

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pendidikan memberikan dampak yang sangat penting terhadap pertumbuhan ekonomi suatu bangsa (Deda, Pacukaj, dan Vardari 2021). Pertumbuhan ekonomi dipengaruhi oleh kualitas tenaga kerja, modal dan kemajuan teknologi. Kualitas tenaga kerja bergantung pada tingginya tingkat pendidikan tenaga kerja, dengan kata lain semakin tinggi pendidikan yang dimiliki oleh tenaga kerja maka semakin berkualitas kinerjanya. Dengan demikian dapat disimpulkan bahwa semakin tinggi kualitas tenaga kerja suatu bangsa semakin tinggi pula pertumbuhannya ekonominya.

Di era perkembangan teknologi yang sangat pesat ini pendidikan dituntut untuk selalu mengikutinya. Teknologi merupakan komponen terpenting dari sebuah sistem informasi, sementara itu saat ini sistem informasi didalam dunia pendidikan dipergunakan disemua proses, mulai dari proses pendaftaran, seleksi, proses belajar mengajar, penilaian, pembayaran sampai dengan proses kelulusan. Sistem informasi didalam pendidikan menghasilkan sejumlah data besar yang berkaitan dengan mahasiswa dalam bentuk data. Sangat penting bagi pemangku kepentingan untuk secara efektif mengubah kumpulan data yang sangat besar ini menjadi pengetahuan yang akan membantu pengajar, administrator, dan pembuat kebijakan untuk menganalisisnya guna meningkatkan kualitas pendidikan. Selain itu, sistem informasi dapat dipergunakan untuk meningkatkan kualitas proses pendidikan dengan memberikan informasi yang tepat waktu kepada pemangku kepentingan (Asif dkk. 2017).

Sistem informasi di perguruan tinggi menghasilkan data yang melimpah, namun minim untuk mengubahnya menjadi informasi yang berguna (Morris 2016). Proses merubah data yang besar tersebut menjadikan informasi atau pengetahuan tidak mungkin dilakukan menggunakan cara-cara manual, oleh karena itu dibutuhkan suatu teknik atau metode yang mampu menggali informasi atau pengetahuan dari data tersebut secara cepat dan tepat. Metode yang dipergunakan untuk mencari informasi atau pengetahuan dari kumpulan data dapat dilakukan

dengan *Data Mining* (DM). DM digunakan untuk menganalisis data historis dari organisasi manapun dengan menggunakan algoritma, yang selanjutnya menemukan informasi tersembunyi dari data tersebut yang tidak mungkin ditemukan secara manual (Kumar dan Salal 2019).

Penerapan DM pada data pendidikan disebut sebagai *Educational Data Mining* (EDM) (Baker dan Yacef 2009). EDM adalah disiplin ilmu yang berkembang, berkaitan dengan perluasan metode DM klasik dan mengembangkan metode baru untuk menemukan data yang berasal dari sistem pendidikan (Alsuwaiket 2018). EDM dalam dua dekade terakhir ini mengalami perkembangan sangat pesat (Hernández-Blanco dkk. 2019; Maphosa dan Maphosa 2020; Romero dan Ventura 2020). Pesatnya penerapan ini dipengaruhi dengan data yang melimpah dan kecepatan perkembangan teknologi komputer yang memudahkan pengguna dalam melakukan pengolahan data. Bahkan EDM menjadi area penelitian tersendiri didalam ilmu DM, dimana peneliti dari belahan dunia membentuk asosiasi yang terhubung didalam portal website www.educationaldatamining.org, setiap tahunnya diadakan pertemuan melalui konferensi dari diberbagai negara yang tergabung didalamnya.

