



Klasifikasi Ketertarikan Mahasiswa Pembelajaran Berbasis Video Business Intelligence Menggunakan Adaboost

Anis Zubair¹

¹ Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang. Jl. Terusan Dieng
No.57-59, Pisang Candi, Kec. Sukun, Kota Malang, Jawa Timur 65146
e-mail: anis.zubair@unmer.ac.id¹

ABSTRAK

Kata Kunci:

AdaBoost
Precision
Recall
F1

Salah satu syarat yang harus dilalui oleh mahasiswa agar dapat lulus dan meraih gelar sarjana S1 adalah membuat skripsi. Pembuatan skripsi yang sesuai dengan apa yang pernah dipelajari selama perkuliahan adalah prioritas utama para mahasiswa semester akhir. Pada masa pandemi kebanyakan perkuliahan diselenggarakan tidak dalam bentuk tatap muka secara langsung. Namun pada masa pandemi kebanyakan perkuliahan diselenggarakan dalam bentuk pembelajaran berbasis video. Dengan dua alasan di atas penelitian ini berfokus untuk mengetahui klasifikasi ketertarikan mahasiswa untuk membuat skripsi berdasarkan mata kuliah dalam bentuk pembelajaran berbasis video yang pernah ditempuh. Metode yang digunakan membuat klasifikasi ini adalah metode boosting adaptif yang populer dengan istilah AdaBoost. AdaBoost dipilih karena menunjukkan kinerja yang baik pada beberapa penelitian sebelumnya. Akurasi klasifikasi metode ini ditentukan menggunakan Precision, Recall, dan F1. Hasil penelitian menunjukkan bahwa Precision tertinggi adalah Precision untuk kelas Tidak, Recall tertinggi adalah Recall untuk kelas Ragu-ragu, dan F1 tertinggi adalah F1 untuk kelas Ragu-ragu dan Tidak.

ABSTRACT

Keyword:

AdaBoost
Precision
Recall
F1

One of the requirements that must be passed by students in order to graduate and earn a bachelor's degree is to write a thesis. Writing a thesis that is in accordance with what has been learned during college is a top priority for final semester students. During the pandemic, most lectures are not held face-to-face. However, during the pandemic, most lectures are held in the form of video-based learning. With the above two reasons, this research focuses on knowing the classification of student interest in writing a thesis based on courses in the form of video-based learning that has been taken. The method used to make this classification is an adaptive boosting method which is popularly known as AdaBoost. AdaBoost was chosen because it showed good performance in several previous studies. The classification accuracy of this method is determined using Precision, Recall, and F1. The results showed that the highest Precision was Precision for the No class, the highest Recall was Recall for the Doubtful class, and the highest F1 was F1 for the Doubtful class and No.



PENDAHULUAN

Business Intelligence merupakan salah satu jenis pekerjaan yang banyak ditawarkan di LinkedIn. Tercatat bahwa Shopee, MNC Media, dan Zenius merupakan beberapa perusahaan yang menawarkan posisi sebagai Business Intelligence LinkedIn, 2021 [1]. Di Program Studi Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang Business Intelligence diajarkan sebagai mata kuliah yang ditawarkan pada semester 7. Di era pandemi, sebagai pengganti tatap muka di kelas, para mahasiswa mengikuti kuliah dengan menggunakan bermacam-macam media. Beberapa media yang digunakan dalam proses mengajar-belajar ini meliputi Whatsapp, Google Classroom, Google Meet, Microsoft Teams, Zoom, Youtube, dan lain-lain.

Di Program Studi Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang Business Intelligence diajarkan, selain menggunakan Whatsapp dan Zoom, menggunakan video yang ditayangkan di Youtube. Pemilihan pembelajaran menggunakan video di Youtube karena video adalah tren media pembelajaran di era revolusi industri 4.0 Putry, 2020 [2]. Istilah industri 4.0 pertama kali digunakan oleh para pakar di Hannover Fair pada April 2011 Kinzel, 2012 [3]. Di era revolusi industri 4.0 ada lima teknologi yang menjadi pilar utama dalam mengembangkan sebuah industri siap digital. Masing-masing adalah Internet of Things, Big Data, Artificial Intelligence, Cloud Computing dan Additive Manufacturing Aptika, 2021 [4].

