

**KLASIFIKASI KELULUSAN MAHASISWA MENGGUNAKAN
ALGORITME *HYBRID (K-NEAREST NEIGHBOR* DAN *NAÏVE*
*BAYES)***

Skripsi



Disusun Oleh :

Muhammad Fahmi As Shiddiqi

190602082

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH GRESIK
2023**

KATA PENGANTAR

Segala puji syukur ke hadirat Allah yang telah memberikan rahmat, nikmat dan kesehatan sehingga laporan skripsi yang berjudul ‘Klasifikasi Kelulusan Mahasiswa Menggunakan Algoritme *Hybrid (K-Nearest Neighbor Dan Naïve Bayes)*’ dapat dilaksanakan dengan cukup baik.

Adapun Laporan ini berisi tentang klasifikasi kelulusan mahasiswa dilihat dari perolehan nilai Indeks Prestasi Semester, Indeks Prestasi Kumulatif, Semester, Organisasi, dan kelas. Dengan mengimplementasikan Algoritme *K-Nearest Neighbor* dan Algoritme *Naïve Bayes* dalam melakukan klasifikasi.

Terselesaikannya laporan proposal skripsi ini dengan baik dapat terjadi berkat dukungan, motivasi, petunjuk, bimbingan dan do’a dari berbagai pihak. Oleh karena itu, ucapan terima kasih yang sebesar-besarnya penulis sampaikan kepada:

1. Allah SWT atas segala nikmat, ridho dan kuasanya.
2. Nabi Muhammad SAW sebagai panutan teladan dalam menjadi insan manusia yang baik dan benar.
3. Keluarga. Bapak Machsun. Ibu Hidayatul Annisak, Kakak Muh. Islahul Azmi, dan Adik Akhrianil Hami Maisuroyya dan Rizqina Kautsarina yang selalu terus mendo’akan, memberikan semangat, dan terus mendukung penuh hingga detik ini.
4. Bapak Harunur Rosyid, St, M.Kom. selaku Dekan Fakultas Teknik Universitas Muhammadiyah Gresik.
5. Ibu Henny Dwi Bhakti, S.Si., M.Si. selaku Ketua Program Studi Teknik Informatika Universitas Muhammadiyah Gresik dan Dosen Pembimbing yang senantiasa meluangkan waktunya untuk memberikan bimbingan, arahan serta masukan bagi penulis dalam penyelesaian proposal skripsi ini.
6. Ibu Umi Chotijah, S.Kom., M.Kom. selaku Dosen Wali Mahasiswa Teknik Informatika Universitas Muhammadiyah Gresik Angkatan 2019.
7. Bapak dan Ibu Dosen Program Studi Teknik Informatika yang telah memberikan segenap ilmu pengetahuan kepada penulis.

8. Rekan-rekan seperjuangan mahasiswa teknik informatika angkatan 2019 dan Himpunan Mahasiswa Teknik Informatika yang telah menjadi rumah dan keluarga baru yang selalu bersemangat dan berjuang dalam mewujudkan visi-misi bersama selama menjadi mahasiswa aktif Teknik Informatika Universitas Muhammadiyah Gresik.
9. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah memberikan bantuan dan dukungan serta do'a selama menjalani perkuliahan.
Permohonan maaf yang sebesar-besarnya jika dalam pengembangan program serta penulisan laporan proposal skripsi ini terdapat kekurangan dan keterlambatan baik yang Saya sengaja maupun tidak Saya sengaja. Sehingga saran dan kritik yang sifatnya membangun dari semua pihak berguna bagi penulis untuk memperbaiki skripsi ini. Sejatinya tidak ada yang sempurna di muka Bumi ini, sehingga dengan belajarnya melalui kesalahan-kesalahan yang didahului semoga menjadikannya esok yang lebih baik lagi. Aamiin Ya Rabbal 'Alamin.

