

SKRIPSI

**PREDIKSI WAKTU PERKULIAHAN YANG DIBUTUHKAN
MAHASISWA INFORMATIKA DENGAN MELIHAT NILAI
PADA BEBERAPA SEMESTER PERTAMA**



Fadel Muhammed Amien

NPM: 2013730030

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2020**

UNDERGRADUATE THESIS

**LECTURE TIME PREDICTION INFORMATICS STUDENTS
REQUIRE BASED ON FIRST SEMESTERS GRADES**



Fadel Muhammed Amien

NPM: 2013730030

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2020**

LEMBAR PENGESAHAN

PREDIKSI WAKTU PERKULIAHAN YANG DIBUTUHKAN MAHASISWA INFORMATIKA DENGAN MELIHAT NILAI PADA BEBERAPA SEMESTER PERTAMA

Fadel Muhammed Amien

NPM: 2013730030

Bandung, 16 Juni 2020

Menyetujui,

Pembimbing

Dr.rer.nat. Cecilia Esti Nugraheni

Ketua Tim Penguji

Anggota Tim Penguji

Luciana Abednego, M.T.

Mariskha Tri Adithia, P.D.Eng

Mengetahui,

Ketua Program Studi

Mariskha Tri Adithia, P.D.Eng

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

PREDIKSI WAKTU PERKULIAHAN YANG DIBUTUHKAN MAHASISWA INFORMATIKA DENGAN MELIHAT NILAI PADA BEBERAPA SEMESTER PERTAMA

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 16 Juni 2020

Fadel Muhammed Amien
NPM: 2013730030

ABSTRAK

Waktu kelulusan seorang mahasiswa Informatika Universitas Katolik Parahyangan mungkin dipengaruhi oleh beberapa nilai mata kuliah pada semester-semester awal. Hasil evaluasi dari nilai pada beberapa mata kuliah di semester-semester awal tersebut mungkin dapat memberikan gambaran yang cukup jelas berapa lama waktu yang dibutuhkan mahasiswa informatika untuk lulus. Untuk melakukan prediksi lama perkuliahan seorang mahasiswa informatika metode yang digunakan adalah dengan algoritma *Naive Bayes*. *Naive Bayes* merupakan salah satu algoritma klasifikasi statistik sederhana yang mudah untuk diimplementasikan serta memiliki waktu pemrosesan yang cepat bila diterapkan pada dataset yang besar, algoritma Naive Bayes juga memiliki tingkat efektivitas yang tinggi.

Algoritma *Naive Bayes* dapat diterapkan untuk memberikan gambaran prediksi mengenai berapa lama waktu kuliah seorang mahasiswa informatika dan merekomendasikan mata kuliah-mata kuliah yang memiliki pengaruh terhadap kelulusan seorang mahasiswa informatika dengan memperlihatkan atribut-atribut yang sesuai.

Pada skripsi ini akan dibangun perangkat lunak yang dapat mengimplementasikan algoritma *Naive Bayes*. Untuk dapat mengimplementasikan akan dicari terlebih dahulu korelasi antar mata kuliah agar dapat mendapatkan fitur mata kuliah yang memiliki pengaruh terhadap tahun kelulusan seorang mahasiswa.

Berdasarkan hasil pengujian yang dilakukan, perangkat lunak yang dibangun mampu melakukan prediksi lama perkuliahan mahasiswa Informatika UNPAR dengan menggunakan algoritma *Naive Bayes*

Kata-kata kunci: *Naive Bayes*, Prediksi, korelasi, lulus, klasifikasi

ABSTRACT

The time of graduation of an informatics student of Parahyangan Catholic University might be influenced by some course scores in the early semesters. The results of the evaluation of grades in several courses in the first semester may provide a fairly clear picture of the amount of time required for informatics students to graduate. To perform a long prediction of the lecture of an informatics student the method used is the Naive Bayes algorithm. Naive Bayes is one of the simple statistical classification algorithms that is easy to implement and has a fast processing time when applied to large datasets, the Naive Bayes algorithm also has a high level of effectiveness.

