

# Perbandingan Klasifikasi DT, RF, GB dan SVM dalam Implementasi CRISP-DM terhadap Kelulusan Mata Kuliah

<sup>1</sup>Shedriko, <sup>2</sup>Muhammad Firdaus  
<sup>1,2</sup>Universitas Indraprasta PGRI  
Jakarta, Indonesia

<sup>1</sup>shedriko@gmail.com, <sup>2</sup>dasurichi@gmail.com

## \*Penulis Korespondensi

Diajukan : 03/11/2025  
Diterima : 23/12/2025  
Dipublikasi : 01/01/2026

## ABSTRAK

Penelitian dilakukan terhadap kelulusan mahasiswa pada mata kuliah Pengantar Teknologi Informasi atau PTI di Universitas XYZ. Universitas ini memiliki keunikan yaitu dengan memiliki jumlah mahasiswa yang relatif besar dibandingkan universitas lainnya, baik domestik maupun mancanegara. Penelitian ini merupakan kelanjutan dari penelitian yang telah dilakukan sebelumnya oleh penulis namun dengan membandingkan empat klasifikasi lainnya yang berbeda. Dengan menggunakan *digital-based method* atau metode berdasarkan digital, penelitian ini membandingkan empat jenis klasifikasi, yaitu Pohon Keputusan atau *Decision Tree* (DT), Hutan (kumpulan pohon) Acak atau *Random Forest* (RT), Penguatan Gradien atau *Gradient Boosting* (GB) dan Mesin Vektor Pendukung atau *Support Vector Machine* (SVM). Dengan mengambil sampel sebanyak 175 mahasiswa yang mengikuti mata kuliah pengantar teknologi informasi, pengujian dilakukan dengan memasukkan keempat klasifikasi tersebut ke dalam “proses standar lintas industri untuk penambangan data” atau CRISP-DM (*Cross-industry Standard Process for Data Mining*). Tujuan penelitian ini adalah mendapatkan metoda klasifikasi terbaik dari empat jenis klasifikasi tersebut. Hasilnya menunjukkan bahwa SVM masih memberikan nilai yang terbaik untuk tingkat akurasi, serta untuk beberapa parameter penilaian lainnya, kecuali untuk ketepatan atau *precision* yang didominasi oleh GB. Kondisi SVM dan GB tersebut menjadi temuan kebaruan atau *novelty* pada penelitian ini.

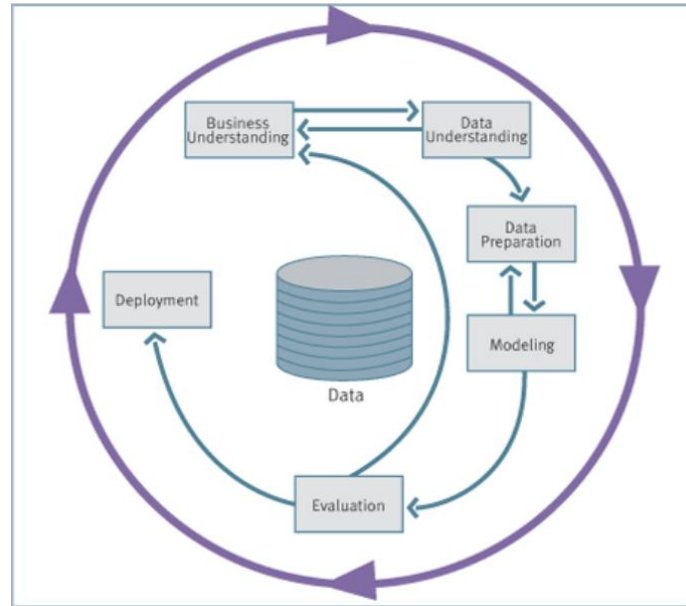
**Kata Kunci:** crisp-dm, klasifikasi, perbandingan, kelulusan, kuliah.