Kontribusi EDM semakin hari semakin berkembang, saat ini EDM merupakan metode yang telah digunakan dalam meningkatkan kualitas Pendidikan (Abu Saa, Al-Emran, dan Shaalan 2019; Al-Emran dkk. 2018; Asif dkk. 2017). EDM telah digunakan dalam mengevaluasi kualitas proses pembelajaran (Bakhshinategh dkk. 2018; Ibitoye, Borokini, dan Alabi 2019). EDM memiliki banyak area, setidaknya terdapat 11 area. Salah satu area EDM yang paling populer dan tertua adalah prediksi performa mahasiswa yang tujuannya adalah untuk memperkirakan nilai yang tidak diketahui dari variabel yang menggambarkan mahasiswa. Dalam pendidikan, nilai-nilai yang biasanya diprediksi adalah performa, pengetahuan, skor, atau nilai (Romero dan Ventura 2010). Prediksi performa mahasiswa sangat penting dalam sistem pendidikan saat ini. Prediksi performa mahasiswa dapat membantu mahasiswa dan pengajar dalam melacak kemajuan mahasiswa (Ragab dkk. 2021). Memprediksi performa mahasiswa dapat menciptakan peluang untuk meningkatkan hasil pendidikan. Dengan pendekatan

prediksi performa yang efektif, instruktur dapat mengalokasikan sumber daya dan instruksi yang lebih akurat. Prediksi awal performa mahasiswa dapat membantu pengambil keputusan untuk memberikan tindakan pada saat yang tepat dan untuk merencanakan pembelajaran yang tepat dalam meningkatkan tingkat keberhasilan mahasiswa (Alyahyan dan Düştegör 2020).

Upaya untuk meningkatkan keberhasilan mahasiswa terus dilakukan oleh para peneliti, sampai dengan saat ini terdapat lebih dari 13.653 dokumen terindeks scopus yang membahas terkait performa akademik siswa. Penelitian terkait performa akademik siswa yang terindek di Scopus pertama kali pada tahun 1954 oleh Reed M. Merrill dalam rangka mengevaluasi performa akademik mahasiswa pada masa percobaan. Namun perkembangan penelitian berkaitan tentang topik ini dari tahun-ketahun tidaklah signifikan. Sampai dengan tahun 1981 penelitian terkait topik ini belum mencapai angka 10 artikel per tahunnya. Namun pada tahun 1983-1992 artikel terkait performa akademik selalu mengalami peningkatan meskipun belum mencapai angka ratusan di setiap tahunnya. Dimulai tahun 1993-2013 artikel selalu selalu bertambah di setiap tahunnya. Tahun 2014 sampai dengan sekarang jumlah artikel terus meningkat, peningkatan signifikan dimulai dari tahun 2018 sampai dengan sekarang, bahkan di tahun 2021 artikel terkait tentang performa akademik siswa berjumlah 2.485 artikel. Perkembangan penelitian dengan fokus pada prediksi performa akademik siswa dalam sepuluh tahun terakhir juga mengalami peningkatan yang luar biasa. Tercatat 1027 artikel yang terindex di Scopus dari tahun 2012-2021. Berdasarkan hal tersebut dapat disimpulkan bahwa penelitian bidang prediksi performa mahasiswa masih sangat menarik dan penting untuk terus dikembangkan.

Penelitian dalam memprediksi performa mahasiswa berusaha untuk menentukan fitur yang saling terkait dan untuk mengidentifikasi alasan yang mendasari mengapa fitur tertentu lebih berpengaruh daripada fitur yang lain (Hellas dkk. 2018). Prediksi performa akademik mahasiswa merupakan bagian dari prediksi performa mahasiswa (Romero, Ventura, Sebastian Pechenizkiy, dan Baker 2010). Performa akademik mahasiswa merupakan komponen terpenting pada institusi pendidikan tinggi, hal ini dikarenakan lembaga pendidikan tinggi dituntut

untuk menghasilkan lulusan yang terampil dengan nilai akademik yang tinggi (Jenitha, Santhi, dan Monisha Privthy Jeba 2021). Upaya untuk memprediksi performa akademik mahasiswa demi menghasilkan lulusan dengan nilai akademik yang tinggi telah dilakukan oleh para peneliti dengan menggunakan berbagai pendekatan.

