

## BAB 2

### TINJAUAN PUSTAKA

#### 2.1. Tinjauan Pustaka

Beberapa analisis tentang prediksi *dropout* telah dilakukan menggunakan berbagai pendekatan *Machine Learning* (ML). Rofi et al. (2024) melakukan penelitian dengan judul "Perbandingan Metode K-NN dan *Random Forest* pada Klasifikasi Mahasiswa Berpotensi *Dropout*" melibatkan penerapan pendekatan *machine learning* dengan dua metode klasifikasi, yaitu *K-Nearest Neighbor* (K-NN) dan *Random Forest*. Tujuan dari penelitian ini adalah untuk mengklasifikasikan kasus *dropout* di lingkungan perguruan tinggi, sebuah isu yang memiliki dampak signifikan terhadap kualitas pendidikan dan akreditasi institusi. Hasil dari penelitian tersebut menunjukkan bahwa algoritma *Random Forest* memberikan tingkat akurasi yang lebih tinggi, mencapai 99.05%, dibandingkan dengan K-NN yang mencapai 98.10%. Lebih lanjut, atribut Persentase Aktif diidentifikasi sebagai faktor yang paling berpengaruh dalam mengidentifikasi mahasiswa berpotensi *dropout* menurut algoritma *Random Forest*. Temuan ini menambah pemahaman tentang efektivitas metode klasifikasi dalam meramalkan mahasiswa yang berpotensi mengalami *dropout*, sementara juga menyoroti pentingnya keterlibatan aktif mahasiswa sebagai salah satu upaya pencegahan yang mungkin. Dengan demikian, penelitian ini memberikan sumbangan signifikan dalam pemahaman mengenai strategi pencegahan terkait masalah *dropout* di konteks pendidikan tinggi [8].

Demikian pula Tamada et al. (2023) dengan judul penelitian "*Predicting Students at Risk of Dropout in Technical Course Using LMS Logs*" Penelitian ini mengadopsi pendekatan teknik *machine learning* untuk memprediksi potensi *dropout* mahasiswa dalam kursus teknis dengan memanfaatkan data log dari *Learning Management System* (LMS). Metode penelitian ini dimulai dengan melakukan analisis yang komprehensif terhadap data log aktivitas mahasiswa di platform online, dengan mempertimbangkan berbagai aspek seperti frekuensi login, durasi sesi, partisipasi dalam diskusi, dan keterlibatan dalam tugas-tugas daring. Dengan menggunakan algoritma *Random Forest*, penelitian ini berhasil mengidentifikasi pola-pola perilaku yang berkorelasi dengan kemungkinan mahasiswa untuk *dropout* dari kursus tersebut. Hasilnya, algoritma prediksi *dropout* yang dikembangkan dalam penelitian ini menunjukkan tingkat akurasi

yang tinggi, yang memungkinkan lembaga pendidikan untuk mengidentifikasi mahasiswa yang berisiko *dropout* dan memberikan intervensi yang tepat guna. Implikasi dari temuan ini adalah adanya kesempatan bagi lembaga pendidikan untuk mengembangkan strategi intervensi yang tepat guna dalam meningkatkan retensi mahasiswa di kursus teknis, seperti penyediaan dukungan tambahan atau penyesuaian kurikulum. Dengan demikian, penelitian ini memberikan kontribusi yang signifikan dalam upaya untuk mengurangi angka *dropout* dalam pendidikan tinggi [7].