Penelitian ini dibuat untuk mengetahui klasifikasi ketertarikan mahasiswa untuk membuat skripsi Business Intelligence setelah mengikuti pembelajaran berbasis video. Metode klasifikasi yang digunakan pada penelitian ini adalah AdaBoost. AdaBoost adalah algoritma boosting yang diperkenalkan pertama kali oleh Freund dan Schapire pada tahun 1997 [5].

Berbagai penelitian terdahulu menunjukkan bahwa AdaBoost memiliki kinerja yang baik sebagai salah satu metode klasifikasi. Wang 2019 [6] menyatakan bahwa algoritma AdaBoost memiliki keunggulan dalam menahan overfitting. Overfitting mengacu pada pemodelan data pelatihan yang terlalu baik. Pemodelan data pelatihan yang terlalu baik di luar dugaan menyebabkan kinerjanya menjadi turun.

Bisri 2015 [7] menggunakan AdaBoost pada decision tree untuk mendapatkan hasil yang optimal dan tingkat akurasi yang baik. Hasil percobaan yang diperoleh menunjukkan bahwa penentuan kelulusan mahasiswa dengan metode decision tree dan AdaBoost terbukti mampu menyelesaikan masalah ketidakseimbangan kelas. Selain dapat menangani masalah ketidakseimbangan kelas, metode decision tree dan AdaBoost terbukti mampu meningkatkan akurasi dan dapat menurunkan tingkat kesalahan klasifikasi.

Hasil penelitian Xiao et al. 2019 [9] menunjukkan bahwa model kombinasi LSTM-AdaBoost mengungguli LSTM (Long Short-Term Memory) dan AdaBoost yang dijalankan secara terpisah.



Selain itu, kombinasi LSTM-AdaBoost juga mengungguli support vector regression, backpropagation neural network, dan stacking LSTM-AdaBoost. Hasil penelitian juga menunjukkan bahwa model kombinasi LSTM-AdaBoost sangat menjanjikan untuk prediksi suhu permukaan laut anomali jangka pendek dan menengah.

Penipuan kartu kredit adalah masalah serius dalam layanan keuangan. Miliaran dolar hilang karena penipuan kartu kredit setiap tahun. Studi penelitian tentang penipuan kartu kredit yang dilakukan oleh Randhawa et al. 2018 [10] menunjukkan bahwa metode hibrid yang menggunakan AdaBoost dan metode voting mayoritas menghasilkan nilai MCC (Matthews Correlation Coefficient) yang sempurna.

Tanah longsor merupakan jenis bencana alam yang umum terjadi yang membawa ancaman besar bagi kehidupan manusia dan ekonomi pembangunan di seluruh dunia. Wu et al. 2020 [11] memetakan kerawanan longsor menggunakan Alternating Decision tree (ADTree), ADTree dengan Bagging (ADTree-Bagging) dan ADTree dengan AdaBoost (ADTree-AdaBoost). Hasilnya menunjukkan tingkat keberhasilan ADTree-AdaBoost adalah keberhasilan terbaik dalam penelitian ini.

METODE

Menurut berbagai metode penelitian yang ada, penelitian yang dilakukan peneliti adalah penelitian eksperimen. Penelitian eksperimen berarti penelitian ini diarahkan untuk mengetahui pengaruh beberapa peubah terhadap peubah lain.

Peubah yang digunakan pada penelitian terbagi menjadi dua. Masing-masing adalah peubah kuantitatif dan kualitatif. Yang termasuk peubah kuantitatif adalah durasi video yang diinginkan, keterbacaan teks, kualitas video, komposisi warna, kejelasan suara, dan kemudahan memahami materi. Sedangkan yang termasuk peubah kualitatif adalah pulau asal, perangkat yang digunakan, background dalam video, dan ketertarikan mengambil skripsi Business Intelligence.

