

## BAB I

### PENDAHULUAN

#### 1.1 Latar Belakang

Pendidikan tinggi merupakan pendidikan jenjang untuk menuntut ilmu setelah pendidikan menengah (SMA/Sederajat) yang terdiri dari berbagai jenjang, yaitu jenjang diploma, sarjana, magister, doktor, dan program profesi. Badan yang menjalankan pendidikan tinggi disebut perguruan tinggi dan dikenal dengan Perguruan Tinggi Negeri (PTN) dan Perguruan Tinggi Swasta (PTS). Bentuk perguruan tinggi ada bermacam-macam, seperti universitas, institut, sekolah tinggi, politeknik, spesialis, dan akademi. (Undang-Undang Republik Indonesia Nomor 12 pasal 1, 2012). Dengan semakin banyak perguruan tinggi yang dimiliki Indonesia, menunjukkan bahwa pemerintah dan masyarakat semakin sadar akan pentingnya sebuah pendidikan, terkhusus pendidikan formal.

Salah satu tolak ukur kualitas sebuah perguruan tinggi adalah dilihat dari akreditasi yang diperoleh. Lembaga yang berwenang untuk menilai akreditasi perguruan tinggi adalah Badan Akreditasi Nasional-Perguruan Tinggi (BAN-PT). Setiap perguruan tinggi berusaha untuk terus memperbaiki mutu universitas dengan meningkatkan pendidikan dan akreditasinya. Salah satu elemen penilaian akreditasi perguruan tinggi adalah lulus tepat waktu. Selain itu, ketepatan lama studi mahasiswa merupakan isu yang penting karena ketepatan tersebut menjadi dasar

efektifnya suatu perguruan tinggi (Untari, 2014). Ketepatan lama studi mahasiswa sudah lama diatur oleh pemerintah dalam peraturan menteri. Mahasiswa dikatakan tepat waktu jika menempuh masa studi sarjana/S1 maksimal selama 4 tahun atau kurang (PERMENDIKNAS No. 232, 2000).

Jumlah Perguruan Tinggi Negeri dan Perguruan Tinggi Swasta sebanyak 4.593 ([pddikti.kemdikbud.go.id](http://pddikti.kemdikbud.go.id)). Universitas Muhammadiyah Semarang (Unimus) adalah salah satu Perguruan Tinggi Swasta yang terletak di Kota Semarang, Provinsi Jawa Tengah. Tanggal 4 Agustus 1999 Menteri Pendidikan dan Kebudayaan Republik Indonesia menerbitkan Ijin Operasional Universitas Muhammadiyah Semarang dengan nomor : 139/D/O/1999. Program studi yang memperoleh ijin operasional pada awal pembukaan tahun 1999 sebanyak 14 program studi meliputi: Program Studi Kesehatan Masyarakat (S-1), Program Studi Statistik (S-1), Program Studi Teknik Mesin (S-1), Program Studi Teknik Mesin (D-3), Program Studi Teknik Elektro (S-1), Program Studi Teknik Elektro (D-3), Program Studi Teknologi Pangan (S-1), Program Studi Teknologi Pertanian (S-1), Program Studi Manajemen (S-1), Program Studi Administrasi Perusahaan (D-3), Program Studi Akuntansi (D-3), Program Studi Bahasa Inggris (S-1), Program Studi Bahasa Inggris (D-3), Program Studi Bahasa Jepang (D-3) ([unimus.ac.id](http://unimus.ac.id)).

Sebagai Perguruan Tinggi yang terbilang masih sangat muda, Unimus yang terakreditasi-B saat ini sudah berkembang dengan seiring bertambahnya Program Studi yang saat ini sudah mencapai 28 Program Studi. Tentunya dengan semakin banyaknya Program Studi yang sudah tersedia dan seiring bertambahnya usia, Unimus sudah menghasilkan lulusan yang tentunya diharapkan berjalan sesuai salah satu misi Unimus, yaitu menghasilkan lulusan yang kompeten, mampu mengisi dan atau menciptakan

lapangan kerja. Untuk menggapai semua itu tentunya Unimus juga mempertimbangkan aspek ketepatan waktu mahasiswanya untuk menempuh lama studi, karena itu merupakan salah satu aspek penilaian akreditasi BAN-PT. Lama studi mahasiswa tentu bisa saja terjadi karena banyak hal atau faktor, misalnya saja dalam penelitian ini ada faktor pekerjaan orangtua dan IPK Mahasiswa yang bersangkutan.