Andrade-Girón et al. (2023) dengan judul penelitian "*Predicting Student Dropout based on Machine Learning and Deep Learning: A Systematic Review*" adalah sebuah penelitian yang meninjau keberhasilan algoritma *Machine Learning* (ML) dan *Deep Learning* (DL) dalam memprediksi *dropout* mahasiswa. Studi ini mencakup peninjauan sistematis terhadap literatur dari basis data elektronik seperti Scopus, IEEE, dan Web of Science hingga Juni 2023, dengan 246 artikel yang diidentifikasi, dan 23 studi yang termasuk dalam ulasan akhir. Penelitian ini mengevaluasi berbagai metrik kinerja seperti akurasi/presisi, sensitivitas/recall, spesifisitas, dan *area under the curve* (AUC). Selain itu, aspek terkait seperti jenis studi, strategi pelatihan, strategi pengujian, *cross-validation*, dan matriks kebingungan juga dipertimbangkan. Hasil penelitian mengungkapkan bahwa algoritma ML yang paling banyak digunakan adalah *Random Forest*, yang hadir dalam 21.73% dari studi yang direview dan mencapai akurasi 99% dalam prediksi *dropout* mahasiswa, lebih tinggi dibandingkan semua algoritma yang digunakan dalam studi yang diulas. Studi ini juga menemukan bahwa algoritma seperti *neural networks*, *decision trees*, *logistic regression*, dan *support vector machines* juga sering digunakan, dengan variasi kinerja yang beragam. Distribusi artikel yang dipilih berdasarkan negara menunjukkan bahwa sebagian besar artikel (21%) diterbitkan di China, diikuti oleh Amerika Serikat (17.39%), serta Korea, India, dan Spanyol (masing-masing 8.69%). Hal ini menunjukkan bahwa *dropout* mahasiswa adalah topik penelitian yang relevan di berbagai belahan dunia. Penelitian ini mengidentifikasi beberapa pola dalam pembagian data untuk pelatihan dan pengujian: 30.43% studi menggunakan 70% sampel untuk pelatihan dan 30% untuk pengujian, sementara 26.08% menggunakan pembagian yang serupa. Selain itu, 17.39% studi menggunakan 80% sampel untuk pelatihan dan 20% untuk pengujian. Secara keseluruhan, algoritma yang sering digunakan dalam studi ini meliputi *neural networks* (39.13%), *decision trees* (56.52%), *logistic regression* (39.13%), *support vector machines* (30.43%), dan *Random Forest*

(47.82%). Algoritma *Random Forest* terbukti memiliki performa terbaik dengan akurasi mencapai 99%. Penelitian ini memberikan gambaran umum tentang algoritma ML dan DL yang menjanjikan dalam memprediksi *dropout* mahasiswa. Algoritma *Random Forest* muncul sebagai algoritma dengan kinerja terbaik dalam studi yang diulas, menunjukkan akurasi yang mengesankan sebesar 99%. Temuan ini mendukung literatur yang ada, yang menunjukkan bahwa *Random Forest* secara konsisten menunjukkan performa superior dalam prediksi *dropout* [9].

Penelitian yang dilakukan oleh Mu'tashim et al. (2023) berjudul "Klasifikasi Ketepatan Lama Studi Mahasiswa dengan Algoritma *Random Forest* dan *Gradient Boosting* (Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta)" merupakan sebuah upaya untuk mengaplikasikan pendekatan data mining dengan menggunakan algoritma *Random Forest* dan *Gradient Boosting*. Penelitian ini memiliki fokus yang jelas, yaitu menganalisis ketepatan lama studi mahasiswa S1 di Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta dengan menggunakan data dari angkatan 2015 hingga 2017. Dengan menggunakan teknik *10 k-fold cross-validation* untuk *Random Forest* dan *3 k-fold cross-validation* untuk *Gradient Boosting*, hasil dari uji coba sampel menunjukkan bahwa algoritma *Random Forest* mencapai tingkat akurasi sebesar 82,64%, sedangkan *Gradient Boosting* mencapai tingkat akurasi sebesar 79,66%. Temuan ini memberikan gambaran yang jelas tentang performa relatif dari kedua metode klasifikasi dalam mengidentifikasi ketepatan lama studi mahasiswa. Selain itu, hasil ini juga dapat memberikan wawasan tambahan bagi pengambil keputusan di Fakultas Ilmu Komputer UPNVJ dalam mengembangkan strategi untuk meningkatkan tingkat kelulusan dan efisiensi akademik mahasiswa. Dengan demikian, penelitian ini bukan hanya memberikan kontribusi pada pengetahuan akademik di bidang data mining dan pendidikan tinggi, tetapi juga memiliki implikasi praktis yang signifikan bagi institusi terkait [10].

