

PREDIKSI KETEPATAN WAKTU STUDI MAHASISWA BIDIK MISI MENGGUNAKAN K-NEAREST NEIGHBOUR

Dedy Priyambodo¹⁾, Aryo Nugroho²⁾, dan Badrus Zaman³⁾

^{1, 2)}Program Studi Sistem Informasi, Fakultas Ilmu Komputer,
Universitas Narotama, Surabaya

³⁾ Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Airlangga, Surabaya

e-mail: dedy.priyambodo@fik.narotama.ac.id¹⁾, aryo.nugroho@narotama.ac.id²⁾,
badruszaman@fst.unair.ac.id³⁾

ABSTRAK

Program Beasiswa Bidikmisi adalah program bantuan Kementerian Pendidikan dan Kebudayaan melalui Direktorat Jendral Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan mulai tahun 2010. Program tersebut ditujukan bagi mahasiswa yang memiliki perkonomian kurang mampu, namun berpotensi untuk melanjutkan ke jenjang yang lebih tinggi (pendidikan tinggi). Dilihat dari jumlah mahasiswa bidikmisi yang diterima di perguruan tinggi dan data histori pendidikan mahasiswa didapatkan data yang dapat memprediksi dan menentukan pola kelulusan mahasiswa penerima bidikmisi pada sebuah perguruan tinggi sehingga dapat dijadikan acuan kemungkinan ketepatan waktu kuliah mahasiswa penerima bidikmisi. Tujuan dari penelitian ini adalah untuk memprediksi ketepatan lulus mahasiswa penerima beasiswa bidikmisi jenjang sarjana pada sebuah universitas dengan menggunakan atribut jenis kelamin, nilai TOEFL, indeks prestasi semester 1 hingga semester 4, rata-rata nilai ijazah SMA, rata-rata nilai ujian nasional, program studi, nilai prestasi keaktifan organisasi. Uji coba dilakukan dengan percentage split data training dan evaluasi 60-40, 70-30, 80-20, dan 90-10, dengan variasi nilai k sebesar 3, 5, 7, 9, 11, 13, dan 15. Hasil dari uji coba didapatkan nilai akurasi tertinggi sebesar 93.93% pada percentage split 70-30 dan nilai k=15.

Kata Kunci: Data mining, Prediksi, Kelulusan Mahasiswa, K-Nearest Neighbour

ABSTRACT

The Bidikmisi Scholarship Program is a program of assistance from the Ministry of Education and Culture through the Directorate General of Higher Education of the Ministry of Education and Culture starting in 2010. The program is intended for students who are economically disadvantaged but have the potential to continue to a higher level (higher education). Judging from the number of Bidikmisi students accepted at tertiary institutions and historical data on student education, data is obtained that can predict and determine the graduation pattern of Bidikmisi recipient students at a university so that it can be used as a reference for the possible timeliness of Bidikmisi recipient students' lectures. The purpose of this study was to predict the accuracy of passing students receiving Bidikmisi scholarships at the undergraduate level at a university by using gender, TOEFL scores, the semester achievement index attributes from semesters 1 to 4, average rating of senior high school, grade national average, study programs, organizational activity achievement scores. The trial was conducted with a percentage split of training data and evaluation of 60-40, 70-30, 80-20, and 90-10, with variations in k values of 3, 5, 7, 9, 11, 13, and 15. The results of the test try to get the highest accuracy value of 93.93% on a percentage split of 70-30 and the value of k = 15.