## I. PENDAHULUAN

Penelitian ini merupakan penelitian lanjutan dari peneliti yang sama yang berkaitan dengan metode CRISP-DM (*Cross-industry Standard Process for Data Mining*). Penelitian ini masih dilakukan di Universitas XYZ pada mata kuliah PTI (Pengantar Teknologi Informasi). Universitas swasta yang berada di wilayah pusat dan selatan kota Jakarta ini memiliki jumlah mahasiswa yang relatif banyak serta mempunyai 5 fakultas, termasuk bidang ilmu terapan. Universitas yang berfokus pada layanan perkuliahan berbiaya murah ini memberikan jaminan akan pendidikan yang berkualitas. Hal tersebut dibuktikan dengan peningkatan kualitas yang berkesinambungan baik terhadap peserta didik maupun tenaga pengajarnya, seperti penugasan terhadap jenjang pendidikan doktoral bagi para dosennya. Kondisi ini diharapkan dapat mendukung visi pendidikan yang berkontribusi terhadap negara dengan basis biaya pendidikan yang terjangkau.

Mahasiswa yang mengikuti mata kuliah Pengenalan Teknologi Informasi (PTI) dengan beberapa kelas yang diambil sebagai *sampelnya*. Dalam data homogen tersebut diambil sejumlah 175 mahasiswa yang akan diamati parameter penilaiannya yang kemudian dianggap sebagai data

*training*. Selanjutnya data tersebut akan diproses menggunakan *application*, yaitu model yang diperoleh dari data training yang digunakan untuk menentukan hasil dari suatu set data *test* baru. Data tersebut dapat dinilai dengan 2 kategori, yaitu deskriptif yang bertugas mendeskripsikan atau mengklasifikasikan sifat-sifat data tertentu ke dalam suatu set data tujuan, dan prediktif yang bertugas melakukan induksi atau rangsangan terhadap suatu set data untuk dapat melakukan proses *prediksi* atau ramalan atau perkiraan. Kemudian data diolah dengan *Cross Industry Standard Process for Data Mining*, yang merupakan kerangka kerja analisis data yang terdiri dari enam tahap: *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* untuk menghasilkan solusi berbasis data yang efektif dan terstruktur (Khumaidi, 2020).



Gambar 1. Model CRISP-DM  
Sumber gambar: Khumaidi, 2020

SVM (*Support Vector Machine*) menggungguli beberapa jenis klasifikasi lain pada penelitian sebelumnya (Shedriko & Firdaus, 2022). SVM memperoleh nilai akurasi tertinggi. Pada penelitian ini digunakan beberapa metode lain yang juga merupakan klasifikasi prediksi dari nilai kelulusan. 3 metode lainnya tersebut adalah *Decision Tree* (DT), *Random Forest* (RF) dan *Gradient Boosting* (GB). *Research gap* pada *Support Vector Machine* adalah bahwa metode ini masih menghadapi keterbatasan dalam hal skalabilitas, interpretabilitas, serta sensitivitas terhadap pemilihan kernel dan parameter. Jabardi (2025) membahas tentang teori SVM, algoritma utamanya, dan bagaimana ia digunakan di dunia nyata, seperti bioinformatika, perbankan, dan pemrosesan citra, di antara bidang-bidang lainnya (Jabardi, 2025). Sementara *research gap* terkait *Decision Tree* adalah perbandingan kinerja klasifikasi ini dengan pendekatan klasifikasi lain menggunakan metrik evaluasi non-akurasi masih terbatas. Mienye dkk (2024) dalam penelitiannya menyajikan tinjauan komprehensif tentang pohon keputusan, termasuk konsep inti, algoritma, aplikasi, dari pengembangan awal mereka hingga algoritma ensemble berkinerja tinggi baru-baru ini dan representasi matematis dan algoritmiknya, yang kurang dalam literatur dan akan bermanfaat bagi peneliti ML (*Machine Learning*) dan ahli industri (Mienye & Jere, 2024). Metode *Random Forest* masih menghadapi keterbatasan dalam hal interpretabilitas, efisiensi komputasi, serta bias pada data tidak seimbang. Selain itu, sebagian besar penelitian belum mengeksplorasi stabilitas *feature importance* dan optimasi hyperparameter secara kontekstual. Rothacher dkk (2024) menggunakan studi simulasi, metode pemilihan variabel diuji mengenai sifat statistik mereka, dan perbandingan antara kinerja mereka dan kinerja model linier konvensional dibuat. Keuntungan dan kerugian dari metode yang diperiksa dibahas, dan rekomendasi praktis untuk penggunaan hutan acak untuk pemilihan variabel diberikan (Rothacher & Strobl, 2024). Sedangkan *Gradient Boosting* adalah meskipun klasifikasi ini terbukti memiliki