Cho et al. (2023) dalam penelitian mereka berjudul "*A Study on Dropout Prediction for University Students Using Machine Learning*" menggunakan berbagai algoritma pembelajaran mesin untuk memprediksi *dropout* mahasiswa di Sahmyook University. Mereka menganalisis data akademik dari 20,050 mahasiswa dan menerapkan algoritma seperti Regresi Logistik, Pohon Keputusan, *Random Forest*, *Support Vector Machine*, *Deep Neural Network*, dan *LightGBM*. Dari hasil eksperimen, algoritma

*LightGBM* menunjukkan performa terbaik dengan F1-score sebesar 0.840, mengungguli algoritma lainnya. Selain itu, mereka meneliti pengaruh teknik *oversampling* seperti SMOTE, ADASYN, dan Borderline-SMOTE untuk mengatasi ketidakseimbangan data, yang merupakan masalah umum dalam prediksi *dropout*. Menariknya, mereka menemukan bahwa penerapan teknik *oversampling* secara umum menurunkan performa algoritma, dengan pengecualian algoritma *Random Forest* yang menunjukkan sedikit peningkatan. Temuan ini menunjukkan bahwa sementara *oversampling* dapat membantu dalam beberapa kasus, itu juga dapat menambah noise pada data, sehingga mempengaruhi akurasi prediksi. Penelitian ini memberikan wawasan berharga tentang efektivitas berbagai algoritma pembelajaran mesin dan teknik *oversampling* dalam konteks prediksi *dropout* mahasiswa [11].

Penelitian tentang prediksi *dropout* mahasiswa juga dilakukan oleh Urbina-Nájera et al. (2022) dengan berjudul "*Predictive Model for Taking Decision to Prevent University Dropout*", para peneliti menggunakan 51,497 instance dengan 26 atribut dari tiga fakultas di UPAEP-Universidad, Meksiko. Mereka menerapkan algoritma jaringan saraf tiruan dan pohon keputusan untuk memprediksi *dropout*, serta menggunakan metode pemilihan atribut dan resampling untuk menyeimbangkan kelas utama. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* memiliki kinerja terbaik dengan *Matthew correlation coefficient* sebesar 87.43% dan akurasi 94.34%, dibandingkan dengan jaringan saraf tiruan yang mencapai *Matthew correlation coefficient* sebesar 53.39%. Penelitian ini memberikan kontribusi yang signifikan dalam upaya pencegahan *dropout* di pendidikan tinggi. Dengan menggunakan algoritma prediktif yang dikembangkan, para peneliti dapat memperkirakan jumlah *dropout* yang mungkin terjadi dalam setiap periode, sehingga membantu pihak terkait dalam mengambil tindakan pencegahan yang tepat. Temuan penelitian ini tidak hanya memberikan wawasan tentang fenomena *dropout* di lingkungan pendidikan tinggi, tetapi juga menawarkan solusi praktis dalam memitigasi dampak negatifnya. Dengan demikian, penelitian ini memiliki potensi untuk meningkatkan retensi mahasiswa dan memperkuat kelangsungan program-program pendidikan tinggi melalui penyediaan dukungan yang lebih efektif bagi mereka yang berisiko mengalami *dropout* [12].