Keywords: Data mining, Classification, Student Graduation, K-Nearest Neighbour

1. PENDAHULUAN

Program Beasiswa Bidikmisi adalah program bantuan Kementerian Pendidikan dan Kebudayaan melalui Direktorat Jendral Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan mulai tahun 2010 [1]. Program tersebut ditujukan bagi mahasiswa yang memiliki perkonomian kurang mampu, namun berpotensi untuk melanjutkan ke jenjang yang lebih tinggi (pendidikan tinggi) [2]. Pendidikan tinggi adalah jenjang pendidikan pada jalur pendidikan formal setelah pendidikan menengah yang dapat berupa program pendidikan diploma, sarjana, magister, spesialis, dan doktor, yang diselenggarakan oleh perguruan tinggi [3]. Program sarjana diarahkan pada hasil lulusan yang menguasai dasar-dasar ilmiah dan keterampilan dalam bidang keahlian tertentu sehingga mampu menemukan, memahami, menjelaskan dan merumuskan cara penyelesaian masalah yang ada di kawasan keahliannya. Dilihat dari banyaknya peminat dari tiap tahun akademik perguruan tinggi dapat memperhatikan berbagai faktor yang berpengaruh terhadap pengelolaan daya tampung mahasiswa, salah satunya ketepatan dari ma-sa studi mahasiswa sesuai dengan waktu yang telah ditentukan. Masa studi adalah waktu studi terjadwal yang harus ditempuh oleh mahasiswa sesuai dengan rentang waktu yang dipersyaratkan.

Data mahasiswa bidikmisi dimungkinkan untuk menghasilkan sebuah data yang dapat diolah sehingga dapat dilakukan prediksi ketepatan lulus mahasiswa jalur bidikmisi. Prediksi yang dilakukan dapat dijadikan acuan untuk memberikan perlakuan khusus untuk mahasiswa sehingga mahasiswa tersebut dapat lulus tepat waktu.

Untuk mendapatkan prediksi ketepatan waktu lulus mahasiswa dibutuhkan sebuah metode pengolahan data yang telah tersedia. Hasil dari pengelolaan pola dapat dianalisis dengan menggunakan konsep data mining. Data mining, sering juga disebut Knowledge Discovery In Database (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar [4].

Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data. Terdapat dua metode dalam pengolahan data. Metode tersebut

adalah metode klasifikasi dan klusterisasi. Klasifikasi adalah suatu cara dalam Data Mining untuk mengklasifikasikan data yang akan kita analisis. Data yang diklasifikasikan dengan pola terawasi selalu memiliki label di setiap kelasnya [5]. Metode Klasifikasi mampu mengolah dataset dengan pelatihan data yang sangat cepat, efektif dan memiliki nilai komputasi yang kompleks adalah k Nearest Neighbor. k Nearest Neighbor adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Percentage split merupakan teknik mengevaluasi dari hasil training data dimana cara ini akan menemukan akurasi terbaik dengan memanfaatkan angka dalam persentasenya [4].

KNN merupakan metode untuk mengklasifikasi dokumen dengan menentukan nilai K-nearest neighbor untuk menentukan label dokumen. K-Nearest Neighbor (KNN) adalah metode model ruang vektor yang berfungsi untuk mengklasifikasikan suatu objek [6]. Pada dasarnya, klasifikasi KNN akan didasarkan pada hipotesis dimana dokumen uji akan memiliki kategori atau label yang sama dengan kategori dokumen pelatihan yang bermain dalam lingkup yang melingkupi dokumen uji [7]. Jadi KNN harus mencocokkan seluruh dokumen untuk menguji dokumen satu per satu, KNN memiliki kelemahan dalam penggunaan waktu proses pengklasifikasian.

Penelitian sebelumnya pernah dilakukan oleh Aryo Nugroho pada tahun 2021 dengan judul A Decision Guidance for Solving Success Rate Political Campaign Using Distance Weighted kNN in Nassi-Shneiderman Framework [8]. Data diolah dengan pemodelan yang dibangun menggunakan machine learning dengan metode klasifikasi k Nearest Neighbor (kNN) dan Distance Weighted kNN (kNN). Hasil penelitian ini menunjukkan bahwa DWkNN pengembangan model telah berhasil memberikan akurasi yang lebih baik. Hasil ini dapat dicapai dalam tahap pelatihan dan pengujian. Pilihan $k = 2$ adalah dilakukan dengan prinsip kehati-hatian untuk menghindari kemungkinan overfitting saat menggunakan $k = 1$. Namun, $k = 2$ menunjukkan hasil yang baik untuk keseluruhan. Himpunan data. Tahapan pengujian menggunakan cross-validation menunjukkan hasil akurasi yang baik di

berbagai n-fold skenario antara 5 dan 20, dan dapatkan hasil terbaik pada 20 kali lipat.

