0,70229	0,97507	0,00000	0,00000	0,49723
	10	0,80522	0,82475	-	-	0,67966	0,97280	0,00000	0,00000	0,47872
	11	0,81304	0,82766	-	-	0,70693	0,98230	0,00000	0,00000	0,48369
	12	0,80870	0,82291	-	-	0,71422	0,97731	0,00000	0,00000	0,47843
	13	0,80261	0,81906	-	-	0,68172	0,97403	0,00000	0,00000	0,45850
	14	0,80522	0,82284	-	-	0,67969	0,97164	0,00000	0,00000	0,48369
	15	0,80522	0,82456	-	-	0,67368	0,97161	0,00000	0,00000	0,48425

Keterangan :

Prop = Proporsi pembagian data
 k = Banyaknya tetangga terdekat
 Acc = Accuracy

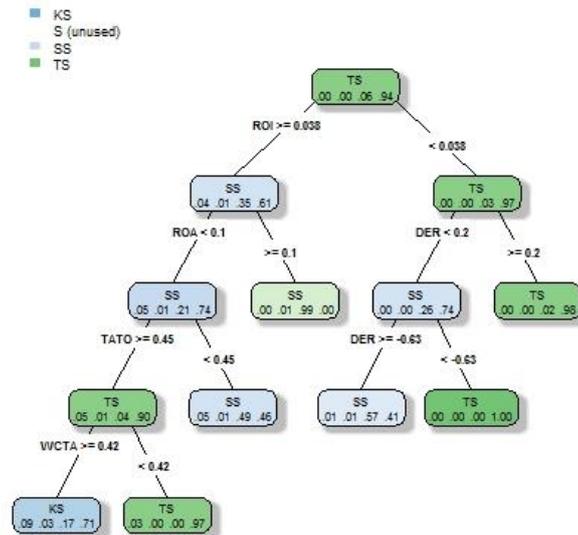
Prec1 = Precision(1)
 Prec2 = Precision(2)
 Prec3 = Precision(3)
 Prec4 = Precision(4)

Rec1 = Recall(1)
 Rec2 = Recall(2)
 Rec3 = Recall(3)
 Rec4 = Recall(4)

Tabel 3 menjelaskan bahwa model KNN terbaik yaitu model dengan proporsi 80%:20% dan $k = 5$ karena paling banyak ketepatan klasifikasi terbesarnya yaitu nilai *accuracy*, *precision*(4), dan *recall*(3) lebih besar daripada model lainnya. Pemilihan nilai k dibatasi pada $k = 15$ karena secara umum setelah $k = 9$ nilai akurasi terus menurun.

4.2. Metode Adaptive Boosting

Pembentukan pohon klasifikasi pertama dalam analisis AdaBoost mengaplikasikan bobot yang sama untuk setiap data latih yaitu $w_n = \frac{1}{N}$ dengan N adalah banyaknya amatan pada data latih dan $n = 1, 2, \dots, N$. Berdasarkan data pada lampiran, data latih sebanyak $60\% \times 575 = 345$ sehingga bobot yang digunakan yaitu $\frac{1}{345} = 0,0029$. Pohon klasifikasi optimal pertama dengan metode CART ditunjukkan pada Gambar 1 berikut :



Gambar 1. Pohon Klasifikasi Optimal Pertama

Gambar 1 digunakan sebagai penentu kelas prediksi suatu data, diperoleh data uji ke-1 dengan kriteria nilai $ROA = -0,03733$; $ROE = -0,3923$; $WCTA = -0,49948$; $TATO = 0,14422$; $DER = 9,50874$; $LDAR = 0,40504$; dan $ROI = -0,00376$ masuk pada $ROI < 0,038$ dan $DER \geq 0,2$ sehingga data uji ke-1 diprediksi masuk ke kelas TS (Tidak Sehat).

Pembentukan pohon klasifikasi pada iterasi ke-2 dilakukan dengan memperbarui bobot amatan yang dihitung berdasarkan output dari iterasi pertama. Diperoleh hasil dari 345 data latih pada iterasi pertama, terdapat 55 amatan yang merupakan *weak learner* sehingga perhitungan *error* iterasi pertama dengan persamaan (2) yaitu :

$$err^{(1)} = \frac{0,0029(1) + 0,0029(0) + \dots + 0,0029(0)}{0,0029 + 0,0029 + \dots + 0,0029} = 0,15942$$

Diperoleh $err^{(1)} = 0,15942 < 1 - \frac{1}{K} = 0,75$ dengan banyaknya kelas pada variabel dependen yaitu $K = 4$ sehingga dilanjutkan menghitung pembobot klasifikasi menggunakan persamaan (3) sebagai berikut :

$$\alpha^{(1)} = \ln \left[\frac{(1 - 0,15942)}{0,15942} \right] + \ln(4 - 1) = 2,76116$$