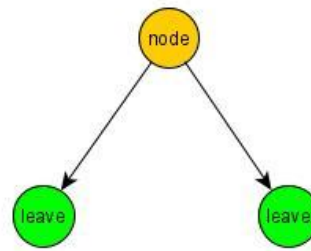
Subjek penelitian ini adalah seluruh mahasiswa Program Studi Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang yang mengikuti kuliah Business Intelligence berbasis video yang dipublikasikan di Youtube. Objek penelitian ini adalah ketertarikan mahasiswa untuk membuat skripsi tentang Business Intelligence. Penelitian ini adalah penelitian sampel jenuh karena subjek penelitian ini adalah seluruh mahasiswa yang mengikuti kuliah Business Intelligence berbasis video yang dipublikasikan di Youtube. Penelitian sampel jenuh dipilih karena peneliti ingin membuat generalisasi dengan kesalahan terkecil. Dengan demikian kesimpulan yang didapatkan dari penelitian ini berlaku untuk semua subjek penelitian.

Metode yang digunakan untuk mengumpulkan data adalah metode kuesioner. Metode kuesioner disebut juga dengan metode pengumpulan data secara tidak langsung, karena peneliti tidak bertanya jawab langsung dengan subjek penelitian. Metode kuesioner yang digunakan adalah metode

kuesioner tertutup. Dengan kata lain jawaban dari setiap pertanyaan kuesioner berupa pilihan yang sudah ditentukan oleh peneliti.

Teknik analisis data yang digunakan adalah klasifikasi AdaBoost. Ada beberapa istilah yang digunakan dalam AdaBoost. Istilah-istilah itu adalah stump dan amount of say.

Pohon keputusan (decision tree) dengan satu titik split adalah stump keputusan (decision stump). Dengan kata lain stump keputusan adalah pohon (tree) dengan satu titik (node) dan dua daun (leave). Stump juga termasuk pembelajar yang lemah (weak learner). Pembelajar yang lemah menghasilkan klasifikasi yang sedikit lebih baik daripada klasifikasi acak (Brownlee, 2017).



Gambar 1. Stump Keputusan

Shin (2020) menyatakan bahwa amount of say adalah rumus yang berkaitan dengan kesalahan total (total error). Rumus amount of say dinyatakan di bawah ini.

$$\text{amount of say} = \frac{1}{2} \log \left(\frac{1 - \text{total error}}{\text{total error}} \right) \quad (1)$$

Langkah-langkah AdaBoost adalah sebagai berikut:

1. Pembuatan inisialisasi bobot tiap-tiap baris. Pembobotan yang digunakan adalah $D_1(i) = 1/n$ untuk $i = 1, \dots, n$.
2. Pembuatan stump keputusan.
3. Penghitungan amount of say.
4. Pengubahan bobot.
5. Pembuatan normalisasi bobot.
6. Pembuatan dataset baru.
7. Jika data belum semuanya terklasifikasi secara benar, kembali ke langkah 1.

Akurasi klasifikasi AdaBoost dihitung menggunakan Precision, Recall, dan F1. Hasil pengujian akurasi klasifikasi AdaBoost ini berdasarkan confusion matrix yang dibuat.

Penelitian-penelitian sebelumnya menunjukkan bahwa Precision dan Recall adalah akurasi yang banyak digunakan pada penelitian klasifikasi. Selain menggunakan istilah Precision dan Recall, beberapa peneliti juga menggunakan istilah lain untuk Precision dan Recall. Powers 2020 [12] menggunakan istilah Recall atau juga disebut Sensitivity dan istilah Precision atau juga disebut Confidence. Juba dan Le (2019) menunjukkan bahwa Precision dan Recall setara dengan akurasi



tertimbang ketidakseimbangan. Lee et al. 2018 [13] menunjukkan bahwa Precision dan Recall dapat digunakan sebagai deteksi anomali berbasiskan range.