Penelitian terbaru oleh Kabathova dan Drlik (2021) mengeksplorasi prediksi *dropout* mahasiswa menggunakan berbagai teknik pembelajaran mesin. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* mencapai tingkat akurasi tertinggi

sebesar 93%, dengan recall mencapai 96% dan presisi sebesar 86%. Algoritma Regresi Logistik juga mencapai tingkat akurasi yang tinggi sebesar 93%, dengan recall sebesar 98% dan presisi sebesar 79%. Di sisi lain, algoritma *Naïve Bayes* menunjukkan hasil yang kurang memuaskan dengan akurasi sebesar 77%, recall sebesar 72%, dan presisi sebesar 93%. Temuan ini menunjukkan bahwa pemilihan algoritma yang tepat dapat memengaruhi hasil prediksi secara signifikan dalam konteks prediksi *dropout* mahasiswa. Selain itu, penelitian ini juga menyoroti penanganan ketidakseimbangan data dalam prediksi *dropout* mahasiswa. Algoritma *Random Forest* dan Regresi Logistik menunjukkan kinerja terbaik dalam mengatasi masalah ini, dengan masing-masing mendapatkan F1-score sebesar 0.91 dan 0.90. Namun, algoritma *Naïve Bayes* dan Jaringan Saraf Tiruan cenderung memberikan hasil yang kurang memuaskan dalam hal ini. Hasil penelitian ini memberikan wawasan yang penting tentang keberhasilan berbagai algoritma dalam memprediksi *dropout* mahasiswa, serta menyoroti pentingnya evaluasi metrik kinerja yang cermat dalam memilih algoritma prediksi yang sesuai untuk digunakan dalam praktek nyata [13].

Lottering et al. (2020) dengan judul penelitian "*A Machine Learning Approach to Identifying Students at Risk of Dropout: A Case Study*" adalah sebuah penelitian yang menerapkan teknik-teknik klasifikasi untuk menentukan mahasiswa yang berisiko *dropout* dari program studi mereka. Penelitian ini menggunakan data dari 4419 mahasiswa yang terdaftar di Fakultas Informasi, Komunikasi, dan Teknologi di Tshwane University of Technology. Algoritma yang digunakan dalam penelitian ini meliputi *Random Forest*, *Support Vector Machines*, *Decision Trees*, *Naïve Bayes*, *K-Nearest Neighbor*, dan *Logistic Regression*. Hasil penelitian mengindikasikan bahwa *Random Forest* mencapai tingkat akurasi tertinggi sebesar 94.14%, dibandingkan dengan algoritma lain. *Support Vector Machines (SVM)* juga menunjukkan kinerja yang baik dengan tingkat akurasi 89.31% dan tingkat spesifisitas 91.25%. Algoritma *Decision Tree (DT)* memiliki tingkat presisi terbaik sebesar 88.46%, yang sangat penting untuk meminimalkan *false negatives*. *KNN* dan *Naïve Bayes* merupakan algoritma dengan kinerja terendah, meskipun *KNN* masih memiliki presisi yang lebih baik dibandingkan *Naïve Bayes*. Penelitian ini menekankan pentingnya penggunaan metode *ensemble* seperti *Random Forest* dalam prediksi *dropout* mahasiswa, karena metode ini dapat meningkatkan akurasi prediksi secara signifikan. Selain itu, penelitian ini juga

menunjukkan bahwa meskipun algoritma prediksi yang digunakan efektif, hasilnya tidak dapat digeneralisasikan ke institusi lain tanpa pengujian lebih lanjut [14].