Penelitian selanjutnya pernah dilakukan oleh Novita priandini pada tahun 2017 dengan judul *Categorizing document by fuzzy C-Means and K-nearest neighbors approach* [9]. Dalam penelitian ini mengelola beberapa dokumen dengan mengkategorikan merupakan salah satu aplikasi Information Retrieval, karena melibatkan text mining dalam prosesnya. Sedangkan teknik kategorisasi dapat dilakukan baik dengan metode Fuzzy C-Means (FCM) maupun K Nearest Neighbors (KNN). Hasil penelitian ini adalah pengelompokan dokumen pelatihan, dihasilkan empat cluster dengan rincian 57 dokumen pada cluster 1, 26 dokumen pada cluster 2, 20 dokumen pada cluster 3, dan 57 dokumen pada cluster 4. Dari 20 dokumen pengujian yang diuji, 14 dokumen tidak memiliki relevansi. ke kategori sebenarnya. Sedangkan 6 dokumen tidak memiliki relevansi dengan kategori sebenarnya. Hasil evaluasi sistem menunjukkan nilai sistem Precision Recall sebesar 0,7. Menggabungkan dua metode menghasilkan F-measures 0,7. Dengan nilai-nilai tersebut sudah cukup membuat sistem memiliki ukuran keberhasilan yang cukup baik.

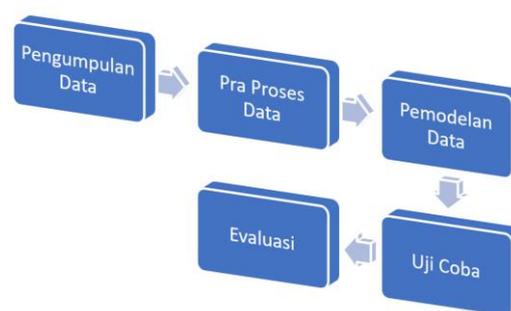
Penelitian selanjutnya dilakukan oleh Niki Dwi Ayuningtyas pada tahun 2020 dengan judul *Problematika Mahasiswa Bidikmisi Prodi Pendidikan Sejarah Angkatan 2013—2017* [10]. Penelitian ini berfokus untuk mendeskripsikan mengenai problematika yang dihadapi mahasiswa bidikmisi selama masa studi. Metode dalam penelitian ini menggunakan deskriptif kualitatif dengan pendekatan studi kasus. Pengumpulan data dalam penelitian ini menggunakan observasi, wawancara, dan dokumentasi. Hasil penelitian menunjukkan bahwa setiap mahasiswa memiliki masalah yang berbeda dengan mahasiswa lainnya dalam menempuh masa studi. Untuk kelulusan, jika dibandingkan lebih banyak mahasiswa bidikmisi yang tidak tepat waktu.

Penelitian ini bertujuan untuk memprediksi ketepatan waktu lulus dari mahasiswa bidikmisi jenjang sarjana pada sebuah universitas. Kemudian, dari hasil klasifikasi ketepatan kelulusan mahasiswa dapat dijadikan sebagai bahan evaluasi untuk percepatan kelulusan mahasiswa bisikmisi jenjang sarjana.