Pendekatan-pendekatan dalam prediksi performa akademik mahasiswa diantaranya yaitu (Amrieh, Hamtini, dan Aljarah 2016) mengusulkan model baru untuk prediksi performa mahasiswa berdasarkan teknik data mining dengan atribut/fitur data baru, yang disebut fitur perilaku siswa. Jenis fitur ini terkait interaksi mahasiswa dengan *Learning Management System* (LMS). Aluko dkk (2018) mengusulkan prediksi performa akademik mahasiswa dengan memanfaatkan informasi yang terkandung dalam prestasi akademik sebelumnya. IPK digunakan sebagai metrik untuk mengukur performa akademik mahasiswa. IPK dibagi menjadi 2 (lulus/gagal), IPK lebih besar dari 2,4 dinyatakan lulus dan IPK kurang dari 2,4 dinyatakan gagal. Helal dkk (2018) mengusulkan model klasifikasi yang berbeda untuk memprediksi performa mahasiswa, menggunakan data yang dikumpulkan dari data pendaftaran mahasiswa dan data aktivitas yang dihasilkan dari sistem manajemen pembelajaran universitas (LMS). Data pendaftaran berisi informasi mahasiswa seperti fitur sosiodemografi, dasar masuk universitas (misalnya melalui ujian masuk atau tanpa ujian) dan jenis kehadiran (misalnya penuh waktu dan paruh waktu). Ramaswami dkk (2019) mengumpulkan data interaksi siswa dengan LMS (Xorro-Q) selama satu semester dan satu mata kuliah, data dibagi dua yaitu partisipasi di dalam kelas dan diluar kelas. Sementara itu (Yaacob dkk. 2019) menggunakan metode *CRoss Industry Standard Process for Data Mining* (CRISP-DM) dalam memprediksi performa mahasiswa untuk dapat mengidentifikasi fitur penting dalam data pendidikan.

Data pendidikan yang digunakan untuk memprediksi performa akademik mahasiswa sebagian besar menggunakan dua jenis kumpulan data yaitu database perguruan tinggi/universitas dan data platform pembelajaran online (Albreiki, Zaki, dan Alashwal 2021). Pada umumnya para peneliti menggunakan data dari mesin pembelajaran (LMS) dan data yang bersumber dari lingkungan pendidikan dalam

melakukan prediksi performa akademik mahasiswa, namun beberapa peneliti menyampaikan bahwa perlu ditambahkan fitur-fitur lain untuk memprediksi performa akademik mahasiswa.

Fitur-fitur yang mendukung performa akademik terus diusulkan oleh banyak peneliti. Romer (1993); Cohn dkk (1998) meneliti hubungan antara tingkat kehadiran mahasiswa terhadap nilai akademik. Abu (2016) menemukan bahwa banyak fitur lain yang memiliki pengaruh signifikan terhadap prediksi akademik mahasiswa diantaranya yaitu; fitur demografis (jenis kelamin), pekerjaan orang tua, pra-pendaftaran (nilai sekolah menengah), dan diskon biaya Universitas. Sementara itu Alsinaidi dkk. (2021) menyampaikan bahwa fitur ekstrakurikuler mempengaruhi performa akademik mahasiswa. Arola Anderson dkk. (2021) menyampaikan bahwa aktifitas organisasi kampus pada semester pertama memiliki pengaruh positif terhadap optimis mahasiswa, membantu mengatasi perubahan (Bohnert, Fredricks, dan Randall 2010), berpengaruh positif terhadap perkembangan dan perilaku anak (Morris 2015), memiliki pengaruh terhadap masa tunggu lulusan dalam mendapatkan pekerjaan (Arifin, Widowati, dan Farikhin 2022).