kinerja prediksi yang tinggi, metode ini masih menghadapi sejumlah keterbatasan, seperti sensitivitas terhadap hyperparameter, risiko overfitting, dan kebutuhan komputasi yang tinggi. Selain itu, penelitian yang mengkaji interpretabilitas model serta kinerja Gradient Boosting pada data tidak seimbang masih terbatas. Bentejac dkk (2020) mengusulkan analisis praktis tentang bagaimana varian baru dari gradient boosting ini bekerja dalam hal kecepatan pelatihan, kinerja generalisasi, dan pengaturan hyper-parameter, menggunakan CatBoost dan XGBoost (Bentéjac et al., 2021). Tujuan penelitian ini adalah mendapatkan metoda klasifikasi terbaik dari 4 jenis klasifikasi tersebut.

## II. STUDI LITERATUR

Beberapa studi dari penelitian-penelitian terdahulu, termasuk penelitian yang dilakukan oleh penulis sendiri, dapat dilihat sebagai berikut:

Penelitian di Universitas XYZ menggunakan metode CRISP-DM untuk membandingkan algoritma klasifikasi *Naive Bayes*, *Neural Network*, *Logistic Regression*, dan *Support Vector Machine* dalam memprediksi kelulusan mahasiswa mata kuliah PTI. Hasilnya, SVM terbukti paling efektif, efisien, dan robust dalam menentukan prediksi kelulusan mahasiswa (Shedriko & Firdaus, 2022). Sedangkan penelitian selama enam tahun di industri manufaktur oleh Bokrantz mengembangkan metode CRISP-DM dengan menambahkan fase *Operation and Maintenance* serta kerangka berbasis tugas untuk mengaitkan keterampilan. Hasilnya menyoroti tantangan biaya tersembunyi, pengelolaan *AI drift*, dan pentingnya kompetensi domain, sains data, serta rekayasa data dalam seluruh siklus proyek AI (*Artificial Intelligence*) (Bokrantz et al., 2024).

Klasifikasi *decision tree* merupakan metode populer dalam representasi data. Berbagai penelitian di bidang medis, teks, citra, dan lainnya telah memanfaatkannya. Artikel yang ditulis Jijo membahas secara rinci algoritma, dataset, dan hasil yang digunakan, serta mengevaluasi pendekatan untuk menentukan model klasifikasi paling akurat dari berbagai jenis data, dengan akurasi hingga 99,93% bila *decision tree* menggunakan repositori *machine learning* sebagai datasetnya (Jijo & Abdulazeez, 2021). Klasifikasi *decision tree* juga dianggap metode populer untuk pengolahan data. Penelitian di berbagai bidang—medis, teks, citra, dan smartphone—menggunakannya. Artikel Sivananda membahas algoritma, dataset, dan hasil secara rinci, mengevaluasi berbagai pendekatan, serta menganalisis keakuratan klasifikasi dan pemanfaatan berbagai jenis dataset (Sivananda et al., 2024).