Langkah selanjutnya yaitu memperbarui bobot dengan persamaan (4), diperoleh :

$$\begin{aligned} w_1^{(2)} &= \frac{0,0029}{1} \exp(2,76116(1)) = 0,04585 \\ w_2^{(2)} &= \frac{0,0029}{1} \exp(2,76116(0)) = 0,0029 \\ &\vdots \\ w_{345}^{(2)} &= \frac{0,0029}{1} \exp(2,76116(0)) = 0,0029 \end{aligned}$$

Diperoleh bobot amatan baru yaitu $w_n = (0,04585; 0,0029; \dots; 0,0029)$ yang selanjutnya digunakan untuk membentuk pohon klasifikasi pada iterasi ke-2. Cara yang sama dilakukan untuk memperoleh bobot amatan iterasi selanjutnya sampai iterasi ke-M.

Penerapan AdaBoost pada prediksi data dengan iterasi maksimal 5 menghasilkan jumlahan $\alpha^{(m)}$ untuk kelas TS(1) sebesar 9,9821; kelas KS(2) sebesar 0, kelas S(3) sebesar 0; dan kelas SS(4) sebesar 2,83255. Hasil akhir prediksi kelas data uji ditentukan oleh jumlahan $\alpha^{(m)}$ terbesar sehingga data uji ke-1 diprediksi masuk kelas 1 (Tidak Sehat).

Model terbaik AdaBoost diperoleh dengan cara *trial and error*, diperoleh rata-rata ketepatan klasifikasi model (masing-masing sepuluh kali perulangan) pada Tabel 4 berikut:

Tabel 4. Ketepatan Hasil Klasifikasi Metode AdaBoost (*Output R 3.6.1*)

Prop	M	Acc	Prec1	Prec2	Prec3	Prec4	Rec1	Rec2	Rec3	Rec4
60%:40%	5	0,80522	0,89247	0,19556	0,12750	0,63785	0,91851	0,15337	0,06151	0,65477
	10	0,80261	0,89691	0,19958	0,17500	0,65801	0,91558	0,16072	0,07579	0,67071
	15	0,81000	0,89336	0,27070	0,07968	0,65397	0,91919	0,17638	0,08889	0,66857
	25	0,83522	0,89994	0,36553	0,02250	0,74475	0,94601	0,23777	0,05333	0,66678
	30	0,83174	0,90448	0,30675	0,07500	0,72496	0,93973	0,21146	0,09679	0,70085
	50	0,82130	0,89100	0,27283	0,10000	0,72882	0,93715	0,18801	0,03333	0,67679
70%:30%	5	0,80925	0,89821	0,16803	0,05000	0,65653	0,92123	0,12730	0,02917	0,69104
	10	0,80231	0,89784	0,16024	0,07500	0,66792	0,90974	0,16651	0,03000	0,66414
	15	0,82832	0,90356	0,35929	0,11667	0,70376	0,94233	0,27678	0,07667	0,67822
	25	0,8104	0,88429	0,22069	0,12024	0,70345	0,93026	0,20152	0,10095	0,63359
	30	0,8237	0,89131	0,37845	0,10000	0,69760	0,94052	0,22063	0,05095	0,69322
	50	0,82717	0,89958	0,37524	0,16762	0,69076	0,93864	0,22083	0,11190	0,66842
80%:20%	5	0,78783	0,89694	0,14076	0,00000	0,61060	0,91081	0,17167	0,00000	0,61229
	10	0,81826	0,90415	0,09000	0,10000	0,70250	0,92048	0,06762	-	0,72942
	15	0,81391	0,89769	0,28667	0,12000	0,69118	0,92526	0,20160	0,05333	0,67547
	25	0,81391	0,90399	0,21381	0,05000	0,64701	0,91729	0,13774	0,02000	0,68789
	30	0,84522	0,90555	0,22190	0,00000	0,76910	0,94593	0,18929	0,00000	0,70433
	50	0,83217	0,90721	0,16583	0,05000	0,68833	0,93882	0,13929	0,02500	0,70595

Keterangan :

Prop = Proporsi pembagian data
M = Iterasi maksimal
Acc = Accuracy

Prec1 = Precision(1)
Prec2 = Precision(2)
Prec3 = Precision(3)
Prec4 = Precision(4)

Rec1 = Recall(1)
Rec2 = Recall(2)
Rec3 = Recall(3)
Rec4 = Recall(4)

Tabel 4 menginformasikan model AdaBoost terbaik yaitu model dengan proporsi pembagian data 80%:20% dan $M = 30$ karena paling banyak nilai ketepatan klasifikasi terbesarnya yaitu nilai *accuracy* dan *precision(4)* lebih besar dibandingkan model lainnya.