Naive Bayes Algorithm can be provide a predictive picture of the length of time a lecturer of informatics students and recommend courses that have an influence on the graduation of an Informatics student by showing the appropriate attributes.

In this undergraduate thesis, a software is developed to implement Naive Bayes algorithm. To be able to implement the correlation sought between subjects in advance in order to get the features of courses that have an influence on the graduation year of a student.

Based on the test done, it can be concluded that the software can implement Naive Bayes algorithm.

Keywords: Naive Bayes, Prediction, corellation, classifier

Dipersembahkan kepada keluarga, teman-teman IT 2013 dan diri sendiri

KATA PENGANTAR

Puji syukur kehadiran Allah S.W.T karena atas berkat dan rahmat-Nya penulis dapat menyelesaikan tugas dan tanggung jawab dalam menyusun skripsi yang berjudul "Prediksi Waktu Perkuliahan yang Dibutuhkan Mahasiswa Informatika dengan Melihat Nilai Pada Beberapa Semester Pertama". Penulis menyadari dalam penyusunan skripsi ini, penulis tidak luput dari bantuan beberapa teman dan kerabat, untuk itu penulis ingin mengucapkan terima kasih kepada :

- Keluarga yang selalu memberikan dukungan moral dan semangat terutama mamah yang selalu ada dalam suka dan duka sehingga penulis dapat menyelesaikan tanggung jawab sebagai seorang mahasiswa.
- Ibu Natalia selaku dosen pembimbing yang sudah banyak mengingatkan, memberikan arahan dan membimbing penulis dengan sangat sangat sabar.
- Ibu Mariskha dan Bu Luci atas saran dan masukan yang telah diberikan kepada penulis sebagai dosen penguji.
- Sipit yang sudah bersedia menyediakan waktu luangnya untuk menjadi mentor penulis dalam segala hal yang diperlukan untuk menyelesaikan skripsi ini.
- Fadhil yang juga telah menjadi mentor penulis sehingga penulis bisa lebih memahami hal yang diperlukan dalam pengerjaan skripsi ini.
- Vica yang selalu menjadi motivasi dan selamanya jadi penyemangat penulis untuk menyelesaikan skripsi dan yang telah membuat penulis menjadi lebih dewasa.
- Ilham yang selalu ada di sisi penulis teman seperjuangan dan semoga kita sukses di masa depan.
- Gavril, Jepus, Caca, Rahel, Glorya yang selalu membuat bahagia suasana hati penulis, walaupun kadang menyedihkan juga.
- Bima, Om, dan rekan-rekan kontrakan yang jadi teman seperjuangan penulis. Goodluck untuk kalian semua

Semoga bantuan dan dukungan yang telah kalian berikan kepada penulis dibalas oleh-Nya dengan berkali-kali lipat dalam hal lain yang kalian butuhkan. Akhir kata, penulis meminta maaf atas kekurangan dan kesalahan dalam penyusunan skripsi ini. Semoga karya yang penulis hasilkan ini dapat bermanfaat bagi semua pihak yang membaca dan memerlukan.

Bandung, Juni 2020

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xix
DAFTAR TABEL	xxi
DAFTAR KODE PROGRAM	xxiii
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Tujuan	2
1.4 Batasan Masalah	2
1.5 Metodologi	2
1.6 Sistematika Pembahasan	2
DAFTAR NOTASI	1
2 LANDASAN TEORI	5
2.1 <i>Predictive Analytics</i>	5
2.2 Teori Peluang	5
2.3 Peluang Bersyarat	6
2.3.1 Teorema Bayes	6
2.3.2 Naive Bayes Classifier	7
2.4 Korelasi Antara Dua Variabel atau Lebih	8
2.5 <i>Cross Validation</i>	10
2.6 <i>Mean Square Error</i>	10
3 ANALISIS	11
3.1 Analisis Penyelesaian Masalah	11
3.2 Analisis Data Set Nilai Mahasiswa Informatika	13
3.3 Analisis Pemilihan Variabel Mata Kuliah yang Akan Digunakan di <i>Naive Bayes</i>	13
3.4 Analisis Perangkat Lunak	15
3.4.1 Diagram Aktivitas	15
3.4.2 Diagram Kelas	16
3.4.3 Analisis Input Output	17
4 PERANCANGAN	19
4.1 Kebutuhan Masukan dan Keluaran	19
4.2 Rancangan Antarmuka	19
4.3 Diagram Kelas Rinci	20
4.4 Rincian Metode	22