Prediksi tren saham sulit dilakukan hanya dengan analisis empiris. Penelitian oleh Yin menggunakan data historis empat perusahaan AS dan mengoptimalkan model *random forest* dengan metode D-RF-RS dan *exponential smoothing*. Hasilnya, akurasi meningkat 0,17, lebih tinggi dari *light gradient boosting*, menunjukkan keunggulan *random forest* untuk prediksi jangka menengah dan panjang (Yin et al., 2023). Penelitian ini mengembangkan model *random forest* untuk estimasi biomassa di atas tanah (*AGB*) hutan Pinus Yunnan, China, menggunakan data Sentinel-2 dan algoritma Boruta. Model QRFb menunjukkan akurasi tertinggi ( $R^2$  0,88, RMSE 29,56 Mg/ha), mengurangi kesalahan estimasi, dan memberikan wawasan baru untuk penelitian biomassa dan siklus karbon (Zhang et al., 2024).

Algoritma *gradient boosting* modern seperti XGBoost, LightGBM, dan CatBoost menawarkan kecepatan dan akurasi tinggi. Studi yang dilakukan Bentejac ini membandingkan kinerja, kecepatan pelatihan, dan *hyperparameter*. Hasil menunjukkan CatBoost unggul dalam akurasi dan AUC, LightGBM tercepat, dan XGBoost menempati posisi kedua dalam kecepatan dan akurasi (Bentéjac et al., 2021). Implementasi *tree-ensemble* penting untuk prediksi, termasuk harga rumah. Studi yang dilakukan oleh Sibindi ini mengembangkan model hibrida LGBM-XGBoost untuk mengurangi *overfitting* dan meningkatkan akurasi dengan optimasi *hyperparameter* Bayesian. Hasilnya, model hibrida lebih unggul dibanding LGBM, XGBoost, Adaboost, dan GBM, dengan MSE 0,193, MAE 0,285, dan MAPE 0,156 (Sibindi et al., 2023).

Perkembangan teknologi dan data menuntut pemrosesan cepat dan andal. Penelitian oleh Abdullah ini membahas penggunaan algoritma *Support Vector Machine (SVM)* di berbagai bidang seperti pengenalan wajah, diagnosis penyakit, analisis teks, dan keamanan jaringan. Hasilnya menunjukkan SVM mencapai akurasi tinggi, terutama pada pengenalan wajah dan identifikasi penyakit (Abdullah & Abdulazeez, 2021). *Support Vector Machine (SVM)* merupakan

metode klasifikasi terawasi yang efisien. Penelitian oleh Valero-Carreras ini mengusulkan model SVM baru dengan *feature selection* dan membandingkannya dengan model *soft margin* klasik menggunakan metrik seperti AUC, Kappa, dan F-Score. Keduanya diselesaikan menggunakan pendekatan metaheuristik karena kompleksitas modelnya (Valero-Carreras et al., 2023).

### III. METODE

Penelitian dilakukan dengan menggunakan data dari mahasiswa yang mengikuti mata kuliah PTI (Pengantar Teknologi Informasi) yang berjumlah 175 orang, yang diperoleh melalui pengamatan parameter nilai input dan output, dengan menggunakan *digital-based method* (Tachaiphum, 2022). Metode ini bergantung pada alat-alat teknologi yang digunakan, dalam hal ini adalah *Orange Data Mining*. Dengan merangkai berbagai jenis model yang berkaitan dengan bermacam-macam klasifikasi dari *tool* tersebut yang sesuai kebutuhan, diperoleh keseluruhan nilai yang digunakan pada penelitian ini. Adapun evaluasi pada penelitian ini dilakukan dengan membandingkan 4 jenis klasifikasi berbeda, yaitu *Decison Tree*, *Random Forest*, *Gradient Boosting* dan *Support Vector Machine*. Pengujian dilakukan dengan mengimplementasikan ke 4 klasifikasi tersebut pada CRISP-DM.