Gresik, 6 Januari 2022

Penulis

ABSTRAK

Kelulusan merupakan tujuan akhir bagi setiap orang dalam rangkaian proses menyelesaikan studi dalam memperoleh gelar, penghargaan ataupun yang lainnya. Pada Program Studi Teknik Informatika Universitas Muhammadiyah Gresik sendiri terdapat banyak mahasiswa yang lulus tidak tepat waktu (terlambat) dan mahasiswa yang *Drop Out* (DO), hal tersebut merupakan masalah penting bagi Program Studi, dikarenakan jumlah total dari mahasiswa yang mendaftar tidak sebanding dengan jumlah total mahasiswa yang lulus akan berdampak pada penilaian Akreditasi dimana akan berdampak pada menurunnya calon mahasiswa baru yang akan mendaftar di Program Studi Teknik Informatika Universitas Muhammadiyah Gresik. Sehingga sangat penting bagi Program Studi mengetahui klasifikasi kelulusan mahasiswa. Informasi tersebut dapat menjadi dasar dalam membuat rencana strategis dalam rangkaian pembelajaran supaya meningkatkan jumlah mahasiswa dengan masa studi lulus tepat waktu. Dalam penelitian ini akan dilakukan klasifikasi kelulusan mahasiswa Program Studi Teknik Informatika Universitas Muhammadiyah Gresik menggunakan algoritme *K-Nearest Neighbor* dan *Naïve Bayes*. Hasil pengujian evaluasi performa menggunakan *Confusion Matrix* didapatkan nilai rata-rata *precision* dan *recall* dari kedua algoritme. Algoritme *K-Nearest Neighbor* mendapatkan rata-rata nilai *precision*, dan *recall* lebih tinggi dari algoritme *Naïve Bayes* yaitu sebesar 85,1% dan 92,9%. Sedangkan untuk algoritme *Naïve Bayes* mendapatkan rata-rata nilai *precision* dan *recall* lebih rendah yaitu sebesar 73% dan 73,5%. Sementara hasil pengujian menggunakan *K-Fold Cross Validation* mendapatkan hasil bahwa algoritme *K-Nearest Neighbor* dinilai lebih baik dari algoritme *Naïve Bayes* dengan melihat nilai hasil dari *accuracy*, *precision*, dan *recall* algoritme *K-Nearest Neighbor* mendapatkan hasil yang lebih baik dengan nilai *accuracy* sebesar 84,2% nilai *precision* sebesar 92,3% dan nilai *recall* sebesar 82,6%.

Kata Kunci : Klasifikasi, Kelulusan Mahasiswa, *K-Nearest Neighbor*, *Naïve Bayes*, *Confusion Matrix*, *K-Fold Cross Validation*..

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN	iii
KATA PENGANTAR	iv
LEMBAR PERNYATAAN KEASLIAN SKRIPSI	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xi
DAFTAR PERSAMAAN	xiv
ABSTRAK	xv
ABSTRACK	xvi
BAB 1 PENDAHULUAN	1
1.1. LATAR BELAKANG.....	1
1.2. RUMUSAN MASALAH	3
1.3. TUJUAN PENELITIAN	4
1.4. BATASAN MASALAH	4
1.5. MANFAAT PENELITIAN	5
1.6. METODOLOGI PENELITIAN	5
1.7. SISTEMATIKA PENULISAN	6
BAB 2 TINJAUAN PUSTAKA	8
2.1. <i>DATA MINING</i>	8
2.2. KLASIFIKASI	9
2.3. ALGORITME <i>K-NEAREST NEIGHBOR</i>	10
2.4. ALGORITME <i>NAÏVE BAYES</i>	11
2.5. AKURASI	11
2.6. PENELITIAN TERKAIT.....	12

BAB 3 ANALISIS DAN PERANCANGAN SISTEM	18
3.1. ANALISIS SISTEM.....	18
3.1.1. Representasi Data	18
3.1.2. <i>Preprocessing</i> Data.....	22
3.2. PERANCANGAN SISTEM.....	25
3.2.1. <i>Flowchart</i> Sistem Klasifikasi Kelulusan Mahasiswa Universitas Muhammadiyah Gresik	25
3.2.2. <i>Flowchart</i> Algoritme <i>K-Nearest Neighbor</i>	26
3.2.3. <i>Flowchart</i> Algoritme <i>Naïve Bayes</i>	27
3.2.4. Diagram Konteks	28
3.2.5. Diagram Berjenjang.....	28
3.2.6. Data Flow Diagram.....	29
3.3. PERANCANGAN ANTARMUKA SISTEM.....	31
3.4. PERANCANGAN PENGUJIAN	33
3.4.1. Perhitungan Algoritme <i>K-Nearest Neighbor</i>	33
3.4.2. Perhitungan Algoritme <i>Naïve Bayes</i>	36
3.4.3. Evaluasi Performa <i>Confusion Matrix</i>	48
3.4.4. Pengujian <i>Black Box</i>	50
BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM.....	51
4.1. IMPLEMENTASI SISTEM	51
4.1.1 Halaman Login	51
4.1.2 Halaman Pilih Metode.....	51
4.1.3 Halaman Data Training	52
4.1.4 Halaman Data Testing	52
4.1.5 Halaman Klasifikasi	53
4.1. PENGUJIAN SISTEM.....	54
4.2.1. Pengujian <i>Confusion Matrix</i> Algoritme <i>K-Nearest Neighbor</i>	56
4.2.2. Pengujian <i>Confusion Matrix</i> Algoritme <i>Naïve Bayes</i>	58
4.2.3. Perbandingan Hasil Pengujian <i>Confusion Matrix</i>	60