2. METODE PENELITIAN

Penelitian ini adalah eksperimen untuk menentukan klasifikasi ketepatan waktu kelulusan

mahasiswa bidikmisi jenjang sarjana berdasarkan atribut jenis kelamin, rata-rata nilai ijazah, rata-rata nilai UN, nilai toefl, nilai prestasi organisasi, program studi mahasiswa dan indek prestasi semester 1 sampai dengan 4 mahasiswa. Tahap penelitian menggunakan teknik data mining dengan metode K-Nearest Neighbor. Sedangkan, untuk pengujian nilai akurasi dari metode menggunakan presentase split untuk mencari nilai optimal dari metode yang digunakan. Secara garis besar penelitian dilakukan dengan 5 tahap, yaitu pengumpulan data, pra proses data, pemodelan data, uji coba, dan evaluasi [11]. Tahapan ini dapat dilihat pada Gambar 1.



Gambar. 1. Tahap Penelitian

A. Pengumpulan Data

Pengumpulan data adalah prosedur yang sistematis dan standar untuk memperoleh data yang diperlukan. Pengumpulan data dilakukan dengan cara observasi dokumen dan laporan yang ada pada sebuah perguruan tinggi negeri di kota Surabaya. Data yang dikumpulkan adalah data histori pendidikan mahasiswa bidikmisi jenjang sarjana mulai tahun 2013 sampai dengan tahun 2017.

B. Pra Proses Data

Tahap pra proses data adalah tahap mentransformasikan data mentah kedalam format data yang sesuai dengan kebutuhan analisis dalam penelitian. Proses ini bertujuan menyiapkan dan mengolah data dengan teknik-teknik tertentu sehingga terbentuk data yang siap diolah pada proses selanjutnya. Teknik yang dilakukan dalam proses ini diantaranya adalah:

1) Data Validation

Data validation, untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*). *Missing data* terlihat.

2) *Data integration and Transformation*

Data integration and Transformation, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal.

3) *Data size reduction and dicrtization*

Data size reduction and dicrtization, untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informatif.

C. *Pemodelan Data*

Tahap pemodelan data adalah pengolahan data dengan metode K-Nearest Neighbor sesuai dengan tahapan yang diperlukan. Dalam tahap ini dilakukan pemrosesan data sebanyak 1593 data. Data tersebut adalah data dari hasil pra proses data. Data yang diperoleh selanjutnya diolah sesuai dengan langkah-langkah pada metode K-Nearest Neighbor untuk menghasilkan sebuah model.

Algoritma K-Nearest Neighbor merupakan metode klasifikasi yang mengelompokan data baru berdasarkan jarak data baru itu kebeberapa data/tetangga (neighbor) terdekat [12]. Teknik K-Nearest Neighbor dengan melakukan langkah-langkah yaitu, mulai input: Data training, label data traning, k, data testing.

- a. Untuk semua data testing, hitung jaraknya ke setiap data training
- b. Tentukan k data training yang jaraknyapaling dekat dengan data
- c. Testing

| Tahun Angkatan Mahasiswa | Laki Laki | | Perempuan | |
|--------------------------|-------------|------------|-----------|-------------|
| | MOLOR | TEPAT | MOLOR | TEPAT |
| 2013 | 24 | 33 | 19 | 159 |
| 2014 | 15 | 66 | 18 | 316 |
| 2015 | 5 | 78 | 9 | 346 |
| 2016 | 0 | 99 | 0 | 319 |
| 2017 | 0 | 25 | 0 | 62 |
| JUMLAH TOTAL | 44 | 301 | 46 | 1202 |
| TOTAL | 1593 | | | |

- d. Periksa label dari k data ini
- e. Tentukan label yang frekuensinya palingbanyak
- f. Masukkan data testing ke kelas dengan frekuensi paling banyak

Untuk menghitung jarak antara dua titik x dan y bisa digunakan jarak Euclidean seperti pada persamaan 1.

$$d(X_1, X_2) = \sum_l \left| \frac{n_{1i}}{n_1} - \frac{n_{2i}}{n_2} \right| \dots \text{(persamaan 1)}$$

Yang mana $X_1, 1 =$ adalah attribute kategori, dan n_{1j}, n_1 mewakili frekuensi yang sesuai.