### IV. HASIL DAN PEMBAHASAN

Nilai yang digunakan bagi kelulusan mahasiswa dalam suatu mata kuliah pada umumnya adalah kehadiran, tugas, ujian tengah semester dan ujian akhir semester, disamping nilai-nilai lainnya seperti keaktifan di kelas dan kuis. Implementasi CRISP-DM yang dilakukan adalah sebagai berikut:

#### 1. Pemahaman Bisnis

Fase pemahaman bisnis dalam memprediksi kelulusan mahasiswa mencakup identifikasi tujuan utama, seperti meningkatkan tingkat kelulusan dan efisiensi akademik. Peneliti memahami kebutuhan institusi, faktor-faktor yang memengaruhi kelulusan seperti nilai, kehadiran, dan aktivitas belajar, serta menentukan pertanyaan analitis yang relevan agar model data mampu mendukung pengambilan keputusan pendidikan. Hal tersebut dapat dilakukan dengan menggunakan model yang tepat beserta *tools* yang tepat pula.

#### 2. Pemahaman Data.

Fase ini berfokus pada pengumpulan, eksplorasi, dan evaluasi kualitas data yang tersedia. Data dapat mencakup nilai akademik, kehadiran, usia, latar belakang sosial, serta aktivitas perkuliahan. Tahap ini bertujuan mengenali karakteristik, pola awal, dan potensi masalah seperti data hilang atau tidak konsisten, sehingga diperoleh pemahaman mendalam untuk mendukung tahap pemodelan secara akurat. Adapun kriteria data yang digunakan adalah sebagai berikut: nilai akhir  $\geq 56$ , absensi  $\leq 80\%$ , nilai tugas  $\geq 60$ , harus mengikuti setiap kuis yang diberikan, serta aktif di kelas. Nilai subyektifitas tetap berperan dengan mempertahankan obyektifitas, seperti bila ada nilai yang hilang atau nilai ujian akhir mahasiswa yang buruk sedangkan si mahasiswa aktif di kelas dan menunjukkan nilai yang baik saat di kelas, akan diberikan pertimbangan secara subyektif oleh di pengajar.

#### 3. Persiapan Data

Fase ini mencakup proses pembersihan, transformasi, dan integrasi data agar siap dianalisis. Data mahasiswa seperti nilai, kehadiran, aktivitas akademik, dan latar belakang diseleksi serta dikonversi ke format yang konsisten. Data yang tidak relevan, duplikat, atau hilang diperbaiki agar model prediksi kelulusan dapat dibangun secara akurat dan menghasilkan wawasan yang valid. Beberapa sampel data lulus dapat dilihat pada tabel berikut.