Metrik atau rumus yang digunakan untuk pengujian klasifikasi berdasarkan metrik yang digunakan oleh Han [14]. Metrik tersebut menggunakan istilah P sebagai kependekan dari Positives, N sebagai kependekan dari Negatives, TP sebagai kependekan dari True Positives, TN sebagai kependekan dari True Negatives, FP sebagai kependekan dari False Positives, dan FN sebagai kependekan dari False Negatives. Letak TP, FN, FP, dan TN dinyatakan pada tabel di bawah ini.

Tabel 1. Confusion Matrix 2 x 2

		Predicted	
		A	b
Actual	a	TP	FN
	b	FP	TN

Rumus-rumus di bawah ini berdasarkan letak TP, FN, FP, dan TN yang dinyatakan pada tabel sebelumnya. Classification Accuracy juga disebut Recognition Rate dinyatakan dengan rumus di bawah ini.

$$CA = \frac{TP + TN}{P + N} \quad (2)$$

Precision yang juga disebut Confidence dinyatakan dengan rumus di bawah ini.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

Recall yang juga disebut Sensitivity atau True Positive Rate dinyatakan dengan rumus di bawah ini.

$$Recall = \frac{TP}{P} \quad (4)$$

F1 yang juga disebut F atau F-score atau Harmonic Mean of Precision and Recall dinyatakan dengan rumus di bawah ini.

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (5)$$

Nilai F1 terletak di antara 0 dan 1. Nilai 1 adalah nilai terbaik dan 0 adalah nilai terburuk. Jika F1 memiliki nilai yang mendekati terbaik mengindikasikan bahwa model klasifikasi yang digunakan memiliki precision dan recall yang baik.

Penelitian ini menggunakan tiga kelas. Masing-masing adalah Ya, Tidak, dan Ragu-ragu. Oleh karena itu klasifikasi yang digunakan adalah klasifikasi multikelas. Klasifikasi multikelas adalah klasifikasi yang melibatkan lebih daripada dua kelas [15].



HASIL DAN PEMBAHASAN

Hasil-hasil utama penelitian ini dinyatakan dalam bentuk tabel. Sedangkan pembahasannya dinyatakan dalam bentuk narasi yang menyertai tabel.

Deskripsi Data

Data yang didapatkan sebanyak 154 baris dan 10 kolom. Deskripsi data menurut tiap-tiap kolom dinyatakan pada tabel di bawah ini.

Tabel 2. Deskripsi Data

Nama peubah	Type data	Nilai data
Pulau Asal	kualitatif	Sumatera, Kalimantan, Jawa, Nusa Tenggara & Bali, Sulawesi, Maluku & Papua
Perangkat yang digunakan	kualitatif	Desktop, Laptop, HP
Durasi video yang diinginkan	kuantitatif	1, 2, 3
Backsound dalam video	kualitatif	Ya, Tidak
Keterbacaan teks	kuantitatif	1, 2, 3, 4, 5
Kualitas video	kuantitatif	1, 2, 3, 4, 5
Komposisi warna	kuantitatif	1, 2, 3, 4, 5
Kejelasan suara	kuantitatif	1, 2, 3, 4, 5
Kemudahan memahami materi	kuantitatif	1, 2, 3, 4, 5
Tertarik mengambil skripsi Business Intelligence?	kualitatif	Ya, Tidak, Ragu-ragu

Durasi video yang diinginkan diberi nilai 1, 2 dan 3. Nilai 1 berarti kurang dari 5 menit, nilai 2 berarti antara 5 sampai dengan 10 menit, dan nilai 3 berarti lebih dari 10 menit.

Keterbacaan teks, kualitas video, komposisi warna, kejelasan suara, dan kemudahan memahami materi diberi nilai 1, 2, 3, 4 dan 5. Nilai 1 berarti kurang sekali, nilai 2 berarti kurang, nilai 3 berarti cukup, nilai 4 berarti baik, dan nilai 5 berarti baik sekali.