Penelitian terdahulu mengenai ketepatan lama studi mahasiswa menggunakan metode algoritma *Naive Bayes* oleh Arief Jananto (2013) menghasilkan hasil uji coba diperoleh tingkat kesalahan prediksi berkisar 20% sampai dengan 50% dengan data *training* dan *testing* diambil secara *random*. Namun rata-rata tingkat kesalahan berkisar 20% hingga 34%. Tinggi rendahnya tingkat kesalahan dapat disebabkan oleh jumlah *record* data dan tingkat konsistensi dari data *training* yang digunakan. Sedangkan hasil prediksi dari ketepatan lama studi dari mahasiswa angkatan 2008 adalah sebesar 254 mahasiswa diprediksi “Tepat Waktu” dan sisanya yaitu 4 orang diprediksi “Tidak Tepat Waktu”.

Selanjutnya, penelitian terdahulu mengenai ketepatan lama studi mahasiswa menggunakan metode *Regresi Logistik Biner* dan *Naive Bayes Classifier* oleh Erene Fajrila (2018) menghasilkan klasifikasi yang menggunakan *Regresi Logistik Biner* menggunakan variabel yang signifikan tersebut mendapatkan akurasi sebesar 82,29%, sedangkan hasil klasifikasi menggunakan algoritma *Naive Bayes Classifier* 89,10%. Hal ini dapat dikatakan pengklasifikasian menggunakan *Naive Bayes Classifier* lebih baik dibandingkan *Regresi Logistik Biner*.

Penerapan *Support Vector Machine* dalam dunia medis oleh Rachman dan Purnami (2012) yang digunakan untuk mengetahui tingkat keganasan dari penyakit kanker payudara (*breast cancer*). Penelitian tersebut dilakukan karena di Indonesia adalah Negara berkembang yang penderita penyakit tersebut sangat banyak. Hal itu dibuktikan dengan data Sistem Informasi Rumah Sakit (SIRS) 2007 yang menunjukkan kanker payudara menempati urutan pertama pada pasien rawat inap di seluruh rumah sakit Indonesia. Penelitian tersebut menggunakan pendekatan 2 metode, yaitu regresi logistik ordinal dan SVM. Hasil tingkat akurasi untuk menentukan keganasan kanker payudara untuk metode SVM memiliki akurasi yang lebih tinggi dari pada metode regresi logistik ordinal. Tingkat akurasi menggunakan SVM adalah sebesar 98,11% dan akurasi untuk regresi logistik ordinal hanya sebesar 56,60%. Tingkat akurasi tersebut didapatkan dengan melakukan uji coba untuk pembagian data *training* dan *testing* sebanyak 3, yaitu 50% *training* dan 50% *testing*, 75% *training* dan 25% *testing*, dan 80% *training* dan 20% *testing*. Selain itu, uji coba mencari parameter  $C = 10, 100, \text{ dan } 1000$ , serta nilai  $\gamma = 1, 2, \text{ dan } 3$ . Setelah uji coba, didapatkan bahwa pembagian data optimum di 75% *training* dan 25% *testing* menggunakan nilai parameter  $C=1000$  dan nilai  $\gamma=1$ .