Biswas et al. (2019) melakukan penelitian dengan judul "*Predicting Post-Graduation Dropout Using Machine Learning: A Case Study in Bangladesh*" penelitian tersebut mengaplikasikan berbagai algoritma klasifikasi untuk mengidentifikasi faktor-faktor yang menyebabkan mahasiswa pascasarjana di Bangladesh berisiko *dropout* atau gagal menyelesaikan studi mereka. Penelitian ini menggunakan data asli dari mahasiswa pascasarjana untuk membangun algoritma prediksi. Penelitian ini melibatkan tujuh algoritma klasifikasi, yaitu *Naïve Bayes*, *Multilayer Perceptron*, *Logistic Regression*, *Locally Weighted Learning (LWL)*, *Random Forest*, *Random Tree*, dan *Part*. Setiap algoritma dievaluasi menggunakan matriks kebingungan dan tujuh metrik kinerja: *akurasi*, *sensitivitas*, *presisi*, *spesifisitas*, *F1 score*, *False Positive Rate (FPR)*, dan *False Negative Rate (FNR)*. Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes*, *LWL*, dan *Part* memiliki tingkat akurasi tertinggi sebesar 86.36%, sementara *Random Tree* memiliki kinerja terendah dengan akurasi sebesar 74.24%. Analisis lebih lanjut berdasarkan *F1 score*, yang mempertimbangkan keseimbangan antara *presisi* dan *recall*, menunjukkan bahwa *LWL* memiliki kinerja terbaik dengan *F1 score* sebesar 92.56%, sedikit lebih tinggi dibandingkan *Naïve Bayes* (92.43%) dan *Part* (91.89%). Penelitian ini menyoroti pentingnya penggunaan beberapa metrik evaluasi kinerja, terutama ketika dataset tidak seimbang dan biaya kesalahan positif (*false positives*) dan negatif (*false negatives*) berbeda signifikan. Kinerja superior *LWL* dalam konteks ini menunjukkan bahwa metode ini sangat efektif untuk prediksi risiko *dropout*. Penelitian ini memberikan dasar yang kuat bagi institusi pendidikan untuk mengembangkan strategi intervensi yang proaktif dan berbasis data guna meningkatkan retensi mahasiswa pascasarjana [15].

Agrusti et al. (2019) dengan judul penelitian "*University Dropout Prediction through Educational Data Mining Techniques: A Systematic Review*" adalah sebuah penelitian yang bertujuan untuk mengidentifikasi studi-studi yang menggunakan teknik data mining pendidikan untuk memprediksi *dropout* mahasiswa di program studi tradisional. Studi ini mengacu pada strategi Europe 2020 yang menyoroti bahwa tingkat *dropout* di negara-negara Eropa merupakan masalah utama yang harus dihadapi dalam waktu dekat. Berdasarkan statistik Eurostat tahun 2017, rata-rata 10.6% anak muda (berusia 18-24 tahun) di EU-28 meninggalkan pendidikan dan pelatihan secara dini.

Penelitian ini mengidentifikasi 241 studi terkait dari katalog Scopus dan Web of Science (WoS), dari mana 73 studi dipilih berdasarkan kriteria inklusi dan eksklusi yang telah ditentukan. Dari studi-studi yang dipilih, diidentifikasi enam teknik klasifikasi data mining yang digunakan untuk memprediksi dropout mahasiswa, yaitu *Decision Tree*, *K-Nearest Neighbour*, *Support Vector Machines*, *Bayesian Classification*, *Neural Networks*, dan *Logistic Regression*, serta satu kategori untuk teknik-teknik minor yang disebut "Miscellanea". Penelitian ini menemukan bahwa teknik data mining yang paling sering digunakan adalah *Decision Tree* (67%), diikuti oleh *Bayesian Classification* (49%), *Neural Networks* (40%), dan *Logistic Regression* (34%). Selain itu, diidentifikasi 14 alat data mining yang digunakan dalam studi-studi tersebut, dengan WEKA, SPSS, dan R sebagai alat yang paling sering digunakan. Penelitian ini menyoroti bahwa prediksi *dropout* universitas merupakan minat yang tinggi bagi komunitas peneliti akademik dan bahwa teknik dengan tingkat presisi tinggi sedang dikembangkan untuk mengatasi masalah krusial ini. Namun, tidak ditemukan studi tentang *dropout* yang menggunakan *Convolutional Neural Network* (CNN), sebuah algoritma yang sangat efisien dan lebih sering digunakan dalam penelitian pengenalan gambar [16].

