

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Perguruan tinggi cenderung menjadi tolak ukur yang digunakan masyarakat untuk menentukan kualitas pendidikan mahasiswa sebagai sumber daya manusia. Umumnya, perguruan tinggi dinilai unggul jika mahasiswa yang dihasilkan berkualitas dan memiliki prestasi yang baik. Namun sebaliknya, kegagalan mahasiswa berdampak negatif bagi mahasiswa maupun perguruan tinggi. Saat ini, isu kegagalan mahasiswa yang merupakan tantangan di perguruan tinggi adalah kasus *drop out*. *Drop out* dinyatakan sebagai tindakan pinalti bagi mahasiswa yang tidak mampu menyelesaikan pendidikan sesuai dengan batas waktu yang ditentukan. Jika dibandingkan dengan mahasiswa yang lulus tepat waktu, pendidikan yang diterima sangat minim dan berpengaruh secara signifikan pada kualitas mahasiswa dan institusi (Kovačić, 2010).

*Drop out* bukan merupakan masalah yang baru tetapi cukup serius dan menarik perhatian peneliti dilihat dampaknya terhadap penurunan akreditasi universitas dan dampak negatif terhadap lingkungan sosial dimana kesempatan meraih pendidikan bagi orang lain hilang begitu saja. Penelitian sudah banyak dilakukan dengan memanfaatkan teknologi untuk menemukan cara mencegah terjadinya isu *drop out* ini, yang saat ini disebut *Educational Data Mining* (Patil & Kumar, 2017). *Educational Data Mining* (EDM) merepresentasikan beragam metode algoritma untuk menangani berbagai isu pada sistem pendidikan bahkan menghasilkan pengetahuan baru, khususnya dalam memprediksi variabel atau indikator yang berpengaruh pada munculnya kasus *drop out* di perguruan tinggi.

Pada pendidikan berbasis online, (Yukselturk, Ozekes, & Turel, 2014) menemukan performa metode algoritma *machine learning* lebih baik dalam melakukan prediksi. Yukselturk menggunakan *Genetic Algorithm* (GA) untuk mengevaluasi kemampuan algoritma *k-Nearest Neighbour* (k-NN), *Decision Tree* (DT), *Naive Bayes* (NB) dan *Neural Network* (NN). Algoritma KNN mencapai 87% pada tingkat sensitifitas untuk memprediksi mahasiswa *drop out* menggunakan dataset mahasiswa program pendidikan online. Yukselturk menemukan kesiapan diri mahasiswa di pilihan bidang

Pada penelitian lain, Abu-Oda (Abu-Oda & El-Halees, 2015) meneliti dan menganalisa kemampuan algoritma *Decision Tree* dalam melakukan prediksi dan hasilnya 98.14% akurat. Abu Oda menganalisis data siswa aktif dalam dua tahun pertama dan menemukan bahwa variabel gender diidentifikasi sebagai penyebab kegagalan yang dominan dengan melakukan *association analysis* menggunakan algoritma FP-Growth untuk menemukan pola yang saling terkait pada data mahasiswa yang sangat besar. Tujuannya untuk meningkatkan performansi mahasiswa dan mengurangi jumlah mahasiswa yang *dropout*. Pada dasarnya, kemungkinan dropout tidak hanya terjadi pada siswa reguler atau konvensional, tetapi juga siswa program pembelajaran online memiliki potensi yang tinggi untuk mengundurkan diri (Yukselturk, Ozekes, & Turel, 2014).

Di sisi lain, Márquez (Márquez-Vera, et al., 2016) mengusulkan metode baru untuk mendeteksi lebih awal peluang mahasiswa *drop out* disebut algoritma *Classification Rule Mining* (ICRM) yang dimodifikasi untuk hasil pengukuran yang lebih akurat. Metode ini diadaptasi untuk mendeteksi peluang mahasiswa *drop out* meskipun dengan jumlah data yang kecil. Metode ini berhasil men-

*generate* sekumpulan *rules* atau pemodelan baru untuk menangani data siswa yang tidak seimbang dan menemukan variabel yang tepat. Hasil analisisnya pada 60 atribut data, ada korelasi antara kebiasaan siswa, kondisi sosial dan kinerja siswa. Beberapa indikator seperti IPK, pendidikan orang tua, jam kerja, dan kebiasaan buruk siswa dapat digunakan mengukur potensi terjadinya masalah *dropout*.

Secara umum, kemungkinan *drop out* terjadi juga disebabkan keterbatasan finansial, absensi mahasiswa, pengaruh orangtua, kesempatan kerja, pernikahan dini, demografis dan performansi akademik (Sultana, Khan, & Abbas, 2017). Di Indonesia, statistik data per tahun 2017 (Kemenristekdikti, 2017), jumlah mahasiswa yang terdaftar sejumlah 6.924.511 mahasiswa dengan angka mahasiswa *drop out* sebesar 195.176 mahasiswa. Data menunjukkan bahwa persentase siswa dropout di universitas swasta lebih besar daripada universitas negeri dan beberapa daerah memiliki persentase dropout sangat tinggi.