D. *Uji Coba*

Tahap ujicoba adalah tahap dimana dilakukan ujicoba dari model yang sudah diperoleh menggunakan beberapa skenario untuk mencari nilai optimal dalam penelitian. Ujicoba dilakukan dengan menggunakan presentage split. Dengan pengujian tersebut akan dihasilkan nilai persentase ideal dalam melakukan pengujian data dengan metode K-Nearest Neighbor.

E. *Evaluasi*

Tahapan terakhir dalam penelitian ini adalah tahap evaluasi. Evaluasi dilakukan dengan menganalisa hasil dari pengujian menggunakan *percentage split* dengan mencari nilai optimal sehingga dapat disimpulkan nilai pengujian terbaik dalam proses uji coba yang dilakukan.

3. **HASIL DAN PEMBAHASAN**

F. *Pengumpulan Data*

Data yang digunakan untuk penelitian ini adalah da-ta mahasiswa bidikmisi pada tahun 2013 hingga 2017 sebanyak 1593 data dengan rincian laki-laki sebanyak 345 orang (21.65%), sedangkan perempuan sebanyak 1.248 orang (78,34%). Data ini memiliki atribut nim, jenis kelamin, nilai toefl, Indeks Prestasi Semester 1 hingga semester 4, rata-rata nilai ijazah SMA, rata-rata nilai UN, program studi, nilai prestasi keaktifan organisasi, tanggal masuk mahasiwa, dan tanggal lulus mahasiswa. Dari data yang didapatkan jumlah mahasiswa yang lulus tepat waktu sebanyak 1.503 orang atau 94.35%, sedangkan yang molor sebanyak 90 orang atau 5.65%. Rekapitulasi data yang diperoleh ditunjukkan pada Tabel 1.

TABEL 1
Rekapitulasi Hasil Perolahan Data

G. *Pra Proses Data*

Pra proses data dilakukan untuk mendapatkan data yang siap untuk diproses pada langkah selanjutnya. Pengolahan data menggunakan teknik klasifikasi untuk mendapatkan data sesuai dengan kebutuhan pengolahan data. Dimulai dari eliminasi atribut yang tidak digunakan seperti nim

mahasiswa, tanggal masuk, dan tanggal lulus mahasiswa.

Sebelum dilakukan eliminasi pada tanggal masuk dan lulus mahasiswa dilakukan transformasi data yaitu pengolahan data lama studi mahasiswa dengan menghitung lama studi mahasiswa menggunakan atribut tanggal masuk dan tanggal keluar. Selanjutnya melakukan pengkodean pada program studi mahasiswa, jenis kelamin mahasiswa dan memberikan label kategori kelulusan mahasiswa yaitu mahasiswa lulus tepat waktu dan lulus melebihi batas waktu studi (terlambat). Setelah dilakukan proses tersebut data siap diolah pada proses selanjutnya. Contoh sebagian data yang sudah siap diolah dapat dilihat pada Tabel 2. Data-data ini memiliki 11 atribut, yaitu Kode Prodi, Jenis Kelamin, Nilai Organisasi, Nilai Toefl, Rata Rata Nilai Ijazah, Rata Rata Nilai UN, IPS 1, IPS 2, IPS 3, IPS 4, dan Class.