4.2.4. Pengujian <i>K-Fold Cross Validation</i>	61
4.2.5. Perbandingan Hasil Pengujian <i>K-Fold Cross Validation</i> ..	62
4.2.6. Pengujian <i>Black Box</i>	63
BAB 5 PENUTUP	65
5.1 KESIMPULAN	65
5.2 SARAN.....	65
DAFTAR PUSTAKA	67
LAMPIRAN.....	70

DAFTAR TABEL

Tabel 3.1 Atribut Data (Input).....	19
Tabel 3.2 Atribut Data (Output)	20
Tabel 3.3 Data Training.....	20
Tabel 3.4 Data Testing	21
Tabel 3.5 Tipe Atribut	22
Tabel 3.6 Kategorial Indeks Prestasi Semester	22
Tabel 3.7 Kategorial Indeks Prestasi Kumulatif.....	23
Tabel 3.8 Kategorial MKP, KKN, dan KP	23
Tabel 3.9 Hasil <i>Preprocessing</i> Data Training	23
Tabel 3.10 Hasil <i>Preprocessing</i> Data Testing.....	24
Tabel 3.11 Data Training Perhitungan KNN.....	33
Tabel 3.12 Data Testing Perhitungan KNN	33
Tabel 3.13 Hasil Jarak <i>Eculidean</i> Data Testing	34
Tabel 3.14 Pengurutan Peringkat Nilai Jarak <i>Euclidean</i>	35
Tabel 3.15 Klasifikasi KNN $K = 3$	35
Tabel 3.16 Hasil Klasifikasi KNN $K = 3$	36
Tabel 3.17 Jumlah Data Fitur IPS 1	37
Tabel 3.18 Probabilitas Fitur IPS 1	37
Tabel 3.19 Jumlah Data Fitur IPS 2	37
Tabel 3.20 Probabilitas Fitur IPS 2	38
Tabel 3.21 Jumlah Data Fitur IPS 3	38
Tabel 3.22 Probabilitas Fitur IPS 3	38
Tabel 3.23 Jumlah Data Fitur IPS 4	38
Tabel 3.24 Probabilitas Fitur IPS 4	39
Tabel 3.25 Jumlah Data Fitur IPS 5	39
Tabel 3.26 Probabilitas Fitur IPS 5	39
Tabel 3.27 Jumlah Data Fitur IPS 6	39
Tabel 3.28 Probabilitas Fitur IPS 6	40
Tabel 3.29 Jumlah Data Fitur IPK.....	40