4.3. Ketepatan Hasil Klasifikasi

Langkah yang dilakukan setelah memperoleh nilai *accuracy*, *precision*, dan *recall* metode KNN dan AdaBoost yaitu membandingkan ketepatan klasifikasi dua metode. Rangkuman ketepatan klasifikasi metode KNN dan AdaBoost dalam Tabel 5 berikut :

Tabel 5. Rangkuman Ketepatan Hasil Klasifikasi Metode KNN dan AdaBoost

Metode	Acc	Prec1	Prec2	Prec3	Prec4	Rec1	Rec2	Rec3	Rec4
KNN	0,82087	0,83715	-	-	0,76082	0,97664	0,02500	0,15833	0,53481
AdaBoost	0,84522	0,90555	0,22190	0	0,76910	0,94593	0,18929	0	0,70433

Keterangan :

Acc = Accuracy

Prec1 = Precision(1)

Prec2 = Precision(2)

Prec3 = Precision(3)

Prec4 = Precision(4)

Rec1 = Recall(1)

Rec2 = Recall(2)

Rec3 = Recall(3)

Rec4 = Recall(4)

Tabel 5 memperlihatkan bahwa metode yang lebih baik untuk memprediksi kelas data adalah metode AdaBoost karena model terbaik AdaBoost memiliki lebih banyak nilai *accuracy*, *precision*, dan *recall* yang nilainya lebih besar dibandingkan model terbaik metode KNN.

5. KESIMPULAN

1. Prediksi kinerja perusahaan pada penelitian ini dilakukan dengan membagi data menjadi data latih dan data uji kemudian mengklasifikasikan data menggunakan metode *K-Nearest Neighbor* dan *Adaptive Boosting*.
2. Penentuan nilai k dalam metode *K-Nearest Neighbor* melalui pengujian didapatkan bahwa setelah $k = 9$, akurasi klasifikasi terus menurun sehingga nilai k dibatasi pada $k = 15$. Nilai k yang terlalu besar menyebabkan tetangga terdekat yang terpilih mungkin tidak relevan karena sebenarnya termasuk kelas lain.
3. Analisis menggunakan metode KNN dan AdaBoost dengan sepuluh kali perulangan pada masing-masing model menghasilkan model terbaik KNN adalah model dengan proporsi pembagian 80%:20% dan $k = 5$ dengan *accuracy* sebesar 0,82087 sedangkan model terbaik AdaBoost adalah model dengan proporsi pembagian 80%:20% dan $M = 30$ dengan *accuracy* sebesar 0,84522.
4. Model terbaik AdaBoost memiliki lebih banyak nilai *accuracy*, *precision*, dan *recall* yang lebih besar dibandingkan nilai pada model terbaik KNN sehingga untuk memprediksi kinerja perusahaan di Indonesia tahun 2019 sebaiknya dilakukan menggunakan metode AdaBoost.

DAFTAR PUSTAKA

- Bagaskoro, G., N. dkk. 2018. *Penerapan Klasifikasi Tweets pada Berita Twitter Menggunakan Metode K-Nearest Neighbor dan Query Expansion Berbasis Distributional Semantic*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol. 2, No. 10 Hal. 3849-3855.
- Bursa Efek Indonesia. [online]. www.idx.co.id. (diakses Selasa, 29 Oktober 2019).
- Freund, Y. dan Schapire, R. E. 1999. *A Short Introduction to Boosting*. Journal of Japanese Society for Artificial Intelligence, 14(5) 771-780.
- Fitriyaningsih, I. dan Basani, Y. 2019. *Prediksi Kejadian Banjir dengan Ensemble Machine Learning Menggunakan BP-NN dan SVM*. Jurnal Teknologi dan Sistem Komputer Vol. 7, No. 3 : Hal. 93-97.
- Han, J., Kamber, M., dan Pei, J. 2011. *Data Mining : Concepts and Techniques, Third Edition*. Waltham : Morgan Kaufmann Publishers.
- Johnson, R. A. dan Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis*. New Jersey : Pearson Prentice Hall.
- Lewis, R. J. 2000. *An Introduction to Classification and Regression Trees (CART) Analysis*. Presented at the 2000 Annual Meeting of Society for Academic Emergency Medicine of Sanfrancisco. California.
- Menteri Keuangan Republik Indonesia. 1992. Surat Keputusan Menteri Keuangan Republik Indonesia Nomor 826/KMK.013/1992. Tentang Sistem Penilaian Kinerja BUMN.
- Prasetyo, E. 2012. *DATA MINING : Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta : ANDI.
- Pulloh, J. dkk. 2016. *Analisis Rasio Keuangan untuk Menilai Kinerja Keuangan Perusahaan (Studi Kasus PT. HM Sampoerna Tbk yang Terdaftar di Bursa Efek Indonesia)*. Jurnal Administrasi Bisnis Vol. 33, No.1 : Hal. 89-97.
- Zhu, J. dkk. 2009. *Multi-class AdaBoost*. Statistics and Its Interface, 2, pp.349-360.