TABEL 2
DATASET SIAP OLAH MAHASISWA BIDIKMISI

| Kode Prodi | Jenis Kelamin | Nilai Organisasi | Nilai Toefl | Rata Rata Nilai Ijazah | Rata Rata Nilai UN | IPS 1 | IPS 2 | IPS 3 | IPS 4 | Class |
|------------|---------------|------------------|-------------|------------------------|--------------------|-------|-------|-------|-------|-------|
| 7 | 1 | 100 | 360 | 88 | 75.2 | 2.6 | 3.15 | 3.14 | 3.05 | molor |
| 7 | 1 | 105 | 350 | 79.7 | 75.9 | 2.65 | 3.08 | 2.7 | 1.63 | molor |
| 4 | 2 | 223 | 347 | 84 | 84.8 | 2.17 | 2.66 | 2.64 | 2 | TEPAT |
| 4 | 2 | 174 | 477 | 87.5 | 86.3 | 3.31 | 3.18 | 3.38 | 3.44 | TEPAT |
| 20 | 1 | 168 | 377 | 83 | 80.7 | 3.39 | 3.38 | 3.31 | 3.22 | TEPAT |
| 20 | 2 | 262 | 360 | 91.3 | 90.3 | 3.41 | 3.43 | 3.45 | 3.44 | molor |
| 20 | 2 | 649 | 333 | 86.7 | 80.7 | 3.76 | 3.57 | 3.67 | 3.83 | molor |
| 30 | 2 | 106 | 393 | 84.5 | 84.5 | 2.14 | 2.84 | 2.14 | 2.58 | TEPAT |
| 10 | 2 | 294 | 350 | 82.7 | 77.7 | 3.16 | 3.7 | 3.35 | 3.87 | TEPAT |
| 20 | 1 | 277 | 357 | 82.5 | 79.3 | 2.78 | 2.82 | 3.18 | 3.24 | TEPAT |
| Dst... | | | | | | | | | | 6 |

H. Pemodelan Data

Pada tahap ini dilakukan pemodelan dengan pengujian data kelulusan mahasiswa bidikmisi jenjang sarjana menggunakan metode K-Nearest Neighbor. Metode tersebut digunakan untuk memprediksi ketepatan waktu kelulusan mahasiswa. Setelah melakukan pengolahan data awal, data menjadi 1593 mahasiswa dengan jumlah kasus yang “tepat” waktu berjumlah 1503 mahasiswa sedangkan kasus yang “terlambat”

berjumlah 90 mahasiswa. Dan atribut yang dipakai dalam eksperimen yaitu; jenis kelamin, rata-rata nilai ijazah, rata-rata nilai UN, nilai Toefl, nilai Prestasi, program studi mahasiswa, dan IPK dari semester 1 sampai 4. Hasil pengolahan data dapat dilihat pada Gambar 2.

| Name: CLASS | | Type: Nominal | |
|-----------------|-------|----------------|--------|
| Missing: 0 (0%) | | Unique: 0 (0%) | |
| | | Distinct: 2 | |
| No. | Label | Count | Weight |
| 1 | molor | 90 | 90.0 |
| 2 | TEPAT | 1503 | 1503.0 |

Gambar. 2. Hasil Pengolahan Data

Dengan mengkluster k=1 dari data 1593 mahasiswa, 10 dimensi terdiri dari 2 kelas yaitu “TEPAT” dan “MOLOR”

I. Uji Coba

Uji Coba dilakukan dengan menggunakan 28 skenario. Skenario ditentukan berdasarkan *percentage split*, dan nilai k. Variasi *percentage split* data training dan data testing ada 4, yaitu 60-40, 70-30, 80-20, dan 90-10, sedangkan variasi nilai k ada 7, yaitu k=3, k=5, k=7, k=9, k=11, k=13, dan k=15. Rincian hasil dari uji coba ini ditunjukkan pada Tabel 3.