Tabel 1. Beberapa Data Training dari Mahasiswa yang Lulus

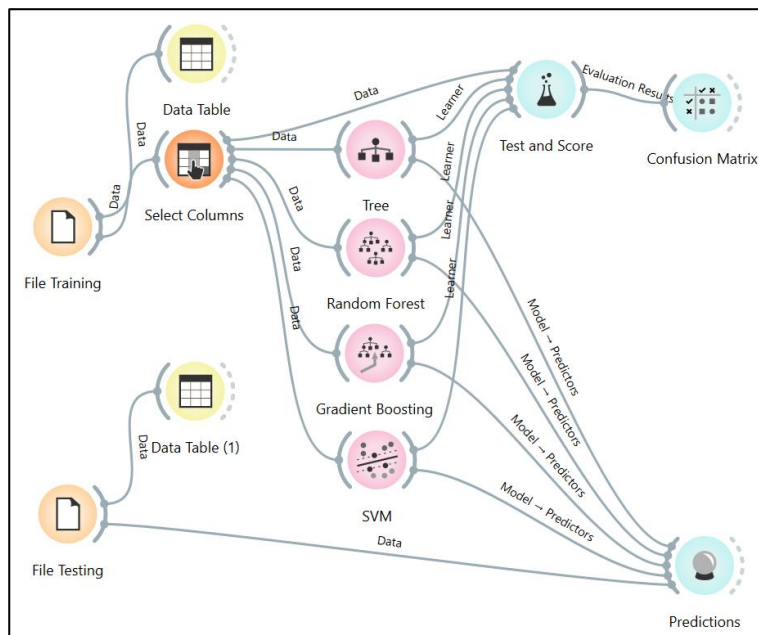
1	Nilai Tugas	Nilai UTS	Nilai UAS	Kelulusan	Target
157	10,0	7,2	7,1	LULUS	1
158	10,0	6,2	4,7	LULUS	1
159	9,0	6,8	4,2	LULUS	1
160	10,0	8,0	6,2	LULUS	1
161	9,0	5,0	6,3	LULUS	1
162	9,0	8,5	7,2	LULUS	1
163	9,0	5,0	6,1	LULUS	1
164	10,0	7,2	6,3	LULUS	1
165	9,0	7,2	6,2	LULUS	1
166	10,0	7,2	7,5	LULUS	1
167	10,0	6,8	4,9	LULUS	1
168	10,0	7,2	6,3	LULUS	1
169	9,0	6,2	5,8	LULUS	1
170	10,0	5,0	4,2	LULUS	1
171	8,0	5,5	5,0	LULUS	1
172	8,0	7,2	6,3	LULUS	1
173	10,0	4,0	5,8	LULUS	1
174	8,0	7,2	5,9	LULUS	1
175	9,0	6,8	5,9	LULUS	1
176	10,0	6,2	4,8	LULUS	1

Sumber tabel: Observasi Penulis

Adapun atribut yang digunakan sebagai *evidence* dalam algoritma yang diperbandingkan pada penelitian ini terdiri dari 3 *item*, yaitu: nilai tugas, nilai ujian tengah semester (UTS), dan nilai akhir semester (UAS)

4. Permodelan

Fase pemodelan ini melibatkan pemilihan dan penerapan algoritma yang sesuai untuk kemudian membandingkannya, yaitu *Decision Tree*, *Random Forest*, *Gradient Boosting* dan *Support Vector Machine*, dalam memprediksi kelulusan mahasiswa. Data yang telah dipersiapkan dibagi menjadi data latih dan uji. Parameter model disesuaikan guna memperoleh performa optimal, sehingga hasil prediksi dapat membantu mengidentifikasi faktor-faktor yang memengaruhi peluang kelulusan mahasiswa.



Gambar 2. Diagram Model dengan Orange Data Mining  
 Sumber Gambar: Penulis

## 5. Evaluasi

Fase evaluasi ini menilai sejauh mana model prediksi bekerja dengan baik. Performa model diuji menggunakan metrik seperti akurasi, error, presisi, recall, dan AUC. Hasil evaluasi dari ke 4 model diperbandingkan, untuk mendapatkan klasifikasi terbaik yang terkait dengan tujuan bisnis, yaitu meningkatkan tingkat kelulusan. Hasil evaluasi tersebut dapat dilihat pada tabel 2,3 dan 4 berikut di bawah.

Tabel 2. Perbandingan akurasi dan error

Classification Algorithms	Accuracy Rate	Error Rate
Decision Tree (DT)	93.1%	6.9%
Random Forest (RF)	92.6%	7.4%
Gradient Boosting (GB)	94.9%	5.1%
Support Vector Machine (SVM)	95.4%	4.6%

Sumber tabel: Penulis

Tabel 2 menunjukkan nilai tingkat akurasi dan error dari hasil perhitungan terhadap ke 4 klasifikasi. SVM dengan tingkat akurasi sebesar 95,4% masih menunjukkan superioritasnya terhadap ke 3 klasifikasi lainnya. Meskipun demikian, ke 3 klasifikasi lainnya tersebut masih lebih baik dari jenis-jenis klasifikasi pada penelitian penulis sebelumnya dengan menunjukkan nilai rata-rata di atas 90% untuk tingkat akurasinya. Sedangkan untuk tingkat error, SVM memberikan hasil yang paling kecil di antara ke 3 jenis klasifikasi lainnya.