Di tahun 2014, Utomo juga telah melakukan penelitian (Utomo, Reimondos, Utomo, McDonald, & Hull, 2014) menggunakan 799 orang yang mengalami dropout di lembaga pendidikan di Jakarta. Menurut Utomo, dropout biasanya terjadi pada orang berusia 12 hingga 19 tahun yang berasal dari daerah pinggiran kota dan pedesaan dengan latar belakang ekonomi rata-rata rendah. Indikator ekonomi yang rendah memicu siswa untuk memilih bekerja dibandingkan dengan melanjutkan studi di perguruan tinggi. Pada penelitiannya, Fahrudin (Fahrudin, Buliali, & Fatichah, 2016) melakukan penelitian terhadap mahasiswa tahun pertama dan menemukan variabel usia, program studi juga memiliki korelasi dan pengaruh pada prediksi mahasiswa dropout.

Pada penelitian ini, studi kasus yang digunakan adalah data mahasiswa Fakultas Ilmu Sosial dan Ilmu Politik Universitas Kristen Indonesia, salah satu universitas swasta di Jakarta. Fakultas Ilmu Sosial dan Ilmu Politik merupakan salah satu fakultas dengan angka penerimaan mahasiswa yang tinggi dimana mayoritas 30% jumlah mahasiswa yang diterima adalah penerima beasiswa dikarenakan keterbatasan finansial dan berasal dari daerah tertinggal. Dari keseluruhan fakultas, selain angka penerimaan mahasiswa yang tinggi, fakultas ini juga memiliki jumlah mahasiswa yang banyak memilih untuk berhenti melanjutkan studi di dua tahun masa studi. Berdasarkan hasil observasi, saat ini fakultas belum memiliki metode evaluasi yang konkrit untuk mengetahui dan menangani munculnya kasus *dropout*. Hal ini mengakibatkan  $\pm 30$  kasus *dropout* atau mahasiswa tidak aktif muncul di setiap tahunnya.

Oleh karena itu, untuk dapat menghasilkan mahasiswa yang berkualitas, fakultas perlu mengidentifikasi indikator yang menyebabkan mahasiswa mengundurkan diri dari masa studi sejak dini sehingga fakultas dapat menangani resiko terjadinya *dropout*. Dalam hal ini, penerapan metode *machine learning* menjadi salah satu cara untuk menemukan variabel atau indikator penyebab terjadinya isu *dropout* di perguruan tinggi. Metode yang diusulkan pada penelitian ini adalah metode *Ensemble Stacking Classification Method* menggunakan *Gradient Boosting* untuk mengidentifikasi variabel akademik dan variabel demografik yang mempengaruhi potensi terjadinya *dropout*. Selain itu, masing-masing variabel juga diuji menggunakan pendekatan algoritma diantaranya *K-Nearest Neighbour*, *Decision Tree*, dan *Naïve Bayes* untuk membandingkan

performansinya dengan metode yang diusulkan yaitu *Ensemble Stacking Classification Method* menggunakan *Gradient Boosting*.

Dari penelitian ini, hasil yang diharapkan adalah untuk menemukan variabel yang memiliki pengaruh signifikan terhadap terhadap prediksi mahasiswa *dropout* sehingga membantu fakultas untuk melakukan evaluasi mahasiswa. Selain itu, untuk membuktikan metode pemodelan *Ensemble Stacking Classification* menggunakan algoritma *Gradient Boosting* teruji baik dengan nilai akurasi prediksi yang tinggi.

## 1.2 Perumusan Masalah

Permasalahan yang diangkat pada penelitian ini adalah :

1. Bagaimana penerapan metode *Ensemble Stacking Classification* dalam mengidentifikasi dan memprediksi variabel yang dapat menentukan mahasiswa *dropout* agar dapat menjadi tolak ukur dalam melakukan evaluasi mahasiswa ?
2. Bagaimana perbandingan tingkat akurasi prediksi mahasiswa *dropout* menggunakan beberapa metode *machine learning* dengan metode yang diusulkan yaitu *Ensemble Stacking Classification* ?

## 1.3 Tujuan Penelitian

Tujuan penelitian ini dilakukan adalah :

- a. Menemukan variabel yang mengakibatkan terjadinya *drop out* di institusi pendidikan tinggi.
- b. Melakukan perbandingan dan evaluasi metode *Ensemble Stacking Classification* dan beberapa metode *machine learning* dengan

mengukur tingkat akurasi prediksi dalam memprediksi penyebab terjadinya *drop out* mahasiswa.

#### **1.4 Manfaat Penelitian**

Penelitian ini dilakukan untuk memberikan manfaat kepada :

- a. Membantu *top management* untuk mencegah dan menemukan solusi awal menangani kasus *drop out* mahasiswa di universitas.
- b. Menemukan indikator penyebab *drop out* dan meminimalisasi pengaruh *drop out* serta membantu mahasiswa untuk menyelesaikan studi tepat waktu.

#### **1.5 Ruang Lingkup**

Ruang lingkup penelitian ini adalah sebagai berikut:

- a. Mengumpulkan data mahasiswa yang akan digunakan sebagai materi uji penentuan keputusan selama 3 tahun terakhir (2016-2018) yaitu data demografik dan kinerja akademik mahasiswa Fakultas Ilmu Sosial dan Ilmu Politik Universitas Kristen Indonesia.
- b. Menggunakan sampel dataset sebanyak 20.000 sampel data mahasiswa.
- c. Menganalisa variabel yang berpotensi sebagai faktor terjadinya *drop out* menggunakan metode *machine learning* yaitu *K-Nearest Neighbour*, *Decision Tree* dan *Naïve Bayes*.