TABEL 3
REKAP HASIL PENGUJIAN MENGGUNAKAN PRESENTAGE SPLIT

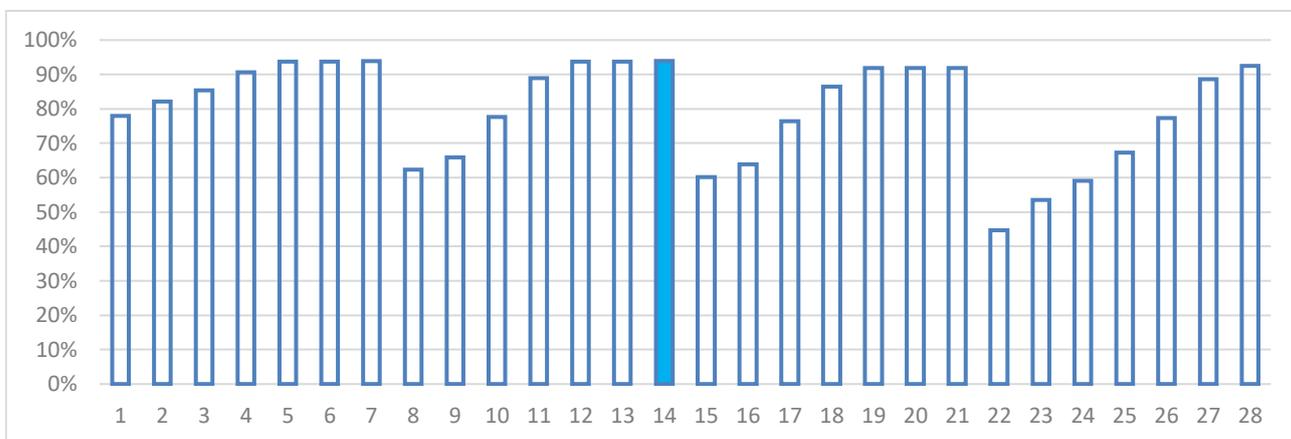
| Skenario | K | Presentage % | Correctly Classified Instances | Incorrectly Classified Instances | Total Number of Instances |
|----------|----|--------------|--------------------------------|----------------------------------|---------------------------|
| 60 | 3 | | 78.02% | 21.98% | 637 |
| 60 | 5 | | 82.10% | 17.90% | 637 |
| 60 | 7 | | 85.40% | 14.60% | 637 |
| 60 | 9 | 60 | 90.58% | 9.42% | 637 |
| 60 | 11 | | 93.72% | 6.28% | 637 |
| 60 | 13 | | 93.72% | 6.28% | 637 |
| 60 | 15 | | 93.88% | 6.12% | 637 |
| 70 | 3 | | 62.34% | 37.66% | 478 |
| 70 | 5 | | 65.90% | 34.10% | 478 |
| 70 | 7 | | 77.62% | 22.38% | 478 |
| 70 | 9 | 70 | 88.91% | 11.09% | 478 |
| 70 | 11 | | 93.72% | 6.28% | 478 |
| 70 | 13 | | 93.72% | 6.28% | 478 |
| 70 | 15 | | 93.93% | 6.07% | 478 |
| 80 | 3 | | 60.19% | 39.81% | 319 |
| 80 | 5 | | 63.95% | 36.05% | 319 |

| | | | | | |
|----|------|--------|--------|--------|-----|
| 17 | K 7 | 76.49% | 23.51% | 319 | |
| 18 | K 9 | 86.52% | 13.48% | 319 | |
| 19 | K 11 | 91.85% | 8.15% | 319 | |
| 20 | K 13 | 91.85% | 8.15% | 319 | |
| 21 | K 15 | 91.85% | 8.15% | 319 | |
| 22 | K 3 | 44.65% | 55.35% | 159 | |
| 23 | K 5 | 53.46% | 46.54% | 159 | |
| 24 | K 7 | 59.12% | 40.88% | 159 | |
| 25 | K 9 | 90 | 67.30% | 32.70% | 159 |
| 26 | K 11 | 77.36% | 22.64% | 159 | |
| 27 | K 13 | 88.68% | 11.32% | 159 | |
| 28 | K 15 | 92.45% | 7.55% | 159 | |

bidikmisi jenjang sarjana. Pengujian dilakukan menggunakan 28 skenario dengan hasil nilai optimal pada pengujian presentage split 70-30 dengan hasil akurasi sebesar 93.93% dengan nilai $k=15$.