Model untuk memprediksi performa akademik mahasiswa telah banyak dikembangkan oleh banyak peneliti, namun masih terdapat beberapa permasalahan yang muncul diantaranya adalah (1) Fitur yang paling berpengaruh terhadap prediksi performa mahasiswa antar peneliti berbeda-beda; Amrieh dkk. (2016) fitur yang paling berpengaruh adalah sumber daya yang dikunjungi, presensi harian, angkat tangan di kelas dan melihat pengumuman. Aluko dkk. (2018) fitur yang signifikan terhadap prediksi performa akademik adalah nilai Matematika, Biologi dan Fisika. Sementara itu Helal dkk. (2018) fitur yang paling berpengaruh adalah jenis kelamin, jenis kehadiran, mode kehadiran. Hal yang berbeda juga disampaikan oleh (Abu Saa dkk. 2019) fitur yang memiliki nilai pengaruh tertinggi adalah nilai akademik dan kinerja kelas, aktivitas *e-learning*, demografi, informasi sosial. Fitur membaca kritis, kompetensi, nilai Bahasa Inggris, penalaran kuantitatif dan nilai Biologi merupakan fitur yang paling berpengaruh terhadap prediksi performa akademik (Arifin dkk. 2021). (2) Belum adanya model prediksi performa akademik

mahasiswa dengan menggabungkan fitur akademik (LMS, nilai) dan nonakademik (demografi, ekonomi, organisasi kampus) secara bersamaan hal ini ditunjukkan pada Tabel 1.1.

Tabel 1. 1 Variabel-variabel prediksi performa akademik mahasiswa

NO	PENULIS	VARIABEL
1	Amrieh dkk., 2016	(1) Fitur demografis. (2) Fitur latar belakang akademik. (3) Fitur perilaku di LMS.
2	Aluko dkk., 2018	Ujian O-level (11 mata pelajaran), cara masuk (JAMB/Entri Langsung) dan skor nilai ujian tersier terpadu (JAMB)
3	Helal dkk., 2018	(1) LMS, (2) fitur sosiodemografi, (3) Dasar penerimaan (ujian masuk atau nilai SLTA) dan (4) Jenis kehadiran (misalnya penuh waktu vs paruh waktu).
4	Yaacob dkk., 2019	Transkrip nilai dan data mahasiswa
5	Sudani dan Ramaswamy, 2019	Kelembagaan, akademik, demografi, dan ekonomi
6	Leo & Yulia (2020)	Demografi, Akademik, LMS (lentera).
7	Arifin dkk, 2021	Informasi akademik, sosial, ekonomi dari 12.411 siswa di Bogota Colombia
8	Halit Karalar dkk, 2021	IPK, LMS (Moodle) dan CMS (Adobe Connect) selama satu semester

Proses pembangunan model yang telah dilakukan oleh para peneliti pada umumnya diawali dengan perbandingan algoritma, hal ini dilakukan untuk mencari algoritma terbaik dalam membangun model prediksi. Kumar dkk. (2012) menyampaikan bahwa tidak ada satu algoritma terbaik dari algoritma lainnya di semua dataset, sehingga perlu membandingkan beberapa algoritma sebelum menggunakannya. Perbandingan algoritma diperlukan sebelum menentukan algoritma yang akan digunakan untuk membangun model prediksi performa akademik mahasiswa (Arifin dkk. 2021). Beberapa peneliti telah membandingkan algoritma untuk mencari algoritma yang tepat diantaranya adalah Aluko dkk.