Tabel 3. Perbandingan analisis statistik

Algoritma	Class	Precision	Recall	AUC
DT	Passed	0.950	0.964	0.890
	Failed	0.853	0.806	0.890
RF	Passed	0.950	0.957	0.965
	Failed	0.829	0.806	0.965
GB	Passed	0.971	0.964	0.967
	Failed	0.865	0.889	0.967
SVM	Passed	0.958	0.986	0.969
	Failed	0.938	0.833	0.969

Sumber tabel: Penulis

Sedangkan pada tabel 3, nampak bahwa GB memberikan nilai precision yang paling tinggi di antara ke 4 jenis klasifikasi. Hal ini menandakan bahwa nilai kelulusan positif (*true positive*) dari GB adalah yang paling tinggi dibanding yang lainnya. Namun demikian, nilai *recall* dan AUC dari hasil perhitungan menunjukkan bahwa SVM juga lah yang paling tinggi. Hal tersebut menunjukkan klasifikasi SVM menyerap nilai positif kelulusan yang lebih baik dari yang lainnya atau sedikit bahkan hampir tidak ada kasus positif yang terlewat. Hal tersebut juga menunjukkan bahwa SVM sangat baik membedakan antara kasus negatif dengan positif terkait kelulusan.

## 6. Penerapan

Fase ini merupakan tahap penerapan hasil model data mining ke lingkungan nyata agar dapat dimanfaatkan oleh pengguna atau organisasi. Model yang telah dievaluasi akan diimplementasikan, baik secara manual maupun otomatis. Pada fase ini juga dilakukan pembuatan laporan, pemantauan kinerja model, serta perencanaan pemeliharaan dan pembaruan agar hasil analisis tetap relevan dan bermanfaat. Pada penelitian ini, fase *deployment* dapat berupa pengujian dari ke 4 klasifikasi yang dapat dilihat pada tabel 4.

Tabel 4. Hasil Pengujian

	Tree	Random Forest	Gradient Boosting	SVM	Homework	MidTest	FinalTest
1	Passed	Failed	Failed	Failed	42	42	42
2	Passed	Failed	Failed	Failed	43	43	43
3	Passed	Failed	Failed	Failed	45	45	45
4	Passed	Failed	Failed	Failed	46	46	46
5	Passed	Failed	Failed	Failed	58	58	58
6	Passed	Passed	Passed	Passed	59	59	59

Sumber tabel: Penulis

Hasil pengujian pada table 4 menunjukkan bahwa RF, GB dan SVM memperoleh kelulusan pada ambang batas nilai 59, di mana mahasiswa dengan nilai  $\geq 59$  dinyatakan lulus. Sedangkan DT memperoleh nilai kelulusan yang sangat tolerir, bahkan dengan nilai 42 pun mahasiswa masih bisa dinyatakan lulus. Hal tersebut disebabkan karena DT kurang dapat membedakan antara kasus negatif dan positif kelulusan dengan baik (AUC), juga penyerapan kasus kelulusan positif (*recall*) yang paling rendah di antara ke 4 jenis klasifikasi.

## V. KESIMPULAN

Dari fase 1 hingga fase 6 implementasi CRISP-DM pada penelitian ini memberikan beberapa Kesimpulan. SVM masih mengungguli ke 3 jenis klasifikasi lainnya. GB memperoleh nilai tertinggi terkait nilai precision dari hasil perhitungan. DT memperoleh nilai terkecil dari kesemua aspek hasil perhitungan, kecuali tingkat akurasi, tingkat error dan *recall*. DT juga memberikan hasil perhitungan kelulusan yang terlalu tolerir dari semua klasifikasi, sehingga dapat dianggap kurang akurat dalam hasil klasifikasi yang diberikan. Dengan demikian, dari penelitian yang sebelumnya serta penelitian ini, penulis masih memperoleh nilai SVM yang terbaik untuk semua parameter perhitungan, kecuali nilai *precision* yang masih didominasi oleh GB.