Tabel 2.1. Daftar Penelitian Terdahulu

No	Fokus	Metode	Hasil	Referensi
1	Membandingkan metode <i>K-Nearest Neighbor</i> (K-NN) dan <i>Random Forest</i> dalam mengklasifikasikan mahasiswa berpotensi drop out di perguruan tinggi	Menggunakan dua metode klasifikasi, yaitu <i>K-Nearest Neighbor</i> (K-NN) dan <i>Random Forest</i>	Algoritma <i>Random Forest</i> memiliki akurasi lebih tinggi (99.05%) dibandingkan dengan <i>K-Nearest Neighbor</i> (K-NN) (98.10%)	Rofi et al. (2024)
2	Memprediksi mahasiswa yang berisiko <i>dropout</i> dalam kursus teknis berdasarkan log LMS, dengan fokus pada kinerja <i>machine learning</i>	Pemanfaatan teknik <i>Machine Learning</i> untuk memprediksi dropout mahasiswa berdasarkan log LMS, dengan penemuan kelompok perilaku yang berkorelasi dengan kinerja.	<i>Random Forest</i> memberikan hasil terbaik dengan 84,47% pada skor F1, dengan penelitian menunjukkan keefektifan metode AI dan ML dalam mencegah <i>dropout</i> .	Tamada et al. (2023)
3	Identifikasi dan prediksi mahasiswa berisiko <i>dropout</i>	Tinjauan sistematis terhadap literatur menggunakan algoritma ML dan	<i>Random Forest</i> memiliki akurasi tertinggi 99%, sering digunakan dalam 21.73% studi yang diulas	Andrade-Girón et al. (2023)

No	Fokus	Metode	Hasil	Referensi
		DL: <i>Random Forest</i> , <i>neural networks</i> , <i>decision trees</i> , <i>logistic regression</i> , <i>support vector machines</i> , dll.		
4	Menganalisis ketepatan lama studi dan penerapan data mining	Penggunaan data mining dengan metode <i>Random Forest</i> dan <i>Gradient Boosting</i>	Algoritma <i>Random Forest</i> dengan 10 <i>k-fold cross-validation</i> memberikan akurasi sebesar 82,64%, Algoritma <i>Gradient Boosting</i> dengan 3 <i>k-fold cross-validation</i> memberikan akurasi sebesar 79,66%	Mu'tashim et al. (2023)
5	Prediksi <i>dropout</i> mahasiswa & Penanganan ketidakseimbangan data	Analisis data akademik menggunakan algoritma pembelajaran mesin: Regresi Logistik, Pohon Keputusan, <i>Random Forest</i> , <i>Support Vector Machine</i> , <i>Deep Neural Network</i> , dan <i>LightGBM</i> . Teknik oversampling: SMOTE, ADASYN, dan Borderline-SMOTE.	<i>LightGBM</i> menunjukkan performa terbaik dengan F1-score sebesar 0.840 dalam memprediksi <i>dropout</i> mahasiswa. Oversampling secara umum menurunkan performa algoritma, kecuali <i>Random Forest</i> yang menunjukkan sedikit peningkatan. Oversampling dapat menambah noise pada data, mempengaruhi akurasi prediksi.	Cho et al. (2023)
6	Prediksi Dropout Mahasiswa	Jaringan Saraf Tiruan dan Pohon Keputusan	Algoritma <i>Random Forest</i> memiliki kinerja terbaik dengan <i>Matthew correlation coefficient</i> sebesar 87.43% dan akurasi 94.34%.	Urbina-Nájera et al. (2022)
7	Prediksi <i>dropout</i> mahasiswa	Analisis data pendidikan menggunakan berbagai algoritma pembelajaran mesin: Regresi Logistik, Pohon Keputusan, <i>Random Forest</i> , <i>Naïve Bayes</i> , <i>Support Vector Machine</i> , dan Jaringan Saraf Tiruan.	<i>Random Forest</i> mencapai akurasi, presisi, dan skor F1 tertinggi masing-masing 93%, 86%, dan 91%. <i>Naïve Bayes</i> memiliki presisi 93% tetapi akurasi keseluruhan rendah 77% karena dipengaruhi oleh false negative. Regresi Logistik dan <i>Random Forest</i> menunjukkan performa terbaik dengan masing-masing nilai akurasi sebesar 93% dan 92%. Algoritma <i>Naïve Bayes</i> dan Jaringan Saraf Tiruan tidak direkomendasikan karena akurasi yang rendah	Kabathova & Drlik (2021)
8	Identifikasi mahasiswa berisiko <i>dropout</i>	Menggunakan algoritma klasifikasi: <i>Random</i>	<i>Random Forest</i> mencapai akurasi tertinggi 94.14%. SVM akurasi 89.31%, DT presisi	Lottering et al. (2020)