Tabel 3.30 Probabilitas Fitur IPK.....	40
Tabel 3.31 Jumlah Data Fitur MKP	40
Tabel 3.32 Probabilitas Fitur MKP	40
Tabel 3.33 Jumlah Data Fitur KKN	41
Tabel 3.34 Probabilitas Fitur KKN	41
Tabel 3.35 Jumlah Data Fitur KP	41
Tabel 3.36 Probabilitas Fitur KP	41
Tabel 3.37 Tabel Data Training ke-1	41
Tabel 3.38 Hasil Klasifikasi Data Training ke-1.....	42
Tabel 3.39 Hasil Klasifikasi Data Training.....	42
Tabel 3.40 Jumlah Data Akhir Fitur IPS 1	43
Tabel 3.41 Nilai Probabilitas Akhir Fitur IPS 1	43
Tabel 3.42 Jumlah Data Akhir Fitur IPS 2.....	44
Tabel 3.43 Nilai Probabilitas Akhir Fitur IPS 2.....	44
Tabel 3.44 Jumlah Data Akhir Fitur IPS 3.....	44
Tabel 3.45 Nilai Probabilitas Akhir Fitur IPS 3.....	44
Tabel 3.46 Jumlah Data Akhir Fitur IPS 4.....	45
Tabel 3.47 Nilai Probabilitas Akhir Fitur IPS 4.....	45
Tabel 3.48 Jumlah Data Akhir Fitur IPS 5.....	45
Tabel 3.49 Nilai Probabilitas Akhir Fitur IPS 5.....	45
Tabel 3.50 Jumlah Data Akhir Fitur IPS 6.....	46
Tabel 3.51 Nilai Probabilitas Akhir Fitur IPS 6.....	46
Tabel 3.52 Jumlah Data Akhir Fitur IPK	46
Tabel 3.53 Nilai Probabilitas Akhir Fitur IPK	46
Tabel 3.54 Jumlah Data Akhir Fitur MKP	47
Tabel 3.55 Nilai Probabilitas Akhir Fitur MKP	47
Tabel 3.56 Jumlah Data Akhir Fitur KKN	47
Tabel 3.57 Nilai Probabilitas Akhir Fitur KKN	47
Tabel 3.58 Jumlah Data Akhir Fitur KP.....	47
Tabel 3.59 Nilai Probabilitas Akhir Fitur KP.....	47
Tabel 3.60 Hasil Pengujian Klasifikasi Data Testing Algoritme <i>Naïve Bayes</i>	48

Tabel 3.61 Evaluasi Performa <i>Confusion Matrix</i>	49
Tabel 3.62 Pengujian <i>Black Box</i>	50
Tabel 4.1 Hasil Pengujian Data Testing	54
Tabel 4.2 Hasil Pengujian <i>Confusion Matrix</i> Algoritme <i>K-Nearest Neighbor</i>	56
Tabel 4.3 Hasil Evaluasi <i>Confusion Matrix</i> Algoritme <i>K-Nearest Neighbor</i>	57
Tabel 4.4 Hasil Pengujian <i>Confusion Matrix</i> Algoritme <i>Naïve Bayes</i>	58
Tabel 4.5 Hasil Evaluasi <i>Confusion Matrix</i> Algoritme <i>Naïve Bayes</i>	59
Tabel 4.6 Perbandingan Hasil Evaluasi Performa <i>Confusion Matrix</i>	60
Tabel 4.7 Hasil Pengujian <i>K-Fold Cross Validation</i> Algoritme <i>K-Nearest Neighbor</i>	61
Tabel 4.8 Hasil Pengujian <i>K-Fold Cross Validation</i> Algoritme <i>Naïve Bayes</i>	61
Tabel 4.9 Perbandingan Hasil Rata-Rata Pengujian <i>K-Fold Cross Validation</i>	62
Tabel 4.10 Hasil Pengujian <i>Black Box</i>	63

DAFTAR GAMBAR

Gambar 2.1. Model klasifikasi	9
Gambar 3.1 <i>Flowchart</i> Sistem Klasifikasi	25
Gambar 3.2 <i>Flowchart</i> Algoritme <i>K-Nearest Neighbor</i>	26
Gambar 3.3 <i>Flowchart</i> Algoritme <i>Naïve Bayes</i>	27
Gambar 3.4 Diagram Konteks Sistem Klasifikasi.....	28
Gambar 3.5 Diagram Berjenjang Sistem Klasifikasi	28
Gambar 3.6 <i>Data Flow Diagram</i> Level 0	29
Gambar 3.7 <i>Data Flow Diagram</i> Level 1 <i>K-Nearest Neighbor</i>	30
Gambar 3.8 <i>Data Flow Diagram</i> Level 1 <i>Naïve Bayes</i>	30
Gambar 3.9 Halaman <i>Login</i>	31
Gambar 3.10 Halaman Data Training.....	32
Gambar 3.11 Halaman Data Testing	32
Gambar 3.12 Halaman Hasil Klasifikasi	33
Gambar 4.1 Halaman Login	51
Gambar 4.2 Halaman Pilih Metode	52
Gambar 4.3 Halaman Data Training.....	52
Gambar 4.4 Halaman Data Testing	53
Gambar 4.5 Halaman Klasifikasi.....	53
Gambar 4.6 Grafik Perbandingan Hasil Pengujian <i>Confusion Matrix</i>	60
Gambar 4.7 Grafik Perbandingan Hasil Pengujian <i>K-Fold Cross Validation</i>	62