4.4.1	Kelas GUIMainMenu	22
4.4.2	Kelas Prediction	22
4.4.3	Kelas NaiveBayes	22
4.4.4	Kelas Mahasiswa	25
4.4.5	Kelas ReadCSV	26
5	IMPLEMENTASI DAN PENGUJIAN	29
5.1	Implementasi Antarmuka	29
5.2	Pengujian Fungsional	31
5.3	Pengujian Ekperimental	32
5.3.1	Pengujian Nilai Mean Square Error dengan 5 Mata Kuliah	33
5.3.2	Pengujian Nilai Mean Square Error dengan 4 Mata Kuliah	34
5.3.3	Pengujian Nilai Mean Square Error dengan 3 Mata Kuliah	35
5.3.4	Pengujian Nilai Mean Square Error dengan 4 Mata Kuliah dengan Kombinasi Mata Kuliah	35
5.3.5	Kesimpulan Pengujian Eksperimental	36
6	KESIMPULAN DAN SARAN	37
6.1	Kesimpulan	37
6.2	Saran	37
	DAFTAR REFERENSI	39
	A KODE PROGRAM	41
	B HASIL EKSPERIMEN	51

DAFTAR GAMBAR

3.1	Diagram Aktivitas	15
3.2	Diagram Kelas Awal	16
4.1	Rancangan antarmuka proses prediksi	20
4.2	Diagram Kelas Rinci	21
4.3	Kelas Prediction	22
4.4	Kelas NaiveBayes	23
4.5	Kelas Mahasiswa	25
4.6	Kelas ReadCSV	26
5.1	Tampilan antarmuka pengguna	30
5.2	Tampilan antarmuka pemilihan dokumen	30
5.3	Tampilan hasil prediksi	32
5.4	Tabel perhitungan MSE iterasi 1	34
B.1	Tabel perhitungan MSE iterasi 1	51
B.2	Tabel perhitungan MSE iterasi 2	52
B.3	Tabel perhitungan MSE iterasi 3	52
B.4	Tabel perhitungan MSE iterasi 4	53
B.5	Tabel perhitungan MSE iterasi 5	53
B.6	Tabel perhitungan MSE iterasi 6	54
B.7	Tabel perhitungan MSE iterasi 7	54
B.8	Tabel perhitungan MSE iterasi 8	55
B.9	Tabel perhitungan MSE per iterasi 9	55
B.10	Tabel perhitungan MSE iterasi 10	56
B.11	Tabel perhitungan MSE iterasi 1	56
B.12	Tabel perhitungan MSE iterasi 2	57
B.13	Tabel perhitungan MSE iterasi 3	57
B.14	Tabel perhitungan MSE iterasi 4	58
B.15	Tabel perhitungan MSE iterasi 5	58
B.16	Tabel perhitungan MSE iterasi 6	59
B.17	Tabel perhitungan MSE iterasi 7	59
B.18	Tabel perhitungan MSE iterasi 8	60
B.19	Tabel perhitungan MSE per iterasi 9	60
B.20	Tabel perhitungan MSE iterasi 10	61
B.21	Tabel perhitungan MSE iterasi 1	61
B.22	Tabel perhitungan MSE iterasi 2	62
B.23	Tabel perhitungan MSE iterasi 3	62
B.24	Tabel perhitungan MSE iterasi 4	63
B.25	Tabel perhitungan MSE iterasi 5	63
B.26	Tabel perhitungan MSE iterasi 6	64
B.27	Tabel perhitungan MSE iterasi 7	64
B.28	Tabel perhitungan MSE iterasi 8	65

B.29 Tabel perhitungan MSE per iterasi 9	65
B.30 Tabel perhitungan MSE iterasi 10	66