No	Fokus	Metode	Hasil	Referensi
		<i>Forest, Support Vector Machines, Decision Trees, Naïve Bayes, K-Nearest Neighbor, dan Logistic Regression</i>	88.46%, KNN dan NB kinerja terendah	
9	Identifikasi mahasiswa berisiko <i>dropout</i>	Menggunakan tujuh algoritma klasifikasi: <i>Naïve Bayes, Multilayer Perceptron, Logistic Regression, LWL, Random Forest, Random Tree, dan Part</i>	<i>Naïve Bayes</i> , LWL, dan Part mencapai akurasi tertinggi 86.36%; LWL memiliki F1 score terbaik 92.56%	Biswas et al.(2019)
10	Prediksi <i>dropout</i> mahasiswa di program studi	Tinjauan sistematis terhadap literatur menggunakan teknik data mining: <i>Decision Tree, Bayesian Classification, Neural Networks, Logistic Regression, dll.</i>	<i>Decision Tree</i> paling sering digunakan (67%), diikuti oleh <i>Bayesian Classification</i> (49%), <i>Neural Networks</i> (40%), dan <i>Logistic Regression</i> (34%)	Agrusti et al. (2019)

Dari tinjauan literatur pada penelitian-penelitian sebelumnya, teridentifikasi aspek kebaruan yang dapat dieksplorasi di STIKes Panti Rapih Yogyakarta. Meskipun berbagai studi telah mengkaji faktor-faktor yang mempengaruhi kinerja akademik dan risiko putus sekolah, masih terdapat kekurangan dalam pemahaman tentang pengaruh kegiatan non-akademik dan akademik, khususnya variabel poin kemahasiswaan yang belum sepenuhnya diteliti dalam konteks STIKes Panti Rapih Yogyakarta.

Penelitian ini bertujuan untuk mengeksplorasi variabel poin kemahasiswaan yang mencerminkan tingkat keterlibatan dan partisipasi dalam kegiatan akademik dan non-akademik, serta kontribusinya terhadap kinerja akademik dan kemungkinan putus sekolah mahasiswa. Dengan mengintegrasikan data dari variabel ini, diharapkan penelitian dapat memberikan wawasan baru yang signifikan mengenai hubungan antara tingkat keterlibatan mahasiswa dengan prestasi akademik. Hal ini akan memungkinkan institusi untuk mengidentifikasi mahasiswa yang berisiko tinggi dan mengambil langkah-langkah preventif yang lebih tepat sasaran, sehingga dapat meningkatkan retensi mahasiswa dan kualitas pendidikan di STIKes Panti Rapih Yogyakarta.

Selanjutnya, penelitian ini juga akan membandingkan beberapa algoritma machine learning, yaitu *Random Forest*, *Naive Bayes*, *K-Nearest Neighbor*, dan *Support Vector Machines* dalam prediksi *dropout*. Melalui perbandingan ini, penelitian bertujuan untuk mengidentifikasi algoritma yang paling efektif dan akurat dalam memprediksi *dropout* berdasarkan data yang tersedia. Hasil dari perbandingan ini diharapkan dapat memberikan pemahaman yang lebih mendalam mengenai kinerja masing-masing algoritma dalam konteks prediksi *dropout* di STIKes Panti Rapih Yogyakarta, serta rekomendasi praktis untuk implementasi algoritma prediksi yang lebih baik di masa mendatang.