(2018) mengatakan bahwa *Support Vector Machine* (SVM) lebih baik dari *Logistic Regression* (LR) dalam hal akurasi untuk memprediksi performa akademik mahasiswa. Kumar dan Salal (2019) menyebutkan bahwa *Decision Tree* (DT) merupakan algoritma terbaik dengan nilai akurasi 98,86% dibandingkan dengan *Neural Network* (NN), *Naïve Bayes* (NB), *K-Nearest Neighbor* (K-NN), dan SVM. Sementara itu, Mengash (2020) mengatakan bahwa *Artificial Neural Network* (ANN) adalah algoritma yang terbaik dibandingkan dengan algoritma DT, SVM, dan NB. Yaacob dkk. (2019) menemukan algoritma NB merupakan algoritma terbaik dibandingkan dengan KNN, DT dan LR. Arifin dkk. (2021) membandingkan algoritma *Generalized Linear Model* (GLM), *Deep Learning* (DL), DT, *Random Forest* (RF), dan SVM dan *Gradient Boosted Trees* (GBT) untuk memprediksi performa akademik mahasiswa, adapun hasil perbandingan menunjukkan bahwa algoritma GBT memiliki tingkat kesalahan terkecil dibanding dengan algoritma lainnya. GBT merupakan mesin pembelajaran jenis *regression* dan *classification*, yang menghasilkan model prediksi berbentuk *ensemble* (*Boosting*) dari model *decision tree* yang lemah seperti, GBT telah berhasil digunakan dalam berbagai penerapan (Fernandes 2019; Friedman 2001; Hew dkk. 2020; Ketui, Wisomka, dan Homjun 2019; Sawant 2019; Zhang dkk. 2019).

GBT bekerja dengan sangat baik pada kumpulan data berukuran sedang. GBT adalah pendekatan *ensemble* dengan melatih banyak pelajar individu, biasanya pohon keputusan. Pohon dilatih secara berurutan, dengan setiap pohon belajar dari kesalahan sebelumnya, tidak seperti di RF di mana pohon dilatih secara bersamaan. *Gradient Descent* digunakan untuk mengumpulkan kontribusi dari setiap individu yang dipelajari selama pelatihan untuk membuat satu pembelajar ansambel yang kuat dari ratusan atau ribuan pembelajar yang lemah (bobot dari masing-masing pohon karenanya akan menjadi parameter model). Untuk menghasilkan model dengan akurasi tinggi GBT membutuhkan nilai *hyperparameter* yang tepat.

Banyak *hyperparameter* tersedia di GBT yang dapat disesuaikan untuk mempengaruhi ansambel umum (seperti *learning rate*) dan pohon keputusan individu (seperti *number of leaves* atau *maximum depth*). Karena ada interaksi yang rumit antara *hyperparameter*, maka sulit untuk menentukan kombinasi

hyperparameter mana yang akan berfungsi paling baik hanya berdasarkan teori. Dengan demikian, penyetelan *hyperparameter* diperlukan karena tidak ada metode lain untuk menentukan nilai *hyperparameter* terbaik selain menguji sejumlah besar kemungkinan kombinasi. Penyetelan *hyperparameters* sangat penting sebelum menerapkan algoritma prediksi (Probst, Boulesteix, dan Bischl 2019; Victoria dan Maragatham 2021; Weerts, Mueller, dan Vanschoren 2020; Zhang, Wang, dan Shen 2022). Kinerja model tergantung pada pemilihan *hyperparameter* (Shekhar, Bansode, dan Salim 2021). Pilihan konfigurasi *hyperparameter* diketahui berdampak signifikan terhadap kinerja model mesin pembelajaran (Zhang dkk. 2022) Untuk menentukan kinerja terbaik dari algoritma mesin pembelajaran, para peneliti telah melakukan penyetelan *hyperparameter* (Duarte dan Wainer 2017; Elgeldawi dkk. 2021; Lorraine, Vicol, dan Duvenaud 2019; MacKay dkk. 2019; Probst, Wright, dan Boulesteix 2019; Rijdsdijk dkk. 2021; Schratz dkk. 2019; Victoria dan Maragatham 2021; Wong dkk. 2019; Zhang dkk. 2022).