DAFTAR PUSTAKA

- [1] “Permenristekdikti-6-2019-Bantuan-Biaya-Pendidikan-bagi-Mahasiswa-Miskin-Berprestasi.pdf.”
- [2] N. D. Ayuningtyas, J. Sayono, and E. S. Andayani, “Problematika Mahasiswa 1B059idikmisi Prodi Pendidikan Sejarah Angkatan 2013—2017,” p. 7.
- [3] “PERATURAN PEMERINTAH



Gambar 3 Hasil pengujian menggunakan 28 skenario

J. Evaluasi

Evaluasi dilakukan dengan melihat hasil dari ke-28 skenario yang merupakan kombinasi *percentage split* dan nilai k . Dari hasil uji coba didapatkan nilai akurasi terbaik pada skenario ke-14 dengan nilai 93.93% yaitu dengan nilai *percentage split* 70-30 dan nilai $k=15$. Secara umum hasil uji coba menunjukkan bahwa semakin besar nilai k , maka nilai akurasi semakin baik. Hal ini terlihat dari berbagai skenario *percentage split*, nilai akurasi terbaik terdapat pada nilai $k=15$. Hasil lengkap dari uji coba dapat dilihat pada Gambar 3.

4. KESIMPULAN

Hasil penelitian menunjukkan bahwa metode K-Nearest Neighbor dapat digunakan sebagai pengolahan prediksi kelulusan mahasiswa

REPUBLIK INDONESIA NOMOR 66
TAHUN 2010.pdf.”

- [4] G. H. Herlambang, A. Nugroho, and B. Zaman, “KLASIFIKASI PERKIRAAN KELULUSAN MAHASISWA JENJANG MAGISTER MENGGUNAKAN METODE NAÏVE BAYES,” *NJCA Nusant. J. Comput. Its Appl.*, vol. 5, no. 1, pp. 40–46, 2020.
- [5] Beni-Suef University *et al.*, “Credibility Detection in Twitter Using Word N-gram Analysis and Supervised Machine Learning Techniques,” *Int. J. Intell. Eng. Syst.*, vol. 13, no. 1, pp. 291–300, Feb. 2020, doi: 10.22266/ijies2020.0229.27.
- [6] A. R. Pratama, M. Mustajib, and A. Nugroho, “Deteksi Citra Uang Kertas dengan Fitur RGB Menggunakan K-Nearest Neighbor,” *J. Eksplora Inform.*, vol. 9, no. 2, pp. 163–172, 2020.

-
- [7] “Rohman - 2015 - MODEL ALGORITMA K-NEAREST NEIGHBOR (K-NN) UNTUK PR.pdf.”
- [8] A. Nugroho, S. Sumpeno, K. Fujisawa, and M. H. Purnomo, “A Decision Guidance for Solving Success Rate Political Campaign Using Distance Weighted kNN in Nassi-Shneiderman Framework”.
- [9] “Priandini et al. - 2017 - Categorizing document by fuzzy C-Means and K-neare.pdf.”
- [10] “Ayuningtyas et al. - Problematika Mahasiswa 1B059idikmisi Prodi Pendi.pdf.”
- [11] B. Zaman, A. Justitia, K. N. Sani, and E. Purwanti, “An Indonesian Hoax News Detection System Using Reader Feedback and Naïve Bayes Algorithm,” *Cybern. Inf. Technol.*, vol. 20, no. 1, pp. 82–94, Mar. 2020, doi: 10.2478/cait-2020-0006.
- [12] A. Sukma, B. Zaman, and E. Purwanti, “Information Retrieval Document Classification with K-Nearest Neighbor,” *Rec. Libr. J.*, vol. 1, no. 2, Art. no. 2, 2015, doi: 10.20473/rlj.V1-I2.2015.129-138.