## VI. REFERENSI

- Abdullah, D. M., & Abdulazeez, A. M. (2021). Machine Learning Applications based on SVM Classification: A Review. *Qubahan Academic Journal*. <https://doi.org/10.48161/qaj.v1n2a50>
- Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. In *Artificial Intelligence Review* (Vol. 54, Issue 3). Springer Netherlands. <https://doi.org/10.1007/s10462-020-09896-5>
- Bokrantz, J., Subramaniyan, M., & Skoogh, A. (2024). Realising the promises of artificial intelligence in manufacturing by enhancing CRISP-DM. *Production Planning and Control*, 35(16), 2234–2254. <https://doi.org/10.1080/09537287.2023.2234882>
- Jabardi, M. (2025). Support Vector and Machines : Theory, Algorithms, and Applications. *Infocommunications Journal*, XVII(1). <https://doi.org/10.36244/ICJ.2025.1.8>
- Jijo, B. T., & Abdulazeez, A. M. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(1), 20–28. <https://doi.org/10.38094/jastt20165>
- Khumaidi, A. (2020). Data Mining for Predicting the Amount of Coffee Production Using Crisp-Dm Method. *Jurnal Techno Nusa Mandiri*, 17(1), 1–8. <https://doi.org/10.33480/techno.v17i1.1240>
- Mienye, I. D., & Jere, N. (2024). A Survey of Decision Trees : Concepts , Algorithms , and Applications. *IEEE Access*, 12(June), 86716–86727. <https://doi.org/10.1109/ACCESS.2024.3416838>

- 
- Rothacher, Y., & Strobl, C. (2024). *Identifying Informative Predictor Variables With Random Forests* (Issue October 2023). <https://doi.org/10.3102/10769986231193327>
- Shedriko, & Firdaus, M. (2022). Penentuan Klasifikasi Dengan Crisp-Dm. *The Indonesian Journal of Computer Science*, 10(11), 826–831.
- Sibindi, R., Mwangi, R. W., & Waititu, A. G. (2023). A boosting ensemble learning based hybrid light gradient boosting machine and extreme gradient boosting model for predicting house prices. *Engineering Reports*, 5(4), 1–19. <https://doi.org/10.1002/eng2.12599>
- Sivananda, M. ., & Kumar, D. G. K. (2024). Classification and Regression Based on Decision Tree Algorithm for Machine Learning. *Interantional Journal of Scientific Research in Engineering and Management*, 08(02), 1–13. <https://doi.org/10.55041/ijrem28533>
- Tachaiphum, N. (2022). *The Global Pandemic and A Shift in Research Methods: Applying Digital-based Methods to Educational Research in Applied Linguistics*. <https://doi.org/10.15663/wje.v27i2>
- Valero-Carreras, D., Alcaraz, J., & Landete, M. (2023). Comparing two SVM models through different metrics based on the confusion matrix. *Computers and Operations Research*, 152(April 2022), 106131. <https://doi.org/10.1016/j.cor.2022.106131>
- Yin, L., Li, B., Li, P., & Zhang, R. (2023). Research on stock trend prediction method based on optimized random forest. *CAAI Transactions on Intelligence Technology*, 8(1), 274–284. <https://doi.org/10.1049/cit2.12067>
- Zhang, X., Shen, H., Huang, T., Wu, Y., Guo, B., Liu, Z., Luo, H., Tang, J., Zhou, H., Wang, L., Xu, W., & Ou, G. (2024). Improved random forest algorithms for increasing the accuracy of forest aboveground biomass estimation using Sentinel-2 imagery. *Ecological Indicators*, 159(February), 111752. <https://doi.org/10.1016/j.ecolind.2024.111752>