DAFTAR TABEL

2.1	Data Set	7
2.2	Tabel <i>rating</i> untuk film	9
2.3	Skema 10 fold Cross Validation	10
3.1	Tabel data set <i>dummy</i>	11
3.2	Tabel sampling data korelasi	13
3.3	Tabel korelasi mata kuliah dengan tahun kelulusan	14
3.4	Tabel korelasi mata kuliah dengan mata kuliah	15
3.5	Analisis Input Output	17
5.1	Tabel data set pengujian fungsional	31
5.2	Tabel perhitungan <i>mean square error</i> dengan 5 mata kuliah	33
5.3	Tabel perhitungan <i>mean square error</i> dengan 4 mata kuliah	34
5.4	Tabel perhitungan <i>mean square error</i> dengan 3 mata kuliah	35
5.5	Tabel perhitungan <i>mean square error</i> dengan 4 mata kuliah dengan kombinasi mata kuliah	36
5.6	Tabel perbandingan hasil percobaan	36

DAFTAR KODE PROGRAM

A.1	readCSV.java	41
A.2	mahasiswa.java	43
A.3	NaiveBayes.java	43
A.4	main.java	45
A.5	GUIMainMenu.java	47

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Waktu kelulusan seorang mahasiswa Informatika UNPAR mungkin dipengaruhi oleh beberapa nilai mata kuliah pada semester-semester awal. Hasil evaluasi dari nilai pada beberapa mata kuliah di semester-semester awal tersebut mungkin dapat memberikan gambaran yang cukup jelas berapa lama waktu yang dibutuhkan mahasiswa informatika untuk lulus. Dengan mengetahui waktu prediksi kelulusan, mahasiswa Informatika UNPAR mungkin dapat menyusun strategi yang lebih baik untuk semester-semester berikutnya. Untuk melakukan prediksi lama perkuliahan seorang mahasiswa Informatika UNPAR metode yang digunakan adalah dengan algoritma *Naive Bayes*.

Dalam teori peluang terdapat teknik *Naive Bayes* di mana teknik ini berguna membangun klasifikasi [1]. Algoritma *Naive Bayes* akan memprediksi kelas *record* baru berdasarkan peluang terbesar untuk menjadi anggota dari klasifikasi yang ada berdasarkan data pelatihan. Algoritma *Naive Bayes* memiliki kekurangan yaitu setiap atribut yang ada saling bebas satu terhadap yang lainnya sehingga, membuat akurasi berkurang. Algoritma *Naive Bayes* juga memiliki kelebihan yaitu perhitungan yang cepat dan efisien, serta tidak membutuhkan data yang banyak untuk melakukan klasifikasi.

Algoritma *Naive Bayes* diharapkan dapat diterapkan untuk memberikan gambaran prediksi mengenai berapa lama waktu kuliah seorang mahasiswa informatika dan merekomendasikan mata kuliah-mata kuliah yang memiliki pengaruh terhadap kelulusan seorang mahasiswa informatika dengan memperlihatkan atribut-atribut yang sesuai, pada skripsi ini atributnya adalah nilai mata kuliah pada empat semester awal. Prediksi dapat dilakukan dengan berpedoman pada data kelulusan mahasiswa Informatika UNPAR, data mahasiswa *drop out* tidak termasuk ke dalam acuan dilakukannya prediksi karena beragamnya faktor mahasiswa mengalami *drop out* sehingga dapat membuat akurasi prediksi berkurang.

Pemilihan mata kuliah apa saja yang mungkin memiliki pengaruh terhadap waktu kelulusan seorang mahasiswa Informatika UNPAR, dengan begitu perlu dilakukan pemilihan mata kuliah-mata kuliah menggunakan uji korelasi yang nantinya akan digunakan untuk algoritma *Naive Bayes*. Pengujian perangkat lunak yang akan dibangun nanti akan dilakukan validasi berdasarkan data yang dimiliki untuk dapat mengetahui seberapa tepat perangkat lunak berjalan.