Berdasarkan hal tersebut diatas maka pada penelitian ini akan membangun sebuah sistem informasi dengan menggunakan model EDM berbasis algoritma GBT yang dioptimasi untuk memprediksi performa akademik mahasiswa dalam hal Indeks Prestasi Semester (IPS) dengan menggabungkan variabel akademik (nilai akademik, LMS) dan nonakademik (data ekonomi, demografi, organisasi kampus) sebagai sumber datanya serta menerapkan metode CRISP-DM untuk membangun model prediksinya.

1.2. Rumusan Masalah

Merujuk pada latar belakang diatas maka rumusan masalah dalam penelitian ini adalah:

1. Belum adanya sistem informasi prediksi performa akademik mahasiswa berbasis model EDM yang secara komprehensif menggabungkan data akademik (nilai akademik, LMS) dan data nonakademik (demografi, ekonomi, organisasi kampus).

2. Fitur-fitur apa saja yang memiliki dampak signifikan terhadap prediksi performa akademik mahasiswa dari penggabungan data akademik dan nonakademik.
3. Dalam membangun model prediksi performa akademik mahasiswa yang mencakup fitur-fitur akademik dan nonakademik, seberapa akurat algoritma GBT setelah dilakukan optimasi *hyperparameter*.

1.3. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Membangun model EDM untuk prediksi performa akademik mahasiswa berbasis algoritma GBT dengan menggunakan data akademik dan nonakademik.
2. Menemukan fitur yang paling berpengaruh dalam memprediksi performa akademik mahasiswa.
3. Meningkatkan akurasi prediksi dengan melakukan optimasi *hyperparameter* pada algoritma GBT.
4. Membangun sistem informasi prediksi performa akademik mahasiswa dengan menggunakan model EDM berbasis algoritma GBT yang dioptimasi.

1.4. Manfaat penelitian

Manfaat dari penelitian prediksi performa mahasiswa ini diantaranya adalah sebagai berikut:

1. Terbentuknya model EDM untuk prediksi performa akademik mahasiswa berbasis algoritma GBT dengan menggunakan data akademik dan nonakademik.
2. Memberikan rekomendasi fitur-fitur yang dapat dipergunakan untuk memprediksi performa akademik mahasiswa.
3. Menemukan peningkatan akurasi prediksi dari algoritma GBT setelah dilakukan optimasi.

4. Memberikan kemudahan terhadap pengguna sistem informasi, dalam memprediksi performa akademik mahasiswa.

1.5. Kontribusi Penelitian

Kontribusi penelitian ini adalah:

1. Membangun model EDM berbasis regresi dengan menggunakan algoritma GBT teroptimasi berbasis GS untuk memprediksi performa akademik mahasiswa dengan menggabungkan fitur akademik dan nonakademik.
2. Menemukan fitur-fitur yang berpengaruh terhadap prediksi performa akademik mahasiswa dari gabungan fitur akademik dan nonakademik.
3. Membangun sistem informasi prediksi performa akademik mahasiswa berbasis algoritma GBT dengan data akademik dan nonakademik sebagai inputan.