Prediksi AdaBoost

Tabel di bawah ini merupakan cuplikan sebagian data hasil prediksi. Data hasil prediksi yang ditampilkan adalah data nomor 2, 11, dan 21. Hasil prediksi 1,00 : 0,00 : 0,00 → Ragu-ragu berarti peluang Ragu-ragu 1,00, peluang Tidak 0,00, dan peluang Ya 0,00. Dengan demikian hasil prediksinya adalah Ragu-ragu. Hasil prediksi 0,00 : 1,00 : 0,00 → Tidak berarti peluang Ragu-ragu 0,00, peluang Tidak 1,00, dan peluang Ya 0,00. Dengan demikian hasil prediksinya adalah Tidak. Hasil prediksi 0,05 : 0,00 : 0,95 → Ya berarti peluang Ragu-ragu 0,05, peluang Tidak 0,00, dan peluang Ya 0,95. Dengan demikian hasil prediksinya adalah Tidak.



Tabel 3. Prediksi AdaBoost

Nomor	Prediksi AdaBoost	Target
.		
.		
2	1,00 : 0,00 : 0,00 → Ragu-ragu	Ragu-ragu
.		
.		
11	0,00 : 1,00 : 0,00 → Tidak	Tidak
.		
.		
21	0,05 : 0,00 : 0,95 → Ya	Ya
.		
.		
.		

Hasil prediksi menggunakan AdaBoost dinyatakan pada tabel confusion matrix di bawah ini. Isian pada tabel menunjukkan banyak data. Baris menunjukkan actual dan kolom menunjukkan predicted. Actual adalah nilai data yang dipilih oleh subjek penelitian. Predicted adalah nilai data yang dihasilkan oleh AdaBoost.

Tabel 4. Confusion Matrix AdaBoost

		Predicted			Jumlah
		Ragu-ragu	Tidak	Ya	
Actual	Ragu-ragu	94	1	1	96
	Tidak	1	35	1	37
	Ya	5	0	16	21
	Jumlah	100	36	18	154

Isi sel tabel di atas menunjukkan ada 94 mahasiswa yang diduga Ragu-ragu dan kenyataannya Ragu-ragu. Ada 35 mahasiswa yang diduga Tidak dan kenyataannya Tidak. Ada 16 mahasiswa yang diduga Ya dan kenyataannya Ya. Ada 5 mahasiswa yang diduga Ragu-ragu dan kenyataannya Ya. Ada 1 mahasiswa yang diduga Ragu-ragu dan kenyataannya Tidak. Ada 1 mahasiswa yang diduga Tidak dan kenyataannya Ragu-ragu. Ada 1 mahasiswa yang diduga Ya dan kenyataannya Ragu-ragu. Ada 1 mahasiswa yang diduga Ya dan kenyataannya Tidak.

Akurasi Prediksi AdaBoost

Akurasi prediksi menggunakan AdaBoost yang digunakan adalah CA (Classification Accuracy), Precision, Recall, dan F1 [16]. Classification Accuracy yang didapatkan adalah 0,942. Sedangkan Precision, Recall, dan F1 dinyatakan pada tabel di bawah ini.



Tabel 5. Akurasi AdaBoost

	Precision	Recall	F1
Ragu-ragu	0,940	0,979	0,959
Tidak	0,972	0,946	0,959
Ya	0,889	0,762	0,821

Precision tertinggi menurut tabel di atas adalah Precision untuk kelas Tidak. Sedangkan Precision terendah adalah untuk kelas Ya. Recall tertinggi menurut tabel di atas adalah Recall untuk kelas Ragu-ragu. Sedangkan Recall terendah adalah untuk kelas Ya. F1 tertinggi menurut tabel di atas adalah F1 untuk kelas Ragu-ragu dan Tidak. Sedangkan F1 terendah adalah untuk kelas Ya. Hasil penelitian ini menunjukkan bahwa F1 memiliki kecenderungan mendekati 1. Dengan demikian hal ini mengindikasikan bahwa model klasifikasi yang digunakan memiliki precision dan recall yang baik.