Penelitian selanjutnya adalah penelitian oleh Octaviani, Wulandari, dan Ispriyanti (2014) dengan menggunakan metode SVM untuk menentukan klasifikasi akreditasi pada SD di Kota Malang. Data akreditasi yang digunakan data akreditasi SD di Kota Malang tahun 2011 sampai 2013 yang didapat dari *website* resmi BAN-S/M dengan 3 status, yaitu Akreditasi A, Akreditasi B, dan Akreditasi C. Sedangkan untuk variabel *independen* terdiri dari 7 komponen standar, yaitu Standar Isi ( $X_1$ ), Standar Proses ( $X_2$ ), Standar Kompetensi Lulusan ( $X_3$ ), Standar Pendidik dan Tenaga Kependidikan ( $X_4$ ), Standar

Sarana dan Prasarana ( $X_5$ ), Standar Pengelolaan ( $X_6$ ), dan Standar penilaian pendidikan ( $X_7$ ). Penelitian tersebut menggunakan metode SVM dengan kernel *Radian Basis Function* (RBF) dan kernel *Polynomial*. Tingkat akurasi yang didapat untuk menentukan akreditasi SD di Kota Malang menggunakan SVM kernel RBF sebesar 93,92% dan untuk SVM kernel *polynomial* adalah sebesar 92,67%.

Penelitian selanjutnya oleh Daqiqil, Astread, dan Mahdiyah (2017) tentang penerapan metode *Random Forest* dan *boosted C5.0* untuk diagnosa kanker payudara. Penelitian tersebut adalah mencari tingkat akurasi dari metode *Random Forest* dan *Boosted C5.0* dan kemudian diterapkan untuk deteksi diagnosa penyakit kanker payudara. Penelitian ini menjelaskan bahwa *Boosted C5.0* adalah pengembangan dari algoritma C4.5 dan ID3 (*Iterative Dichotomiser 3*). Algoritma C5.0 memiliki fitur yang lebih lengkap (*winning* dan *boosting*), lebih cepat, lebih efisien, dan menghasilkan *tree* yang lebih sederhana dari C4.5. Sedangkan *Random Forest* bekerja dengan cara membangun prediktor dengan sekumpulan *decision tree* yang berkembang secara acak pada subruang data *training*. Hasil yang diberikan oleh *Random Forest* untuk klasifikasi adalah modus dari *decision treenya*. Hasil akurasi algoritma *Random Forest* menunjukkan performa yang lebih baik dengan akurasi 95,7% dan C5.0 dengan akurasi 93,7%.

Penelitian lainnya oleh Perdana, Soelaiman, dan Farichah (2017) tentang pengelompokan berkas musik berdasarkan kemiripan karakteristik suara menggunakan metode *Random Forest*, *Multi layer perceptron*, dan J48. Pengelompokan musik berdasarkan karakteristik suara sangat penting bagi penikmat musik. Banyak peminat musik yang menikmati musik berdasarkan *genre* favorit masing-masing. Karena itu dibutuhkan metode ekstraksi fitur yang tepat untuk dapat merepresentasikan berkas musik

berdasarkan *genre* dengan baik. Studi ini melakukan ekstraksi fitur berkas musik dengan mengekstraksi fitur *spectral centroid*, *spectral flux*, *spectrall rolloff*, dan *short time energy* pada tiap berkas musik yang diolah dan kemudian dihitung nilai *mean*, *median*, *skewness*, dan kurtosisnya, dan selanjutnya dikelompokkan menggunakan metode klasifikasi. Tingkat akurasi dari metode klasifikasi tertinggi adalah metode *Random Forest* dengan akurasi 80%, kemudian metode *multi layer percepton* dengan akurasi 77,62%, dan yang terakhir adalah metode J48 dengan akurasi 62,86%. Dari tingkat akurasi, berarti dapat dikatakan bahwa untuk mengelompokkan suara musik berdasarkan *genre* terbaik adalah menggunakan metode *Random Forest*.