1.2 Rumusan Masalah

Berdasarkan deskripsi di atas, ditentukanlah rumusan masalah sebagai berikut :

- Bagaimana cara mengimplementasikan Algoritma *Naive Bayes* untuk memprediksi lama waktu perkuliahan?
- Bagaimana cara menentukan variabel apa saja yang akan dipakai untuk *Naive Bayes*?
- Bagaimana membangun perangkat lunak prediksi waktu perkuliahan yang dibutuhkan mahasiswa informatika dengan melihat nilai pada beberapa semester pertama?

1.3 Tujuan

Untuk menjawab rumusan masalah di atas, maka terbentuklah tujuan dari skripsi ini sebagai berikut :

- Mempelajari prediksi lama waktu perkuliahan yang dibutuhkan mahasiswa informatika untuk lulus berdasarkan nilai pada beberapa semester pertama.
- Menganalisis mata kuliah mana saja yang memiliki pengaruh pada lamanya waktu seorang mahasiswa informatika menyelesaikan perkuliahan.
- Membangun perangkat lunak untuk prediksi lama perkuliahan seorang mahasiswa informatika menggunakan klasifikasi dengan *Naive Bayes*.

1.4 Batasan Masalah

Batasan masalah dari skripsi ini adalah sebagai berikut :

- Prediksi tidak menangani kasus untuk mahasiswa *drop out*.
- Basis dilakukannya prediksi hanya dari nilai-nilai pada mata kuliah semester satu hingga semester empat saja.
- Dokumen yang diproses merupakan dokumen yang memiliki format *.csv* (*Comma Separated Values*).
- Ruang lingkup prediksi hanya pada lingkungan Informatika UNPAR saja.

1.5 Metodologi

Metodologi yang akan digunakan pada skripsi ini adalah sebagai berikut :

1. Melakukan studi literatur mengenai klasifikasi dengan *Naive Bayes*.
2. Melakukan studi literatur mengenai korelasi antara dua variabel atau lebih.
3. Mengolah data set nilai mahasiswa informatika.
4. Melakukan pemilihan variabel mata kuliah yang akan digunakan di *Naive Bayes*.
5. Membuat analisis penyelesaian masalah.
6. Menganalisis kebutuhan perangkat lunak.
7. Melakukan perancangan perangkat lunak serta tampilan perangkat lunak.
8. Melakukan implementasi perangkat lunak.
9. Melakukan pengujian terhadap perangkat lunak yang dibangun.
10. Menulis dokumen skripsi.

1.6 Sistematika Pembahasan

Skripsi ini terbagi menjadi enam bab yang terdiri dari pendahuluan, dasar teori, analisis, perancangan, implementasi dan pengujian, dan kesimpulan. Berikut merupakan sistematika pembahasan dalam skripsi ini.

1. Bab 1 Pendahuluan
Bab ini berisi latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi, dan sistematika pembahasan.
2. Bab 2 Dasar Teori
Bab ini berisi tentang dasar teori mengenai algoritma *Naive Bayes*, Korelasi Antara Dua Variabel atau Lebih, *Cross Validation*, dan *Mean Square Error*.
3. Bab 3 Analisis
Bab ini berisi tentang analisis data set nilai mahasiswa informatika, analisis pemilihan variabel mata kuliah yang akan digunakan di *naive bayes*, analisis masalah, diagram aktivitas, dan diagram kelas awal.

4. Bab 4 Perancangan

Bab 4 berisi perancangan perangkat lunak yang akan dibangun, meliputi kebutuhan masukan dan keluaran, rancangan antarmuka, diagram kelas rinci, dan rincian metode yang digunakan pada perangkat lunak.

5. Bab 5 Implementasi dan Pengujian

Bab 5 berisi implementasi antarmuka perangkat lunak, pengujian fungsional terhadap perangkat lunak yang mengimplementasikan prediksi lama perkuliahan dengan algoritma *naive bayes*, serta eksperimental terhadap perangkat lunak yang mengimplementasikan prediksi lama perkuliahan dengan algoritma *naive bayes*.

6. Bab 6 Kesimpulan dan Saran

Bab 6 berisi kesimpulan dari awal hingga akhir skripsi beserta saran untuk pengembangan selanjutnya.