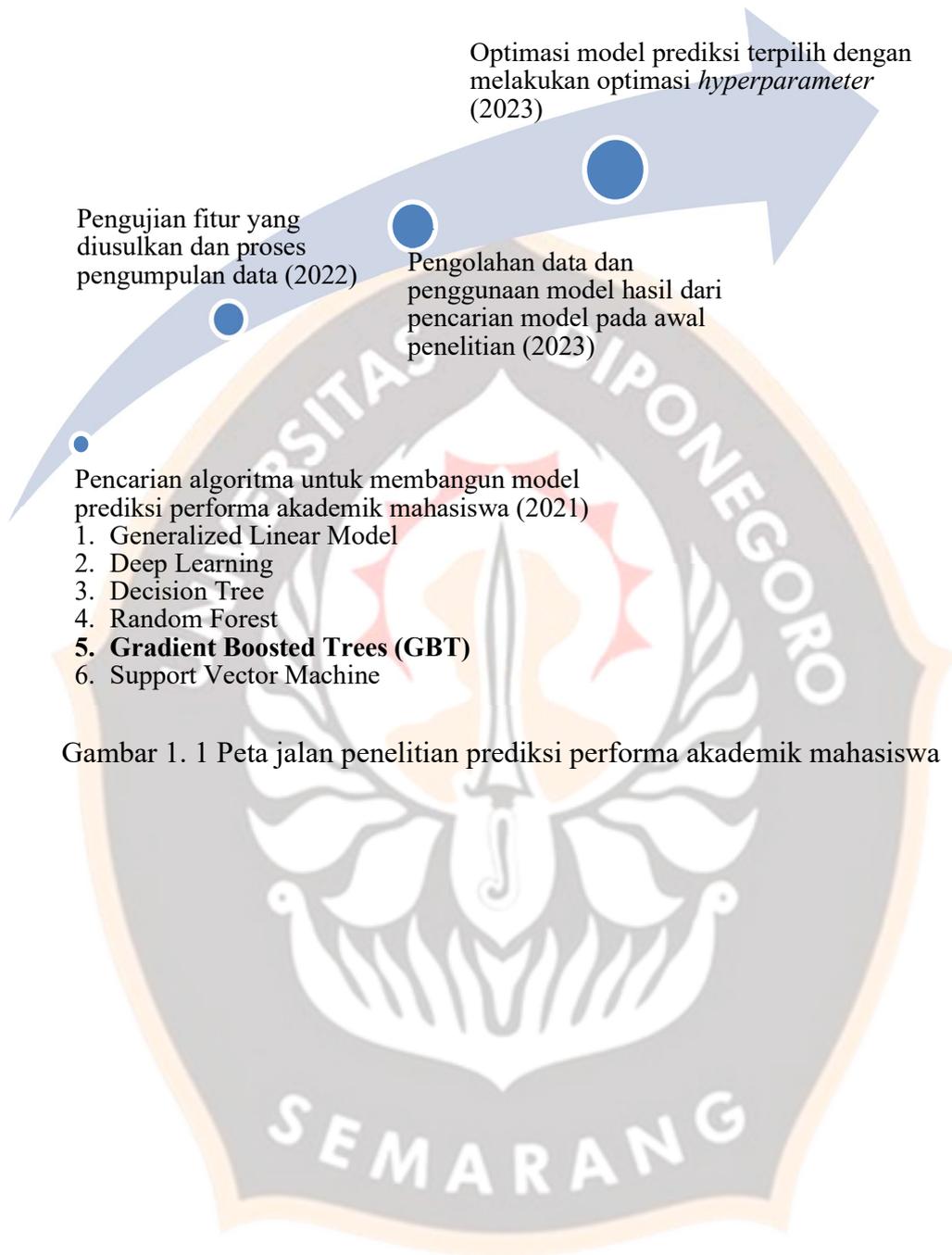
1.6. Luaran penelitian

1. Publikasi artikel pada seminar internasional terindex scopus/wos
2. Publikasi artikel pada jurnal internasional terindex scopus
3. Buku ber-ISBN
4. Sistem informasi prediksi performa akademik
5. HKI Buku dan Aplikasi Komputer
6. Panduan penggunaan aplikasi Sistem Informasi Prediksi Performa Akademik Mahasiswa

1.7. Peta Jalan (Road Map) Penelitian

Agar penelitian berjalan dengan baik maka perlu disusun sebuah peta jalan penelitian, hal ini dapat memberikan gambaran secara jelas apa yang akan dikerjakan oleh peneliti dalam kurun waktu yang telah ditentukan, peta jalan penelitian ditunjukkan pada Gambar 1.1.

SEKOLAH PASCASARJANA



Gambar 1. 1 Peta jalan penelitian prediksi performa akademik mahasiswa

SEKOLAH PASCASARJANA