Klasifikasi adalah metode untuk memprediksi suatu kejadian atau keputusan yang akan datang berada di suatu titik. Klasifikasi merupakan suatu pekerjaan yang melakukan penilaian terhadap suatu objek data untuk masuk dalam suatu kelas tertentu dari sejumlah kelas yang tersedia (Prasetyo, 2012). Metode *Support Vector Machine* (SVM) dan *Random Forest* adalah bagian dari metode klasifikasi. Metode analisis klasifikasi yang digunakan adalah SVM karena memiliki tingkat akurasi yang cukup tinggi dalam hal klasifikasi teks dan cocok untuk data yang memiliki dimensi besar (Naradhipa dan Purwarianti, 2011). Sedangkan *Random Forest* digunakan karena tingkat akurasi dari metode ini lebih tinggi daripada metode *Regresi Logistik* dan *Decision Tree*, yang mana *Random Forest* akan membuat pohon klasifikasi dengan jumlah banyak dan kemudian digabungkan untuk mencari tingkat akurasi yang tinggi (Sartono dan Syafitri, 2010).

Berdasarkan latar belakang dan penelitian terdahulu yang telah dijabarkan, penelitian ini disesuaikan dengan penelitian-penelitian sebelumnya, baik dari segi metode maupun objek penelitian. Dengan adanya penelitian terdahulu, penelitian yang saat ini dilakukan

memiliki arti penting sehingga dapat memberikan informasi dan kontribusi dalam menganalisis terkait ketepatan lama studi seorang mahasiswa Universitas Muhammadiyah Semarang.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah pada penelitian ini yaitu :

1. Bagaimana gambaran secara umum tentang ketepatan lama studi dari alumni Unimus tahun kelulusan 2011-2020?
2. Bagaimana hasil klasifikasi menggunakan metode SVM dan *Random Forest* dalam memprediksi lama studi mahasiswa?
3. Bagaimana tingkat akurasi menggunakan metode SVM dan *Random Forest* dalam memprediksi lama studi mahasiswa?
4. Manakah metode terbaik antara SVM dan *Random Forest* dalam melakukan prediksi ketepatan lama studi mahasiswa?

## 1.3 Tujuan Penelitian

Adapun tujuan penelitian berdasarkan rumusan masalah tersebut yaitu :

1. Mendapatkan gambaran secara umum mengenai ketepatan lama studi dari alumni Unimus tahun kelulusan 2011-2020.
2. Menentukan seberapa tepat metode SVM dan *Random Forest* dalam memprediksi ketepatan lama studi dengan melihat klasifikasinya.

3. Menentukan seberapa tepat metode SVM dan *Random Forest* dalam memprediksi ketepatan lama studi dengan melihat akurasinya.
4. Melihat metode mana yang memiliki tingkat akurasi terbaik untuk memprediksi ketepatan lama studi.

#### 1.4 Manfaat Penelitian

Berdasarkan tujuan penelitian yang ingin dicapai, maka penelitian ini diharapkan mempunyai manfaat sebagai berikut:

1. Manfaat Teoritis

Membantu perkembangan ilmu pengetahuan mengenai metode peramalan (*forecasting*) sehingga dapat digunakan sebagai bahan bacaan dan referensi bagi pembaca dalam melakukan penelitian yang berhubungan dengan peramalan terutama pada metode *Support Vector Machine* (SVM) dan *Random Forest*.

2. Manfaat Praktis

Bagi Peneliti, peneliti mampu menerapkan metode yang sesuai dalam materi yang telah dipelajari serta peneliti mempunyai pengetahuan dan wawasan mengenai peramalan menggunakan *Support Vector Machine* (SVM) dan *Random Forest*.

- a. Bagi pihak terkait yaitu dapat diketahui ketepatan lama studinya mahasiswa selama berkuliah.
- b. Bagi pihak terkait dapat lebih memperhatikan mahasiswa mengenai ketepatan lama studi mahasiswa guna menunjang akreditasi.

## 1.5 Batasan Masalah

Batasan masalah diberikan agar mendapatkan kejelasan pada penelitian ini. Batasan masalah pada penelitian ini yaitu menggunakan data historis alumni Unimus tahun 2011-2020 dengan menggunakan metode *Support Vector Machine (SVM)* dan *Random Forest* dengan penggunaan pengujian data *training* dan data *testing* sejumlah 75% : 25%.