SIMPULAN

Hasil penelitian di atas menunjukkan bahwa sebagian besar mahasiswa yang mengisi kuesioner tidak terlalu tertarik untuk menjadikan Business Intelligence menjadi bahan skripsi. Hal ini dapat dilihat bahwa mahasiswa yang menjawab Ragu-ragu sebanyak 100 dan mahasiswa yang menjawab Tidak sebanyak 36 dari 154 mahasiswa yang disurvei. Akurasi hasil penelitian dapat ditunjukkan oleh nilai F1 yang mendekati 1 yang menunjukkan bahwa akurasi klasifikasi ini termasuk baik.

Hasil penelitian ini juga menunjukkan bahwa pembelajaran Business Intelligence menggunakan video di Youtube masih banyak yang perlu diperbaiki agar dapat dipilih mahasiswa sebagai bahan skripsi. Dengan demikian perlu penelitian faktual dan akurat tentang faktor-faktor yang menyebabkan mahasiswa tidak menjadikan Business Intelligence menjadi bahan skripsi.

DAFTAR RUJUKAN

- [1] LinkedIn. [online] Tersedia di: <<https://id.linkedin.com/jobs/business-intelligence-analyst-jobs?position=1&pageNum=0>> [Diakses 10 November 2021], 2021.
- [2] H. M. E. Putry, et al., Video Based Learning Sebagai Tren Media Pembelajaran di Era 4.0. Tarbiyatuna: Jurnal Pendidikan Ilmiah, 5(1), hal. 1-24, 2020.
- [3] H. Kinzel, Industry 4.0 - Where does this leave the Human Factor? [online] Tersedia di: <<http://www.cujucr.com/downloads/Individual%20Articles/15/vol15%20Holger%20Kinzel.pdf>> [Diakses 10 November 2021], 2012.
- [4] Revolusi Industri 4.0 <https://aptika.kominfo.go.id/2020/01/revolusi-industri-4-0/> [Diakses 10 November 2021].
- [5] Y. Freund & R. E. Schapire, A Decision-Theoretic Generalization of On-Line Learning and



- an Application to Boosting. *Journal of Computer and System Sciences*, 55 (1), hal. 119-139, 1997.
- [6] F. Wang et al., Feature Learning Viewpoint of Adaboost and a New Algorithm. *IEEE Access*, 7, hal. 149890-149899, 2019.
- [7] C. Xiao et al., Short and Mid-term Sea Surface Temperature Prediction. *Remote Sensing of Environment*, 233(8), hal. 1-18, 2019.
- [8] K. Randhawa et al., Credit Card Fraud Detection Using AdaBoost and Majority Voting. *IEEE Access*, 6, hal. 14277 - 14284, 2018.
- [9] Y. Wu et al., Application of Alternating Decision Tree with AdaBoost and Bagging Ensembles for Landslide Susceptibility Mapping. *Catena*, 187, hal. 1 - 17, 2020.
- [10] J. Han, et al., *Data Mining: Concepts and Techniques*, Third Edition. Massachusetts: Morgan Kaufmann Publishers, 2012.
- [11] B. Juba & H. S. Le, Precision-Recall versus Accuracy and the Role of Large Data Sets. *The Thirty-Third AAAI Conference on Artificial Intelligence*, hal. 4039 - 4048, 2019.
- [12] M. Grandini et al. Metrics for Multi-Class Classification: An Overview. *arXiv:2008.05756*, hal. 1 - 17, 2020.
- [13] T. J. Lee et al, Precision and Recall for Range-Based Anomaly Detection. *arXiv:1801.03175*, hal. 1 - 3, 2018.
- [14] J. Brownlee, *Master Machine Learning Algorithms*. Melbourne, 2017.
- [15] T. Shin, A Mathematical Explanation of AdaBoost in 5 Minutes[online] Tersedia di: <<https://towardsdatascience.com/a-mathematical-explanation-of-adaboost-4b0c20ce4382>> [Diakses 17 April 2022], 2020.
- [16] D. M. W. Powers, Evaluation: From Precision, Recall And F-Measure To Roc, Informedness, Markedness & Correlation. *arXiv:2010.16061*, hal. 37 - 63, 2020.